Multi-block, Multi-set, Multi-level and Data Fusion Methods

EIGENVECTOR RESEARCH INCORPORATED

Outline

- Definitions
- Multi-level data
- DOE, crossed and nested designs
- ASCA – ANOVA simultaneous component analysis
  - Example
- MLSCA – Multi-level simultaneous component analysis.
  - Example
- Multi-block
- Alignment
- Examples

EIGENVECTOR RESEARCH INCORPORATED

Definitions

- Single-block: data that is logically contained in a single matrix
- Two-block: two single block data sets that share a common mode (typically the sample mode)
- Multi-block: multiple single blocks that share a common mode
- Multi-set: groups of related samples that have the same variables, typically from designed experiments
- Multi-level: same as multi-set except typically from nested or happenstance designs

3

EIGENVECTOR RESEARCH INCORPORATED

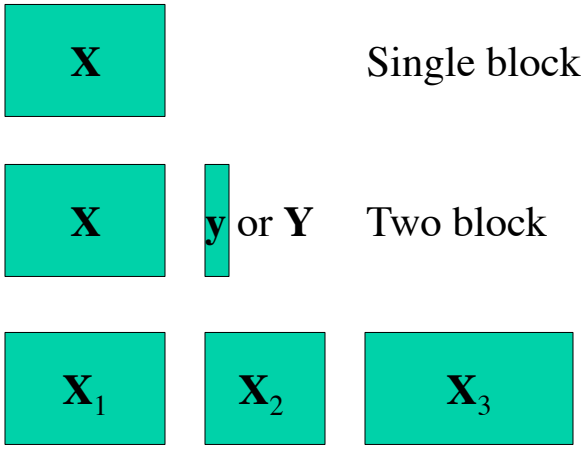
Definitions (cont.)

- Multi-way: Data that is logically arranged in 3-way (or more) arrays
- Data fusion: the process of combining multiple sources of data to improve accuracy
- Alignment: the process of matching the axes (time, wavelength, evolution, spatial) of two data sets along one or more modes

4

EIGENVECTOR RESEARCH INCORPORATED

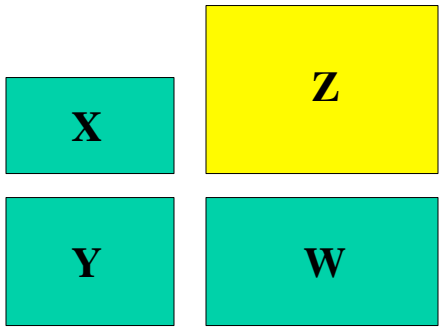
Single, Two and Multi-block



5

EIGENVECTOR RESEARCH INCORPORATED

L- and U- Configurations



Not going to cover this!

S. Wold, S. Hellberg, T. Lundsted, M. Sjöström, H. Wold, (1987), PLS modeling with latent variables in two or more dimensions. In: *Proceedings: PLS Model Building: Theory and Applications*. Symposium Frankfurt am Mein September 23-25, 1987.

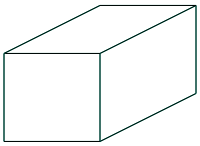
H, Martens, E, Anderssen, A. Flatberg, L. H. Gidskehaug, M. Høy, F. Westad, A. Thybo and M. Martens (2003) : Regression of a data matrix on descriptors of both its rows and of its columns via latent variables: L-PLSR. Computational Statistica & Data Analysis, 48(1), pps 103-123.

H. Martens, (2005) Domino PLS: A framework for multi-directional path modeling. Proc. PLS'05 Intl Symposium "PLS and related methods". (Eds. T. Aluja, J. Casanovas, V.E. Vinzi, A. Morineau, M. Tenenhaus) SPAD Groupe Test&Go), pp125-132.

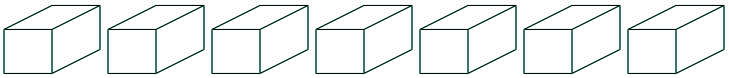
6

EIGENVECTOR RESEARCH INCORPORATED

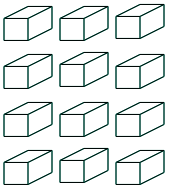
Multi-way



3-way or 3-mode



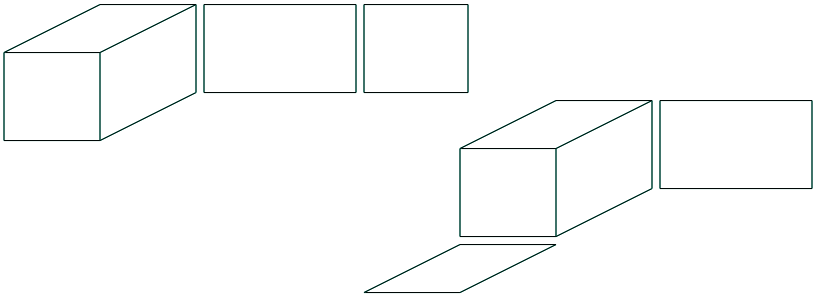
4-way



5-way

EIGENVECTOR RESEARCH INCORPORATED

Combinations



Coupled matrix and tensor factorizations:

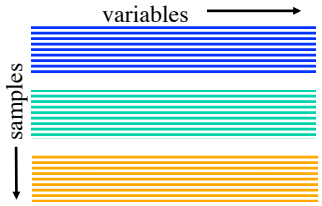
E. Acar, T. G. Kolda, and D. M. Dunlavy. All-at-once Optimization for Coupled Matrix and Tensor Factorizations. KDD Workshop on Mining and Learning with Graphs, 2011.

E. Acar, M. A. Rasmussen, F. Savorani, T. Næs, and R. Bro. Understanding Data Fusion within the Framework of Coupled Matrix and Tensor Factorizations, Submitted (May, 2012)

EIGENVECTOR RESEARCH INCORPORATED

Multi-set Data

- Groups (sets) of related samples which have the same variables.



Differences between groups may hide variability inherent to all samples.
For samples grouped according to an experimental design we can separate variability due to each design factor, and systematic variability independent of the factors. This is the purpose of ASCA and MLSCA

9

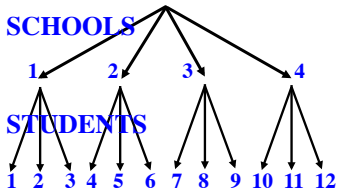
EIGENVECTOR RESEARCH INCORPORATED

Crossed and nested designs

- Crossed (factorial) designs: One or more factors with samples measured for every combination of factor levels.

		Treatment			
		A	B	C	D
Dose	1.1				
	2.0				
	3.5				

- Nested designs: samples belong to groups which are organized hierarchically.



These are both 2-factor designs

10

EIGENVECTOR RESEARCH INCORPORATED

Crossed and Nested Designs

For experimental design with 2 factors, A and B, the data matrix X can be decomposed:

Crossed designs:

$$X = X_{avg} + X_A + X_B + X_{AB} + E$$

Nested design:

$$X = X_{avg} + X_A + X_{B(A)} + E$$

X_{avg}:	matrix with column averages of X for each row.
X_A:	matrix with level averages for factor A. Similarly for X_B.
$X_{B(A)}$:	matrix with level averages for factor B at a given A level.
X_{AB}:	matrix with level averages for interaction between factors A and B.
E:	matrix with residuals.

EIGENVECTOR RESEARCH INCORPORATED

Sum of Squares Decomposition

For such designs the sum of squares can be decomposed into contributions from each factor (and interactions) and the within group (residual):

$$||\mathbf{X}||^2 = ||\mathbf{X}_{avg}||^2 + ||\mathbf{X}_A||^2 + ||\mathbf{X}_B||^2 + ||\mathbf{X}_{AB}||^2 + ||\mathbf{E}||^2$$

offset *----------between----------* *within*

ASCA and MLSCA are exploratory analysis methods which use this separation to isolate variability associated with each factor and reveal systematic variability inherent to the samples but not related to the factors.

EIGENVECTOR RESEARCH INCORPORATED

ASCA

ANOVA Simultaneous Component Analysis

For multivariate datasets based on crossed experimental designs, ASCA applies ANOVA decomposition and dimension reduction (PCA) to :

- Separate the variability associated with each factor.
- Estimate contribution of each factor to total variance.
- Test main factor and interaction effects for significance.
- View scores and loadings for these effects.

Especially useful for high-dimension datasets where traditional ANOVA is not possible.

EIGENVECTOR RESEARCH INCORPORATED

ASCA Method

- X data matrix, with 2 factors A and B.
- Decompose into DOE components

$$\mathbf{X} = \mathbf{X}_{avg} + \mathbf{X}_A + \mathbf{X}_B + \mathbf{X}_{AB} + \mathbf{E}$$

- Build PCA model for each main effect and interaction

$$\mathbf{X} = \mathbf{X}_{avg} + \mathbf{T}_A\mathbf{P}_A^T + \mathbf{T}_B\mathbf{P}_B^T + \mathbf{T}_{AB}\mathbf{P}_{AB}^T$$

- Calculate permutation P-value to estimate each factor's significance.
- Project residuals onto each PCA sub-model.

