

ASCA and MLSCA for Process and Instrument Characterization

ASCA = ANOVA Simultaneous Components Analysis

MLSCA = Multi-Level Simultaneous Components Analysis

©Copyright 2018

Eigenvector Research, Inc.

No part of this material may be
photocopied or reproduced in any form
without prior written consent from
Eigenvector Research, Inc.

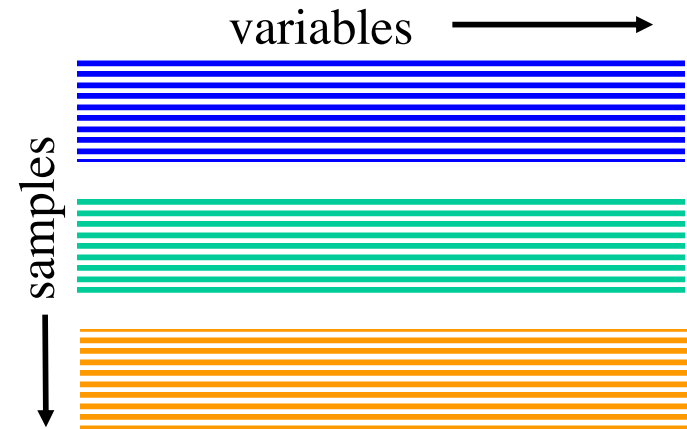


Definitions

- 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

Multi-set Data

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



Differences between groups may hide variability inherent to all samples.

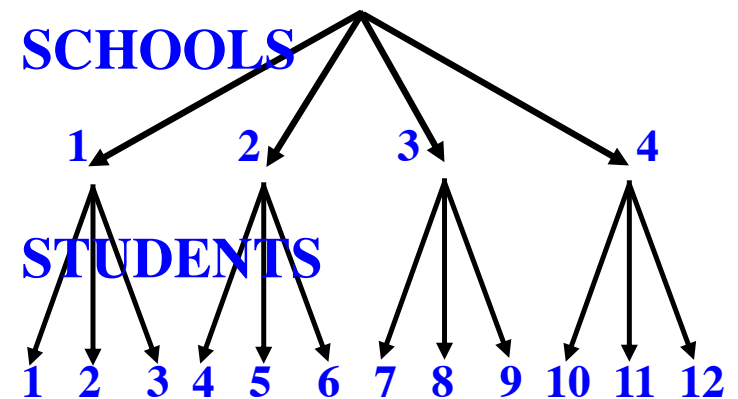
For samples grouped according to an experimental design we can separate variability due to each design factor, and systematic variability independent of the factors. This is the purpose of ASCA and MLSCA

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

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.

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.

ASCA Method

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

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

- Build PCA model for each main effect and interaction

$$X = X_{avg} + T_A P'_A + T_B P'_B + T_{AB} P'_{AB}$$

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

ASCA Example

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

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

ASCA Model

The screenshot shows the ASCA Model software interface. At the top, the window title is "Analysis - ASCA - X, F". The menu bar includes "File", "Edit", "Preprocess", "Analysis", "Refine", "Tools", "Help", and "FigBrowser". Below the menu bar is a toolbar with icons for plotting, data, and analysis. The main workspace displays the model equation: $X = 1m + \text{Time} + \text{Tre...} + (\text{Ti...} + E$. Below the equation is a flow diagram showing "Response" and "DOE" blocks connected to a "Model" block via "P" (Process) nodes. A "Clutter" button is also present. Below the flow diagram is a "View:" section with tabs for "SSQ Table" and "ASCA Settings". The "SSQ Table" is active, displaying the following table:

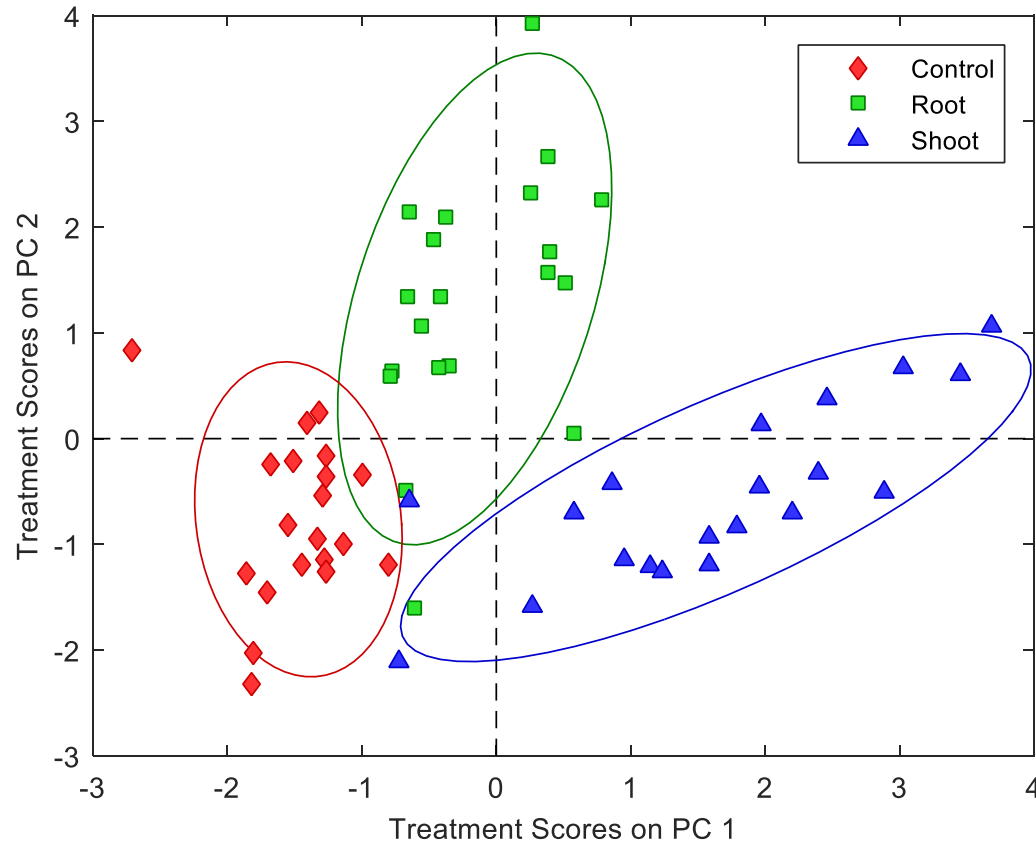
	Term	PCs	Cum Eigen Val	Effect	P-value
1	Time	3	1.52	13.80	0.0100
2	Treatment	2	2.54	23.10	0.0100
3	(Time) x (Treatment)	6	1.49	13.58	0.0100
4	Mean	-	-	0.00	-
5	Residuals	-	-	49.52	-

Below the table is a yellow note box with the following text:

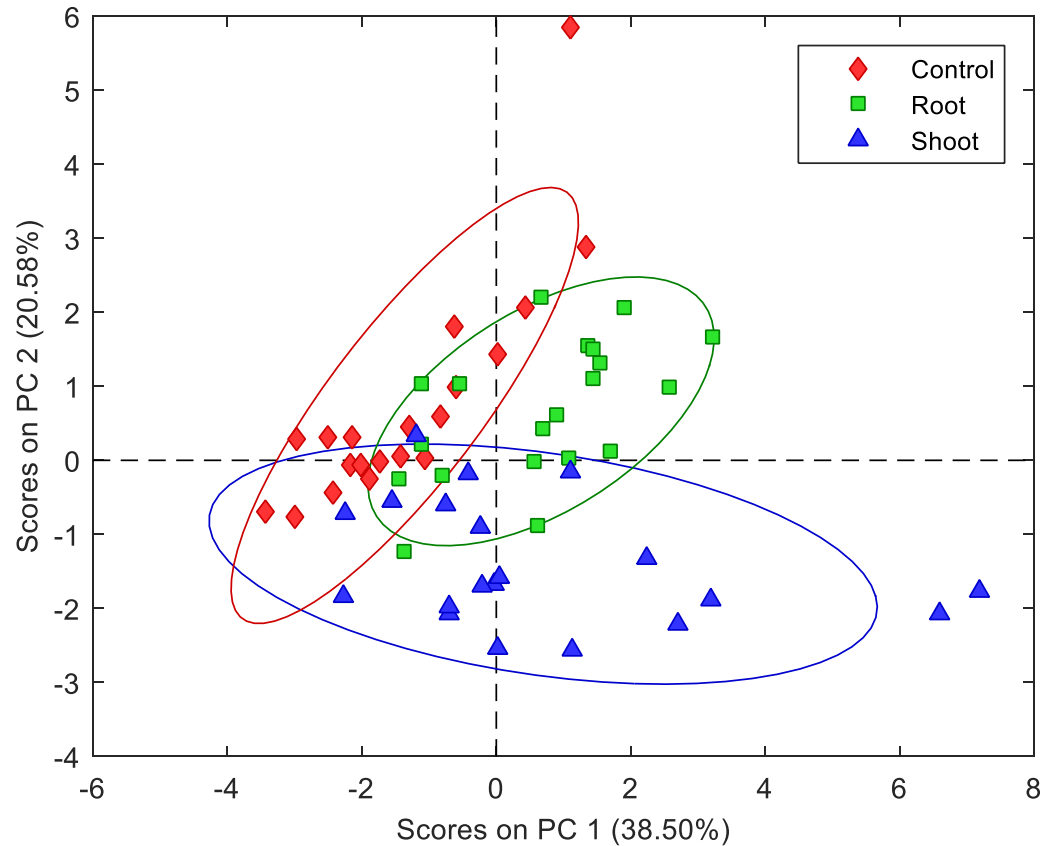
[1 of 1] Note: The x-block appears to be mean centered. This is OK but will cause the "mean" in the effects table to be zero.