EIGENVECTOR RESEARCH INCORPORATED

ASCA Demo data: asca_data

X: Measured glucosinolate levels in cabbage plants,
3 treatments, Control, Root, Shoot.
4 time points, Days 1, 3, 7, and 14.
5 replicates for each time-treatment.
11 measured concentrations.

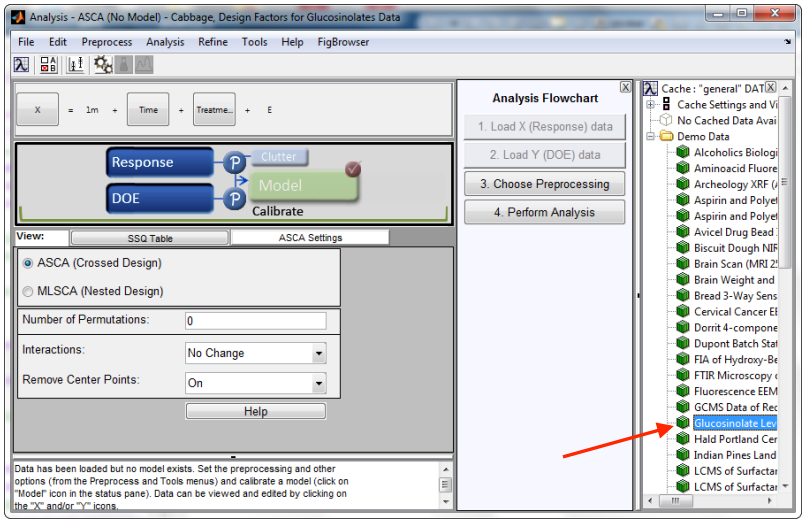
X: (60, 11)
F: (60, 2) design matrix.
See X.description for details.

		Time (Day)			
		1	3	7	14
Treatment	C				
	R	5 replicates each			
	S				

EIGENVECTOR RESEARCH INCORPORATED

15

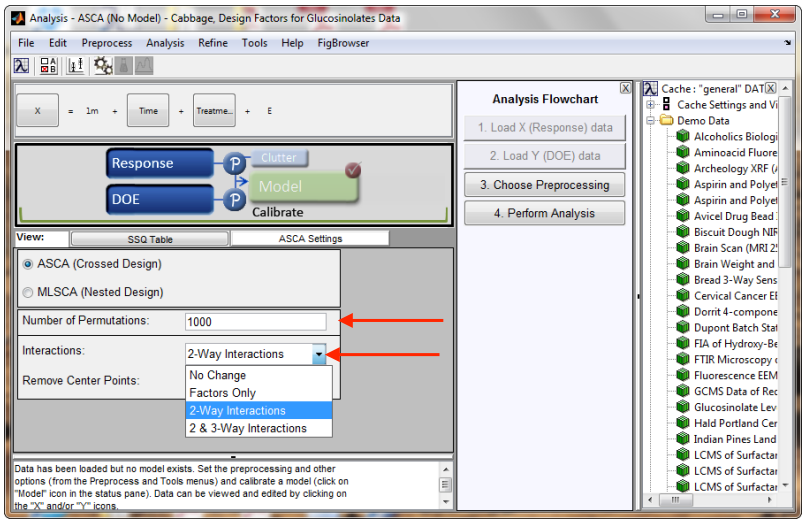
Using ASCA from the GUI



EIGENVECTOR RESEARCH INCORPORATED

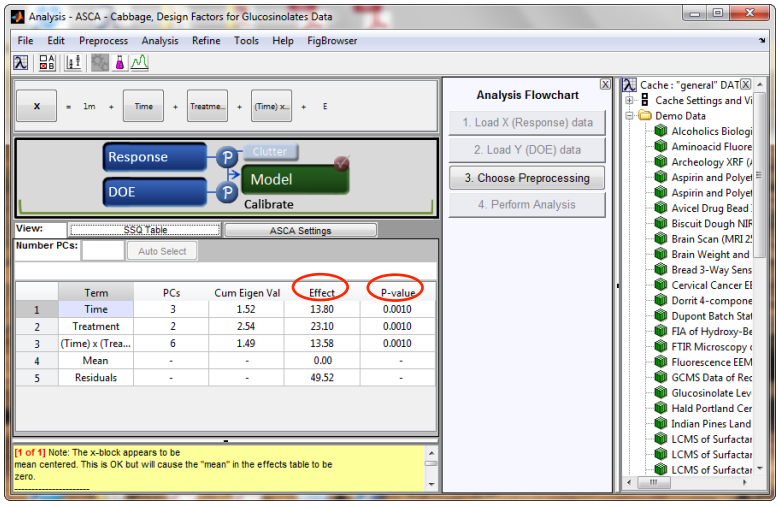
16

Using ASCA from the GUI



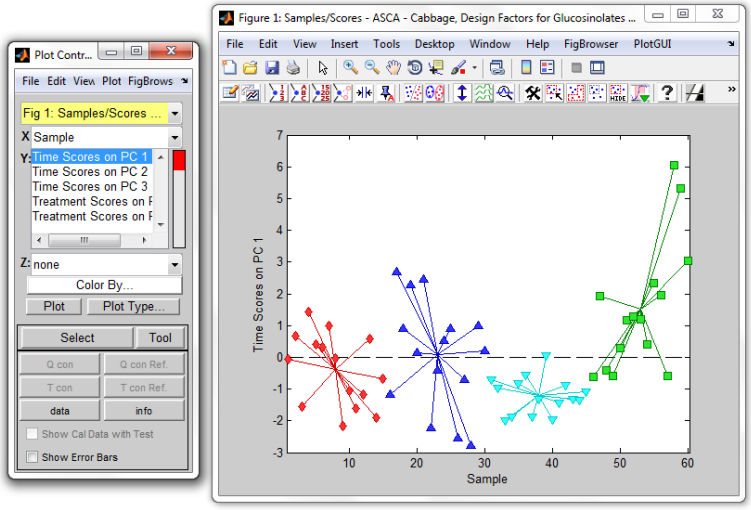
EIGENVECTOR RESEARCH INCORPORATED

Built ASCA



EIGENVECTOR RESEARCH INCORPORATED

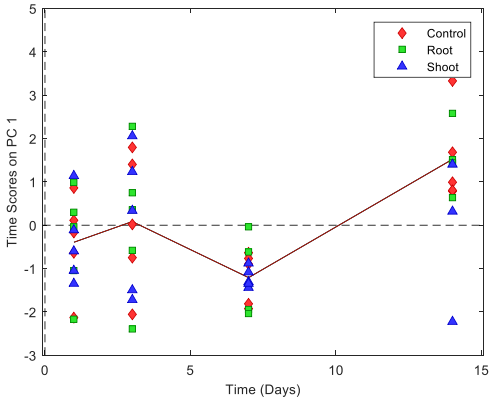
ASCA Scores Plot



EIGENVECTOR RESEARCH INCORPORATED

ASCA Scores Plot

"Time" factor sub-model, PC 1



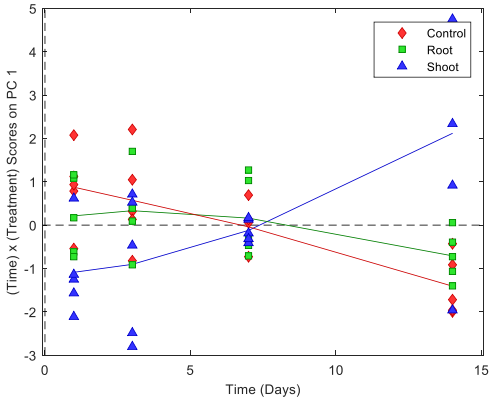
PC 1 of Time dependency common to all Treatments.

Class = Treatment. Connect Classes = Mean at each X

EIGENVECTOR RESEARCH INCORPORATED

ASCA Scores Plot

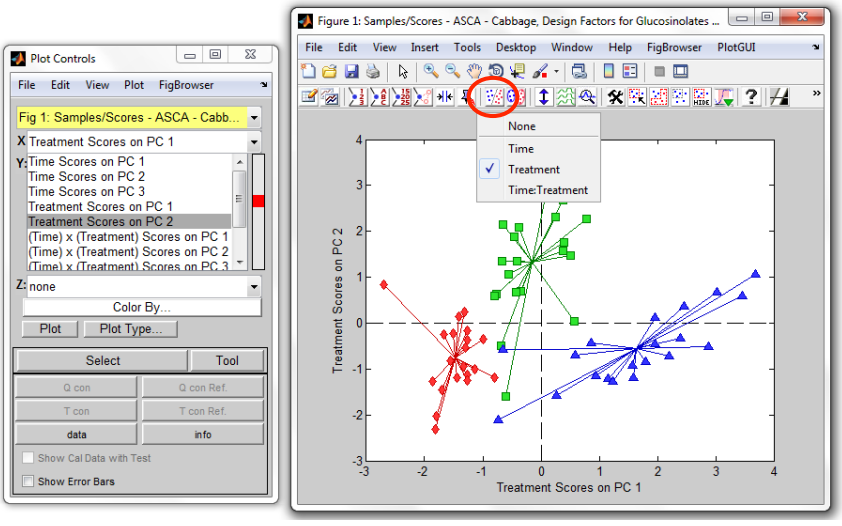
"Time" x "Treatment" interaction sub-model, PC 1



PC 1 of Time dependency at each Treatment level.
Class = Treatment. Connect Classes = Mean at each X

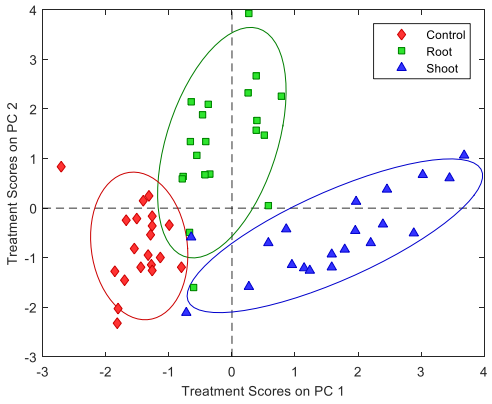
EIGENVECTOR RESEARCH INCORPORATED

ASCA Scores Plot



EIGENVECTOR RESEARCH INCORPORATED

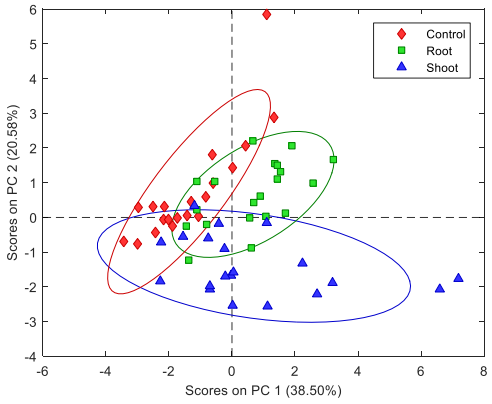
ASCA Treatment Scores Plot



Separating out the Time and Time x Treatment effects highlights the Treatment effect

EIGENVECTOR RESEARCH INCORPORATED

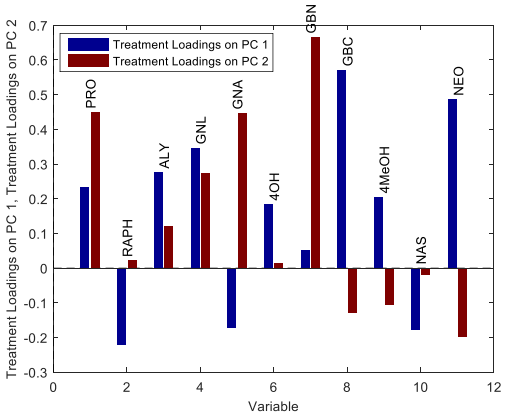
PCA Scores Plot



…better than is seen by simply applying PCA to the data.

EIGENVECTOR RESEARCH INCORPORATED

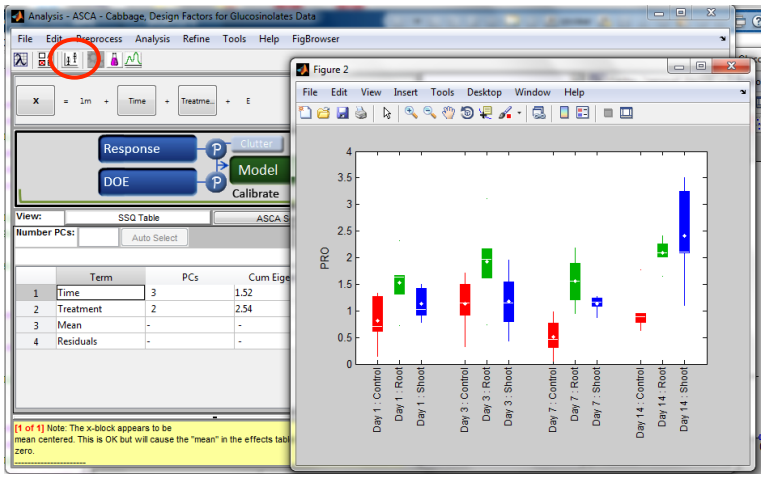
Loadings Plot



EIGENVECTOR RESEARCH INCORPORATED

ASCA Box Plot

To view raw or preprocessed X "Response" data



EIGENVECTOR RESEARCH INCORPORATED

ASCA Conclusions

ASCA allows the variation associated with each factor to be resolved, and to see the main variables involved.

- For a perturbed biological system the Time factor scores reveal the common response, Treatment factor scores show the Treatment effect independent of Time. The Time x Treatment interaction scores show the additional time dependency at each Treatment level.

EIGENVECTOR RESEARCH INCORPORATED

ASCA Conclusions, cont.

- The % contribution of each factor or interaction to the total SSQ shows which effects are important.
- Perturbation P-values for each factor estimates the probability that there is no difference between the factor level averages for this effect.

EIGENVECTOR RESEARCH INCORPORATED

MLSCA

Multi-level Simultaneous Component Analysis

MLSCA is a special case of ASCA applied to data from designed experiments with nested factors.

- Separates variability associated with each factor and residual.
- Estimate contribution of each factor to total sum of squares.
- View scores and loadings for these effects.
- Also builds PCA model on the residuals, or "within" variability. "Within" is often the focus of the analysis.
- Note that "Class Center" pre-processing can achieve same result if there is a single nesting factor.

EIGENVECTOR RESEARCH INCORPORATED

MLSCA

- 2-level data (replicates grouped by 1 factor),

  Samples drawn from N participants.

  X_{ik_i}= measurement for participant i $(i = 1, \dots, N)$, for sample k_i $(k = 1, \dots, K_i)$.
- Use the level information to decompose into a constant, a "between" participants and a "within" participants part:
- $$X_{ik_i} = X_{**} + (X_{i*} - X_{**}) + (X_{ik_i} - X_{i*})$$

  "constant" "between" "within"

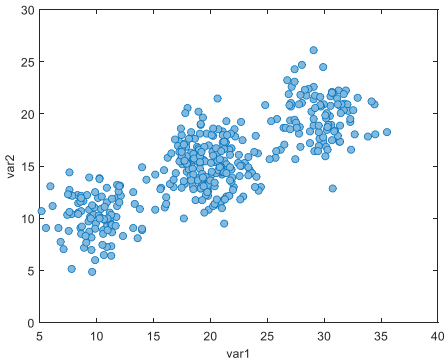
  * indicates "averaging".

EIGENVECTOR RESEARCH INCORPORATED

MLSCA: simple example

MLSCA can be used to reveal systematic variability within grouped samples which can be obscured by inter-group differences.

Example: X: (400,2) 400 samples from 3 individuals.



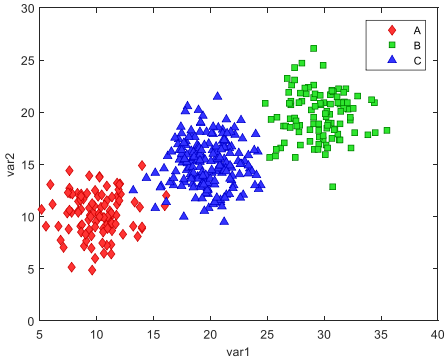
31

EIGENVECTOR RESEARCH INCORPORATED

MLSCA: simple example

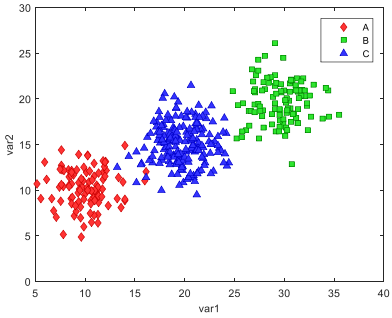
Example: X: (400,2) 400 samples from 3 individuals, A, B, and C.

Need to remove offsets for each individual to see the internal, "within" individual variation.

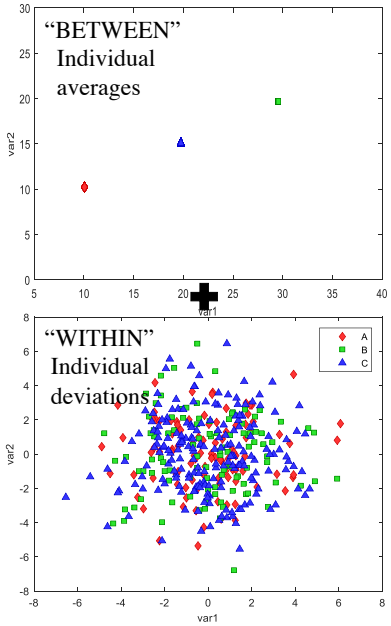


32

EIGENVECTOR RESEARCH INCORPORATED



=



X = average for each individual
+ deviations from that

EIGENVECTOR RESEARCH INCORPORATED

Nested dataset "mlsca_data"

12 engineering variables from a LAM 9600 Metal Etcher over the course of etching 107 wafers.

- Three experiments were run at different times.
- Experiment have 34, 36 and 37 wafers each, for 107 unique wafers.
- 80 samples (replicates) measured for each wafer during etching.
- X is (8560, 12)

	EXPERIMENT												
	1				2				3				
WAFER	1	2	•••	34	35	36	•••	70	71	72	•••	107
80 REPLI-CATES	x	x		x	x	x		x	x	x		x
	x	x		x	x	x		x	x	x		x
	x	x		x	x	x		x	x	x		x
	.	.		.	.	.		.	.	.		.
	.	.		.	.	.		.	.	.		.
	.	.		.	.	.		.	.	.		.
	x	x		x	x	x		x	x	x		x

Nested factors are not crossed.

EIGENVECTOR RESEARCH INCORPORATED

MLSCA Method

- X data matrix, with 2 nested factors A and B.
- Decompose into DOE components

$\mathbf{X} = \mathbf{X}_{avg} + \mathbf{X}_A + \mathbf{X}_{B(A)} + \mathbf{E}$

$\mathbf{X}_A$ contains factor A level averages

$\mathbf{X}_{B(A)}$ contains factor B level averages for each level A

$\mathbf{E}$ are the residuals, "within" component

- Build PCA model for each effect and residual

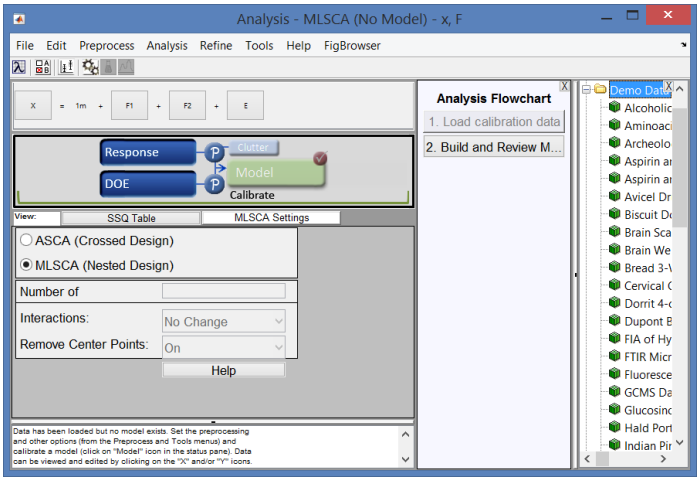
$\mathbf{X} = \mathbf{X}_{avg} + \mathbf{T}_A\mathbf{P}_A^T + \mathbf{T}_{B(A)}\mathbf{P}_{B(A)}^T + \mathbf{T}_E\mathbf{P}_E^T$

constant between A between B within

EIGENVECTOR RESEARCH INCORPORATED

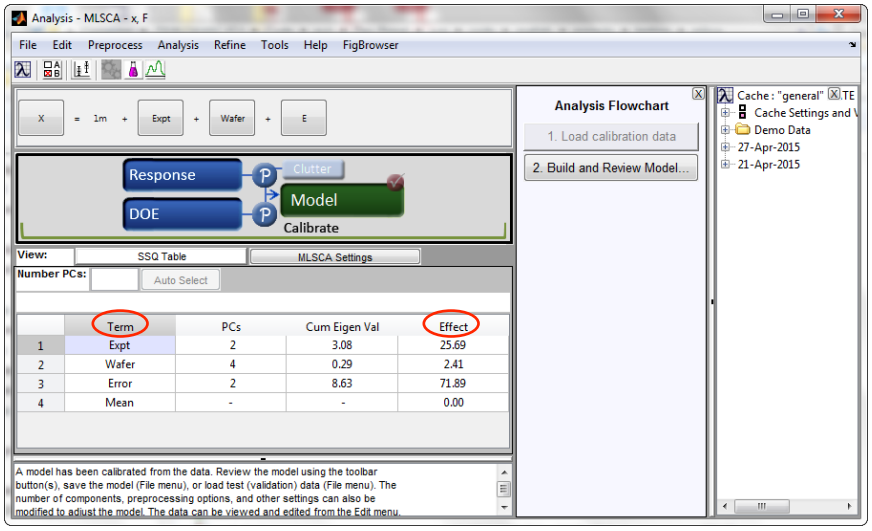
Using MLSCA from the GUI

- MLSCA located under "Design of Experiments" in browse



EIGENVECTOR RESEARCH INCORPORATED

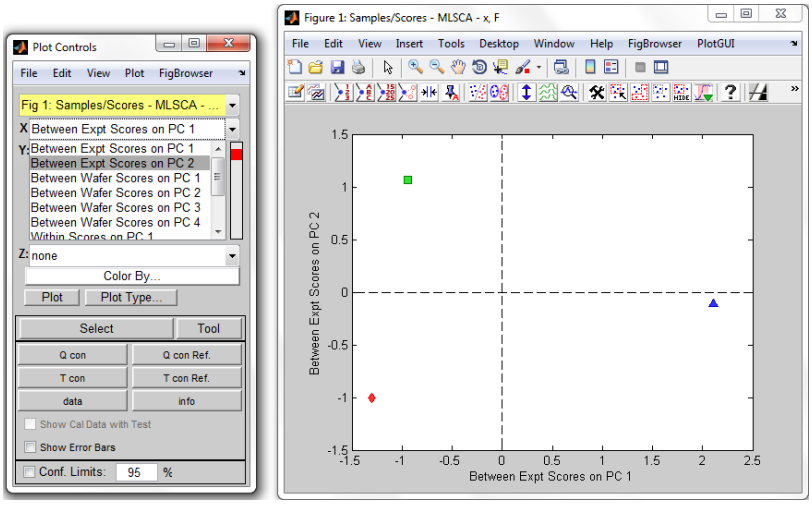
Using MLSCA from the GUI



EIGENVECTOR RESEARCH INCORPORATED

MLSCA Scores Plot

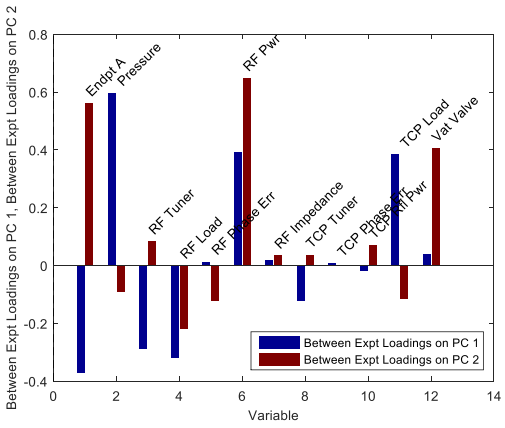
"Experiment" factor sub-model, PC 1 vs 2



EIGENVECTOR RESEARCH INCORPORATED

MLSCA Loadings Plot

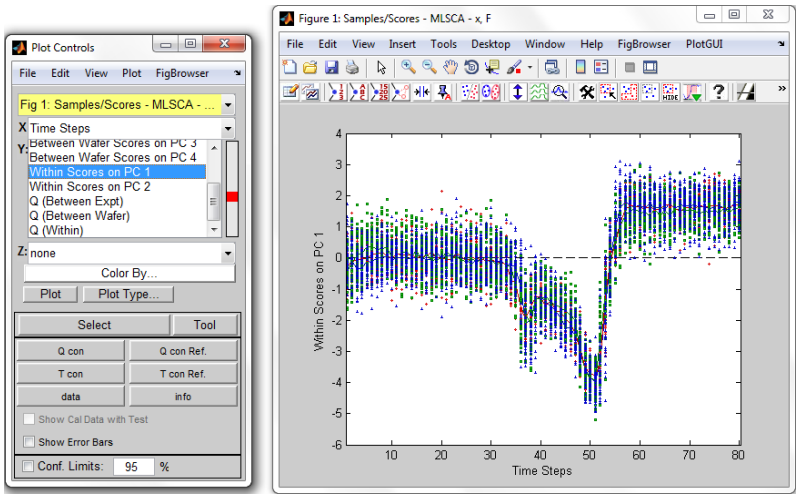
"Experiment" factor sub-model, PC 1 and 2



EIGENVECTOR RESEARCH INCORPORATED

MLSCA Scores Plot

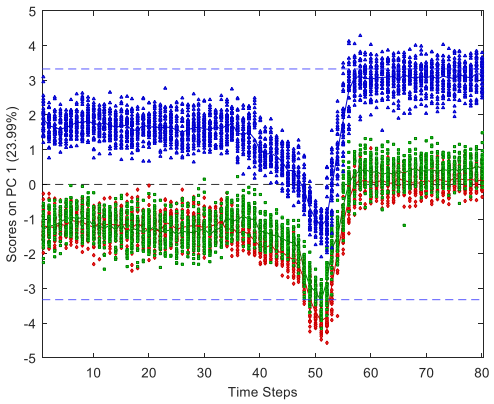
"Within" Residual sub-model, PC 1 vs. time



EIGENVECTOR RESEARCH INCORPORATED

PCA Scores Plot

PC 1 vs. time, Colored by Experiment class



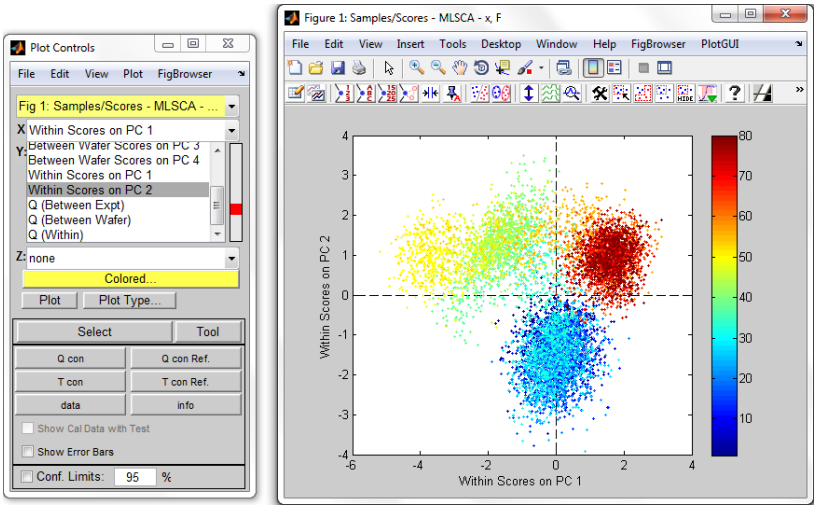
The spike at time step 47-48 is not seen in PC 1. It shows up in PC 2 because the offset between experiments dominates PC1 in simple PCA.

41

EIGENVECTOR RESEARCH INCORPORATED

MLSCA Scores Plot

"Within" sub-model, PC 1 vs 2, colored by time

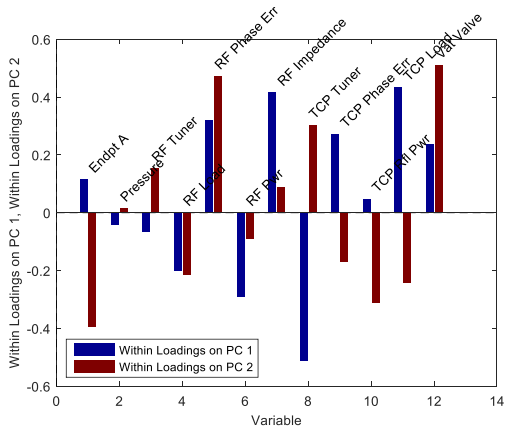


42

EIGENVECTOR RESEARCH INCORPORATED

MLSCA Loadings Plot

"Within" Residual sub-model, PC 1 and 2



EIGENVECTOR RESEARCH INCORPORATED

MLSCA Conclusions

MLSCA allows the variation associated with each nested factor to be resolved, and to see the main variables involved.

- Often used to reveal the inherent "within" group variability of samples after factor effects are removed. For process data this allows separation of within-run variation from between-run variation.
- SSQ contributions show which nested factors are important.

EIGENVECTOR RESEARCH INCORPORATED

ASCA and MLSCA

- MLSCA is a special case of ASCA.
  However, as implemented,
  ASCA = crossed designs,
  MLSCA = nested designs.
- ASCA used to study fixed effect factors while MLSCA focuses on residuals, "within" variability, of nested random effect factors.

EIGENVECTOR RESEARCH INCORPORATED

References

ASCA:

- Smilde, A.K., J.J. Jansen, H.C.J. Hoefsloot, R-J.A.N. Lamars, J. van der Greef, M.E. Timmerman, "ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data", Bioinformatics, 2005, 21, 3043-3048.
- Zwanenburg, G., H.C.J. Hoefsloot, J.A. Westerhuis, J.J. Jansen, and A.K. Smilde, "ANOVA-principal component analysis and ANOVA-simultaneous component analysis: a comparison". J. Chemometrics, 2011.

MLSCA:

- de Noord, O.E., and E.H. Theobald, Multilevel component analysis and multilevel PLS of chemical process data. J. Chemometrics 2005; 301–307
- Timmerman, M.E., Multilevel Component Analysis. Brit. J. Mathemat. Statist. Psychol. 2006, 59, 301-320.
- Jansen, J.J., H.C.J. Hoefsloot, J. van der Greef, M.E. Timmerman and A.K. Smilde, Multilevel component analysis of time-resolved metabolic fingerprinting data. Analytica Chimica Acta, 530, (2005), 173–183.

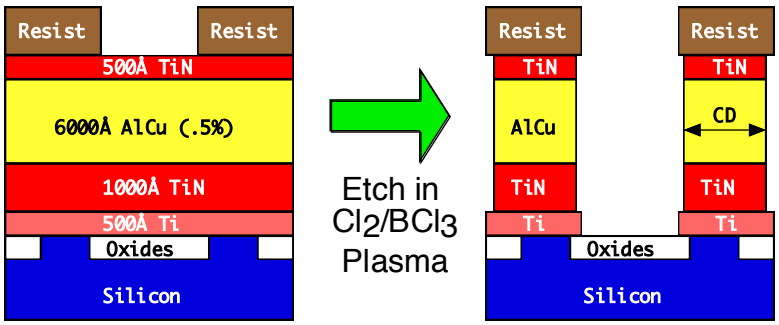
EIGENVECTOR RESEARCH INCORPORATED

Multi-block Data Fusion

- Data fusion can be done at three levels
  - Low level: single model of combined data blocks appropriately scaled/preprocessed
  - Mid level: combining scores from individual data blocks into a consensus model
  - High level: combining predictions from individual models in some sort of voting scheme

EIGENVECTOR RESEARCH INCORPORATED

Example: 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

EIGENVECTOR RESEARCH INCORPORATED

Available Measurements

- Machine State 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
    - broadband plasma emission (often used for endpoint)
    - impedance measurements
- Optical emission spectra
- RF plasma variables

EIGENVECTOR RESEARCH INCORPORATED

Sensitivity of MSPC Models

- Three experiments performed with 21 “induced” faults on:
  - TCP top power
  - RF bottom power
  - Cl2 flow
  - BCl3 flow
  - Chamber pressure
  - Helium chuck pressure
- Data available for Machine State, RF and OES
- Goal: Compare ability of models considered for detecting faults: best case and for routine data

EIGENVECTOR RESEARCH INCORPORATED

Generating Faults

- Set points were changed for controlled process variables
  - very easy to detect set point changes by simply looking at the variable for which the setting was changed, however…
- Data for the controlled variable was adjusted to have the original desired set point
  - the mean set point was reset to the original
- Result is data that looks like a sensor has developed a bias
  - more difficult to detect the fault on the single variable
  - model must detect the fault based on changes in relationships between variables
- Each wafer is analogous to batch in chem process
  - wafter-to-wafer = batch-to-batch

EIGENVECTOR RESEARCH INCORPORATED

Example with Etch Data

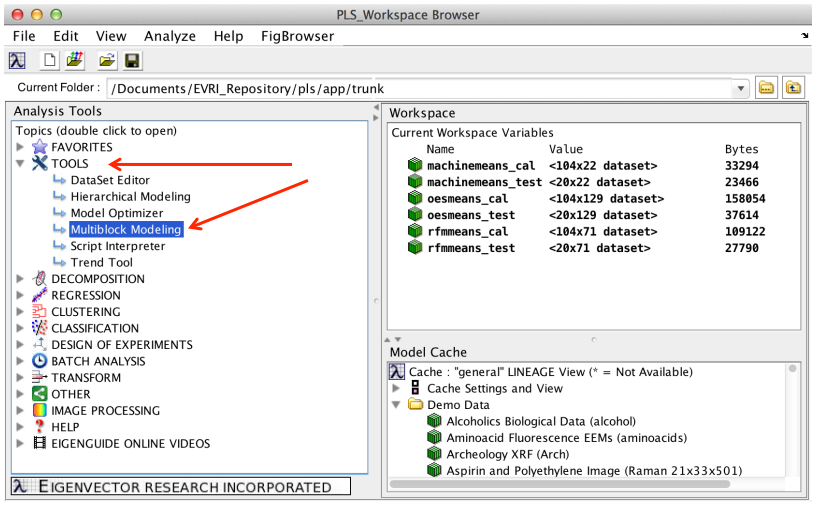
- Available data: Machine, OES and RFM data for 104 normal wafers and 20 induced faults
- Data reduced just to mean over each batch

```
>> clear
>> load Etch_Means
>> whos
  Name                    Size              Bytes  Class
  machinemeans_cal       104x22             33294  dataset
  machinemeans_test       20x22             23466  dataset
  oesmeans_cal           104x129           158054  dataset
  oesmeans_test           20x129            37614  dataset
  rfmmeans_cal           104x71            109122  dataset
  rfmmeans_test           20x71             27790  dataset

>>
```


EIGENVECTOR RESEARCH INCORPORATED

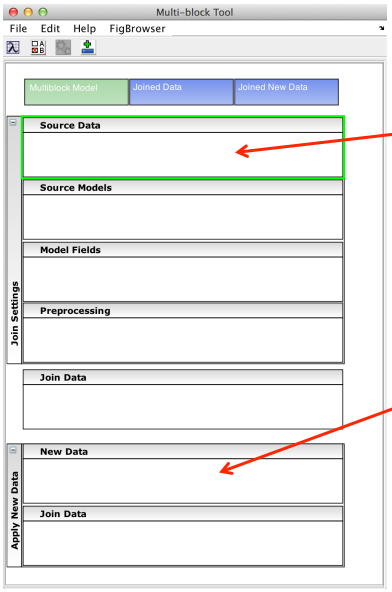
Browse Interface



53

EIGENVECTOR RESEARCH INCORPORATED

Multi-block Tool Interface



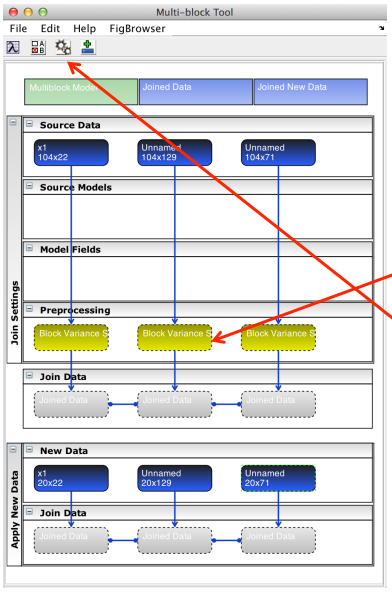
Drag calibration data sets here

Drag test data sets here

54

EIGENVECTOR RESEARCH INCORPORATED

Multi-block Interface Loaded

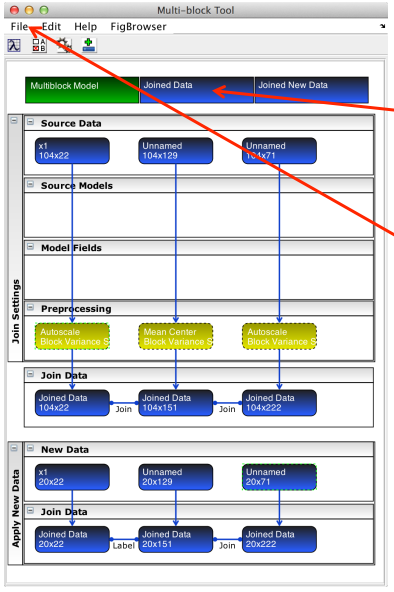


Choose preprocessing for each
Right click, "edit"

Click join data icon

EIGENVECTOR RESEARCH INCORPORATED

Joined Data

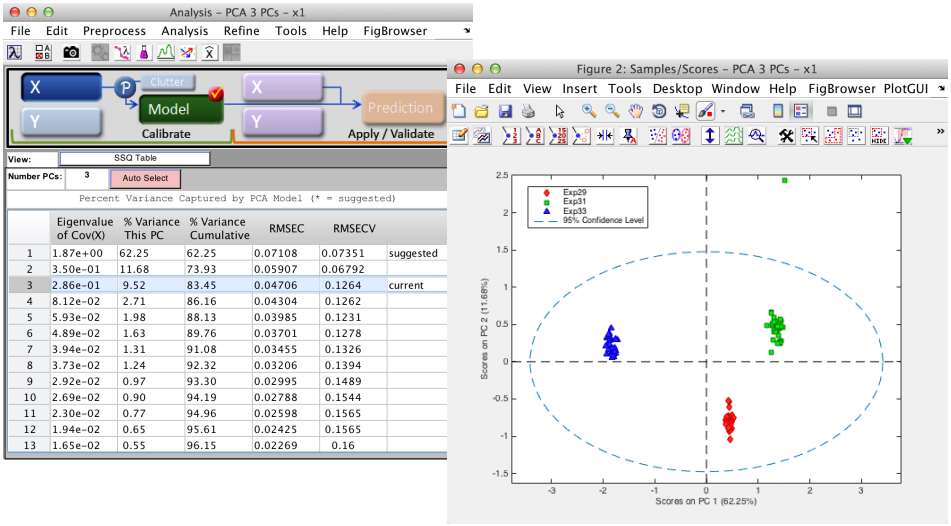


Right click "Joined Data"
Analyze--PCA

File—Save Joined New Data

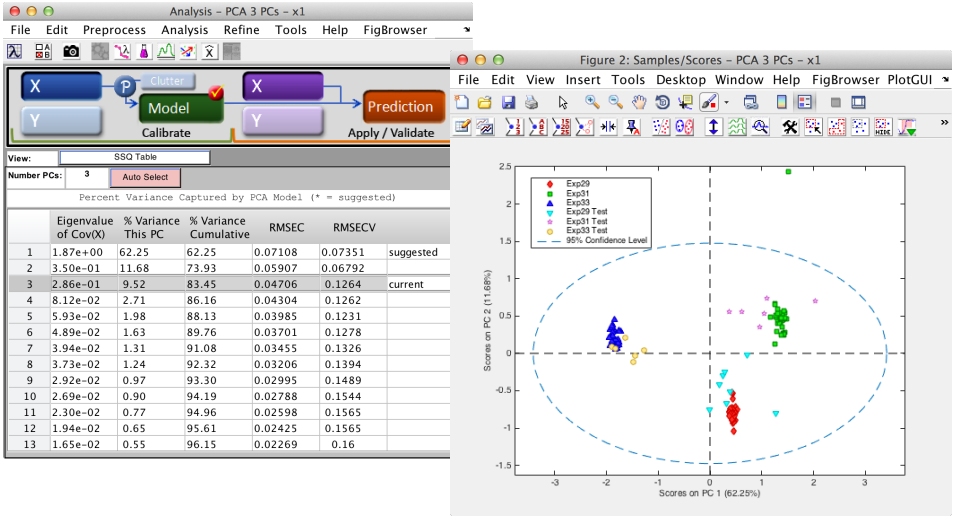
EIGENVECTOR RESEARCH INCORPORATED

Data pushed into PCA



EIGENVECTOR RESEARCH INCORPORATED

With Test Data Loaded



EIGENVECTOR RESEARCH INCORPORATED

Redo at Mid-level

- Develop individual PCA models of data blocks
- Load models into Multi-block tool
- Choose model outputs
- Join and push into PCA

Alignment

- Many data fusion problems involve aligning one block of data with another
- There are many ways of doing alignment!

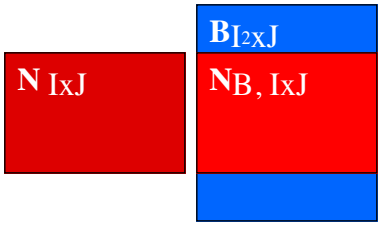
Alignment and Warping Methods

- A veritable smorgasbord of methods available
  - Rank minimization
    - similar to rank annihilation factor analysis (RAFA)
  - 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

Multi-way Alignment

- For test matrix $\mathbf{B}_{I_2xJ}$ and standard matrix $\mathbf{N}_{IxJ}$ (where $I_2>I$) $\mathbf{N}_B$ is the IxJ sub-matrix of $\mathbf{B}$ that minimizes the rank of $[\mathbf{N}|\mathbf{N}_B]$

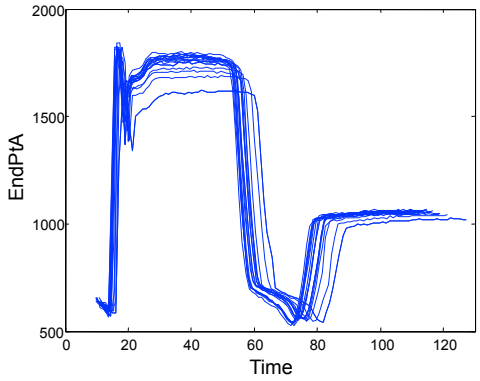


- Can also minimize residuals of projecting N_B onto a bilinear model of N
- Optimize based on sum of eigenvalues to the fourth power

EIGENVECTOR RESEARCH INCORPORATED

Process Data Alignment

- Typical process data has a different number of time points for each wafer processed.
- MPCA requires the same number *M* for each wafer.



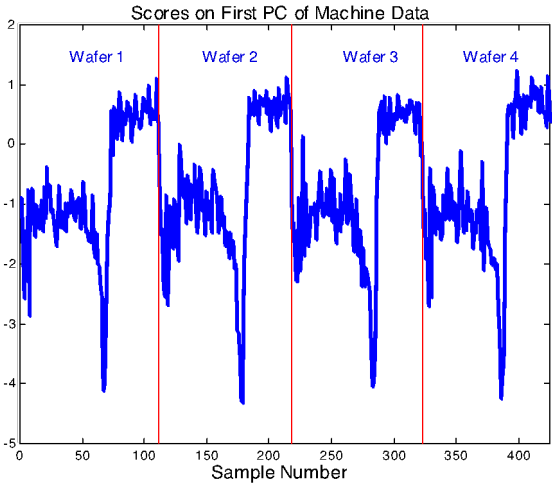
- Misalignment adds rank irrelevant to process monitoring.
  - PCs must account for time shifts in the process data.
  - Irrelevant variance often results in a reduction of model sensitivity.

63

EIGENVECTOR RESEARCH INCORPORATED

PCA Scores of Machine Data

- Use PCA scores to characterize process trace
- Data centered around "peak" in TiN etch or transition from process step 4 to 5
  - simple!



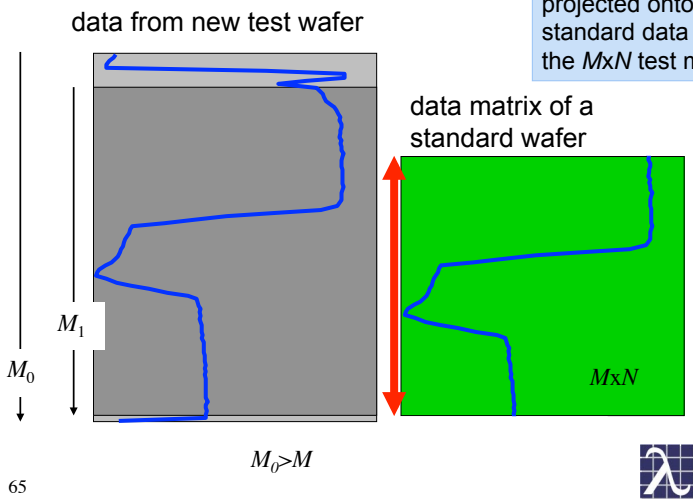
64

EIGENVECTOR RESEARCH INCORPORATED

Alignment Algorithm

- Based on rank minimization

The submatrix of the test wafer that minimizes the residuals when projected onto a PCA model of the standard data matrix is selected as the *MxN* test matrix.

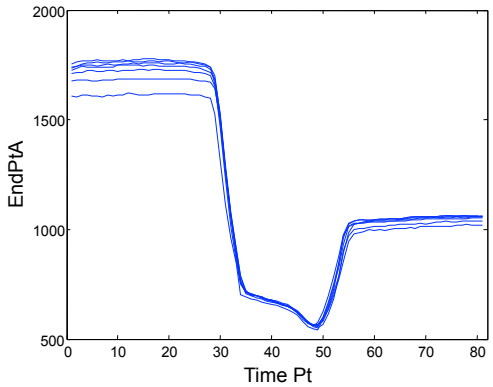


65

EIGENVECTOR RESEARCH INCORPORATED

Aligned Process Data

- Alignment is based on aligning all process variables simultaneously.
  - Data at the very beginning and end of process are not used.
  - Monitoring is based on the "interesting" part of the process.



66

EIGENVECTOR RESEARCH INCORPORATED

Correlation Optimized Warping

- Piecewise preprocessing method
- Allows limited changes in segment lengths, controlled by slack parameter
- Linear interpolation over segments
- Dynamic programming used to optimize correlation between warped sample and reference
- Less flexible than DTW (unless constrained)

EIGENVECTOR RESEARCH INCORPORATED

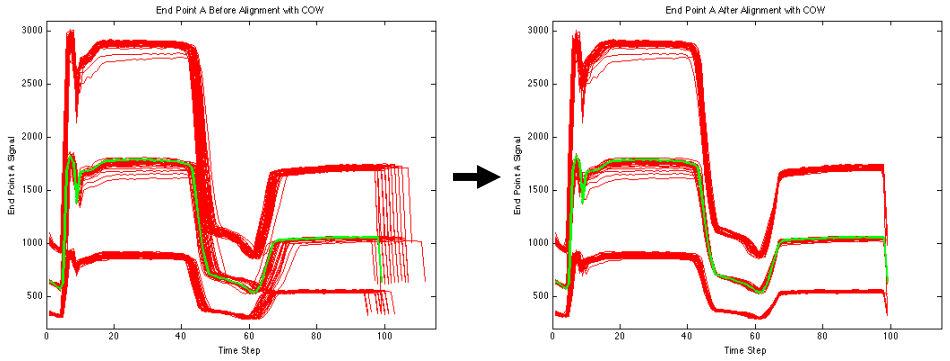
COW References

- 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

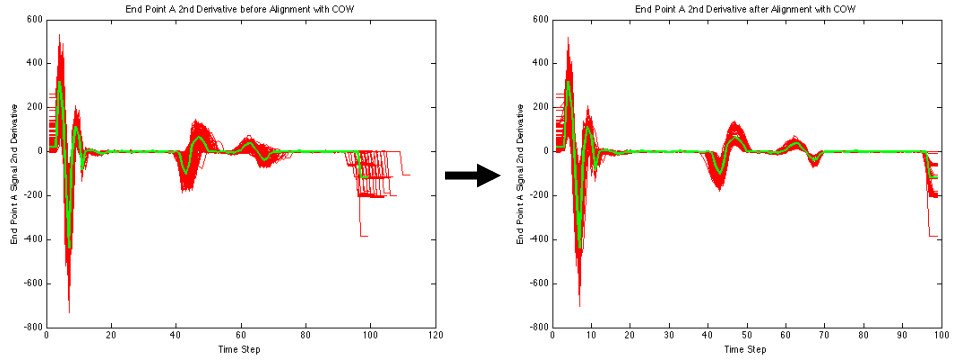


69

EIGENVECTOR RESEARCH INCORPORATED

Hints on COW

May be better to calculate warp with 2nd derivative
Apply calculated warp to other variables
Calculate warp on PCA scores or other latent variable



70

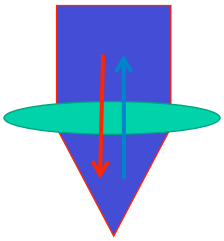
EIGENVECTOR RESEARCH INCORPORATED

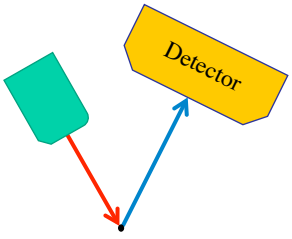
Example from Image Analysis

71

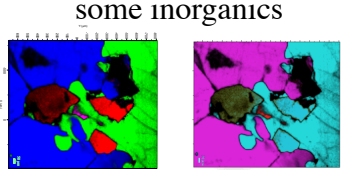
EIGENVECTOR RESEARCH INCORPORATED

Multi-Method Imaging

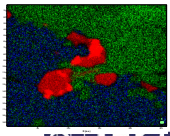




Raman & Fluorescence
Detect molecular species, some inorganics



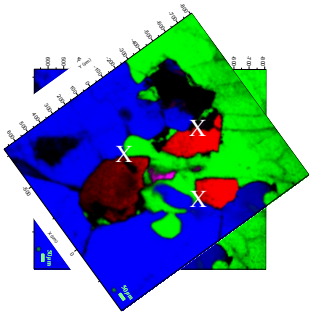
Energy Dispersive X-Ray Fluorescence (EDXRF)
Detects elemental composition

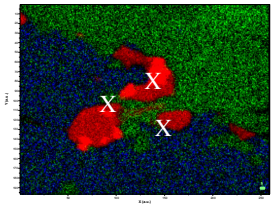


EIGENVECTOR RESEARCH INCORPORATED

Data Fusion

- Alignment of Images
- Balancing of Data Variance
- Concatenation

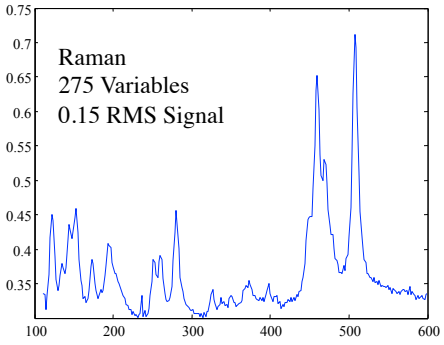


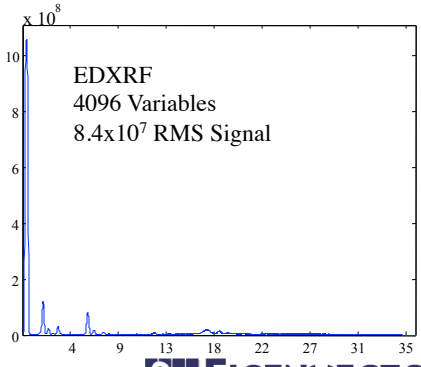


EIGENVECTOR RESEARCH INCORPORATED

Data Fusion

- Alignment of Images
- Balancing of Data Variance
- Concatenation





EIGENVECTOR RESEARCH INCORPORATED

Data Fusion

- Alignment of Images
- Balancing of Data Variance
- Concatenation

Raman
275 Variables
0.15 RMS Signal

EDXRF
4096 Variables
8.4×10^7 RMS Signal

$$\frac{X_{Raman}}{0.15 \times 275} \qquad \frac{X_{EDXRF}}{8.4 \times 10^7 \times 4096}$$

EIGENVECTOR RESEARCH INCORPORATED

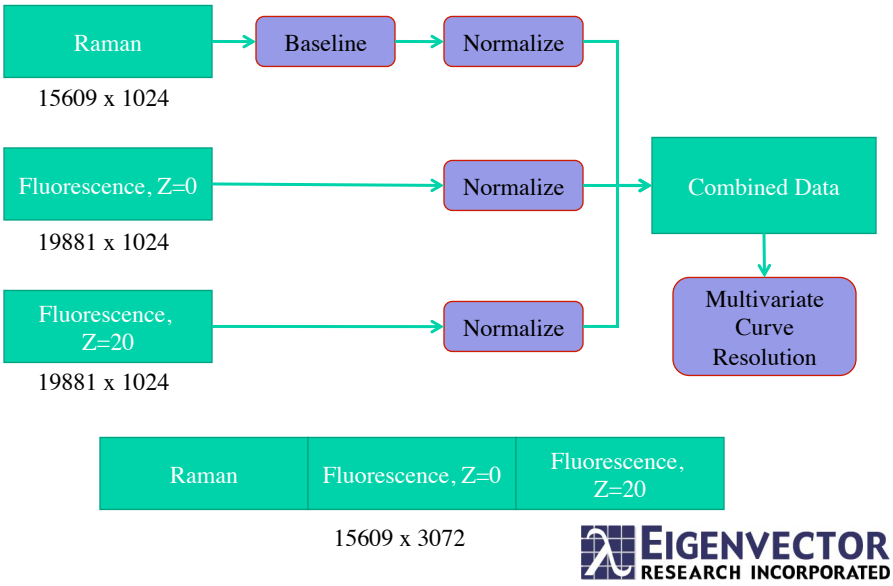
Data Fusion

- Alignment of Images
- Balancing of Data Variance
- Concatenation

$$X_{Fused} = [X_{Raman} \quad X_{EDXRF}]$$

EIGENVECTOR RESEARCH INCORPORATED

Pre-Preprocessing



EIGENVECTOR RESEARCH INCORPORATED

Data Analysis

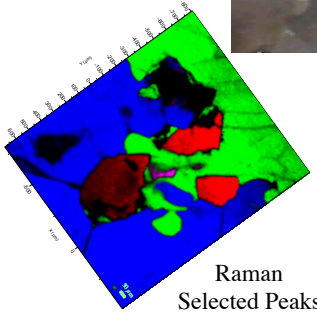
- Multivariate Curve Resolution
- With Contrast "bias" – Pushes solution to a specified edge of the feasible bounds
- Provides solution within noise level which is most consistent with:
  - Image/conc. contrast: best spatial resolution
  - Spectral contrast: best spectral resolution

EIGENVECTOR RESEARCH INCORPORATED

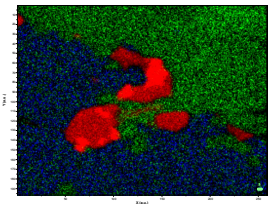
Image of Granite

White Light
Dimensions 2 x 2 mm

George J. Havrilla,
Ursula Fittschen
Los Alamos National Laboratory
Los Alamos, NM



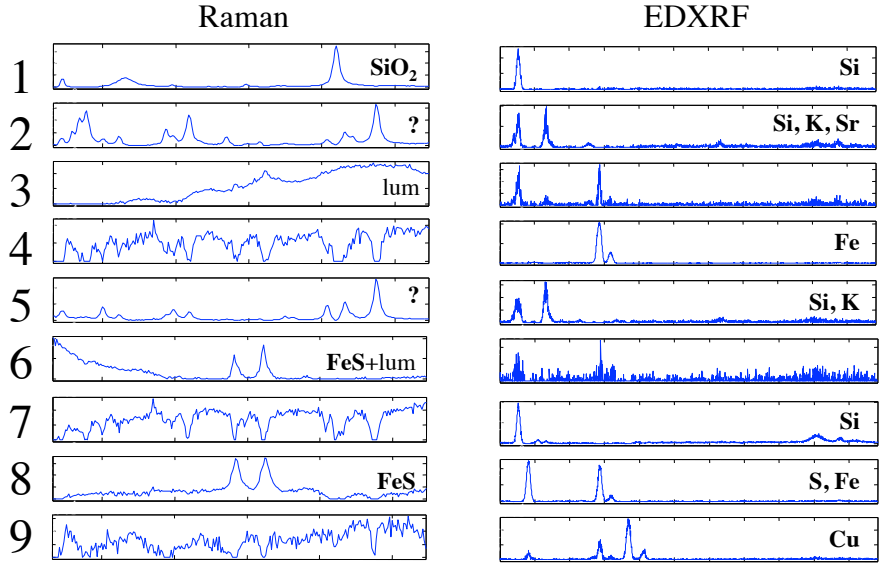
Raman
Selected Peaks



EDXRF
Selected Channels

EIGENVECTOR RESEARCH INCORPORATED

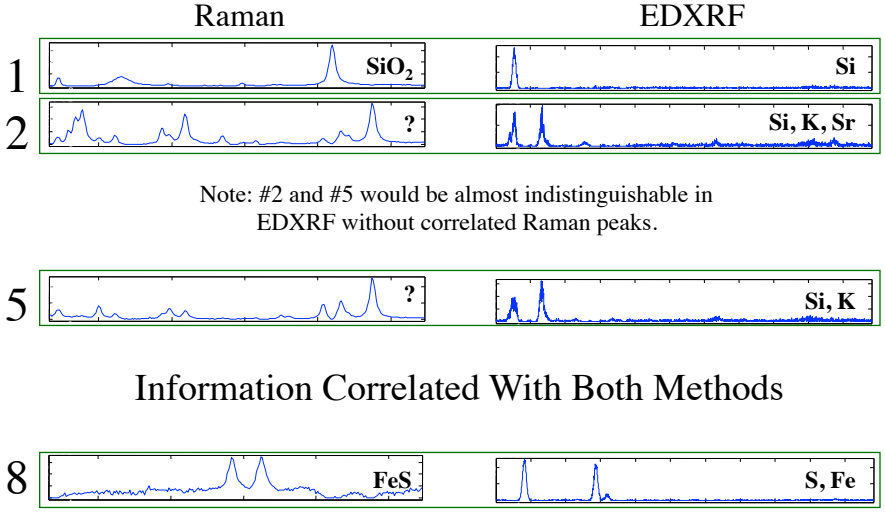
MCR Spectra for Fused Data



Concentration Contrast

EIGENVECTOR RESEARCH INCORPORATED

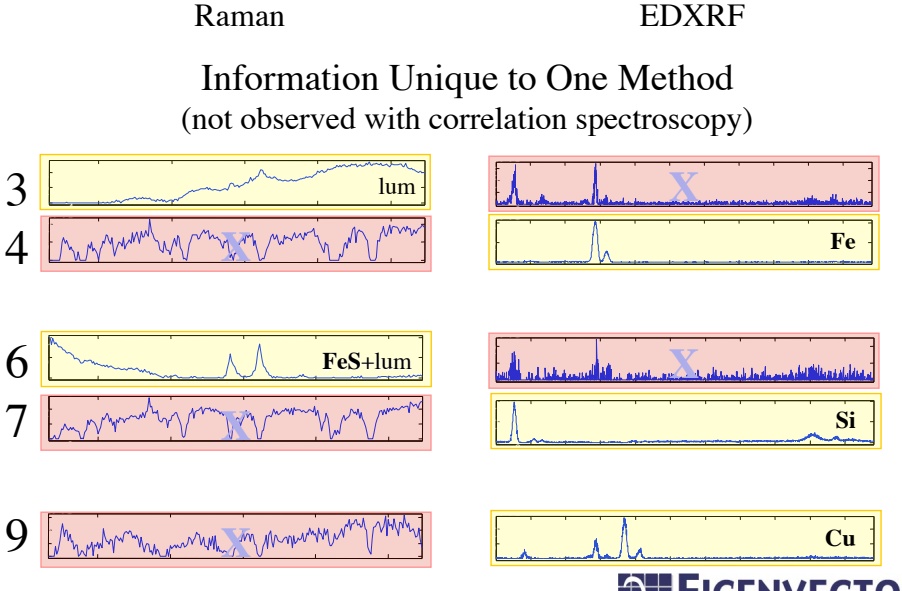
MCR Spectra for Fused Data



Concentration Contrast

EIGENVECTOR RESEARCH INCORPORATED

MCR Spectra for Fused Data



Concentration Contrast

EIGENVECTOR RESEARCH INCORPORATED

Conclusions I

- ASCA
  - for multi-set data typically from designed experiments
- MLASCA
  - for multi-level data typically from happenstance data (often semi-batch)
- ASCA and MLASCA allow new ways to partition and understand variance

EIGENVECTOR RESEARCH INCORPORATED

Conclusions II

- Data Fusion methods combine multi-block data that share a common mode
- Data Fusion can be done at three levels
  - Low Level: joining blocks after preprocessing
  - Mid Level: joining model outputs such as scores
  - High Level: Combine predictions from multiple models in some sort of voting scheme
- Alignment often important
- Often brings out aspects of data that aren't obvious in blocks analyzed separately

EIGENVECTOR RESEARCH INCORPORATED