A model has been calibrated from the data. Review the model using the toolbar button(s), save the model (File menu), or load test (validation) data (File menu). The number of components, preprocessing

ASCA Treatment Scores Plot

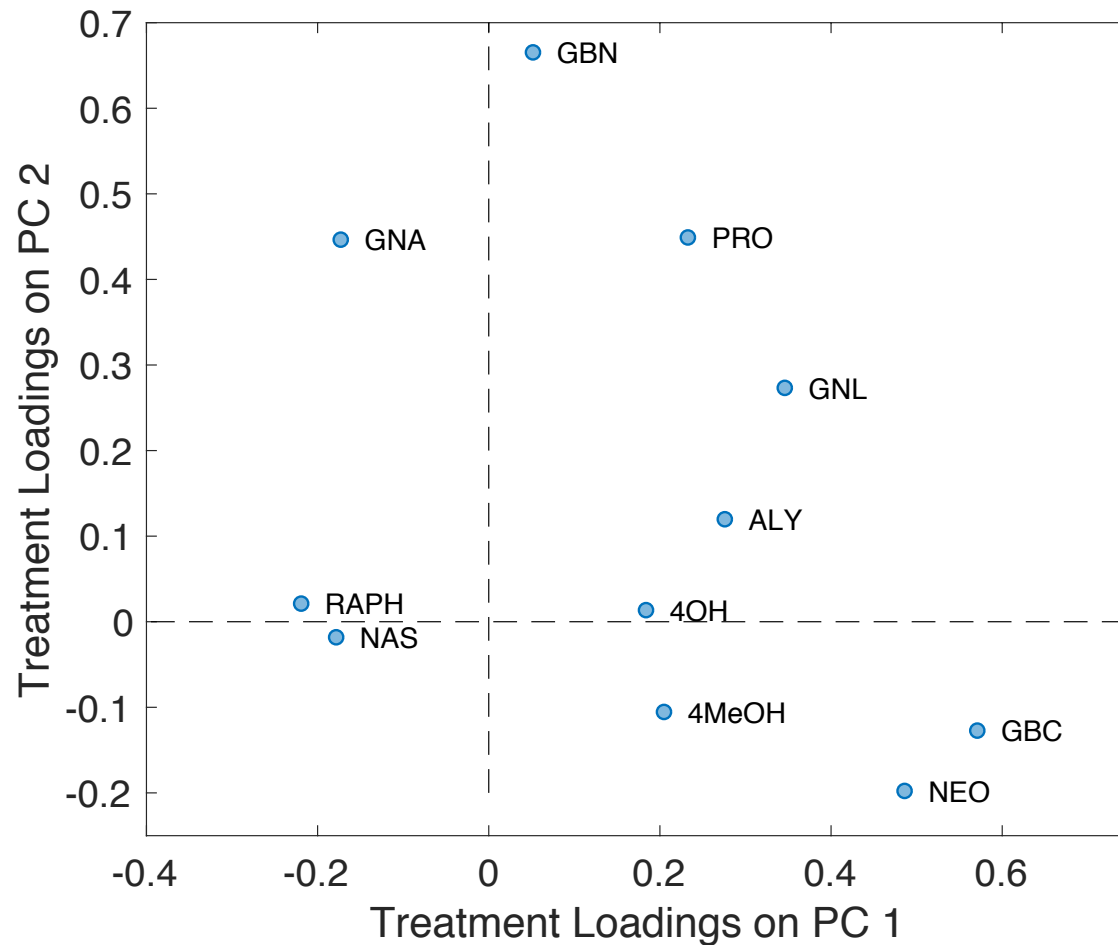


PCA Scores Plot

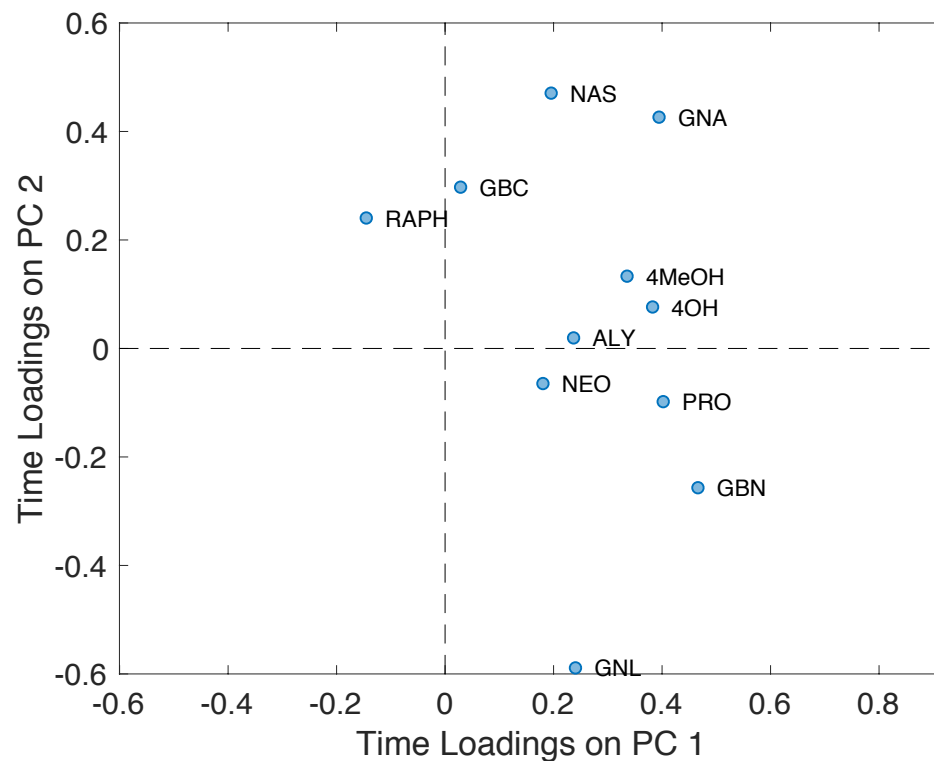
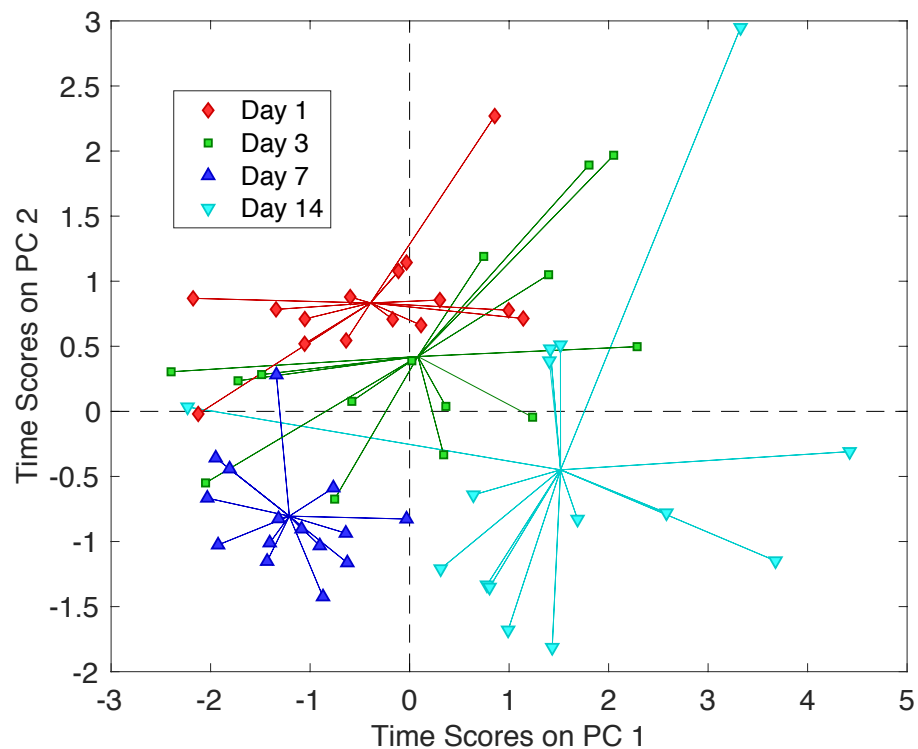


...better than is seen by simply applying PCA to the data.

Treatment Loadings Plot



Time Scores and Loadings



ASCA Conclusions

ASCA resolves variation associated with each factor, and shows main variables involved.

- Time factor scores reveal the common response
- Treatment factor scores show the Treatment effect independent of Time.
- Time x Treatment interaction scores show the additional time dependency at each Treatment level.

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.

MLSCA

Multi-level Simultaneous Component Analysis

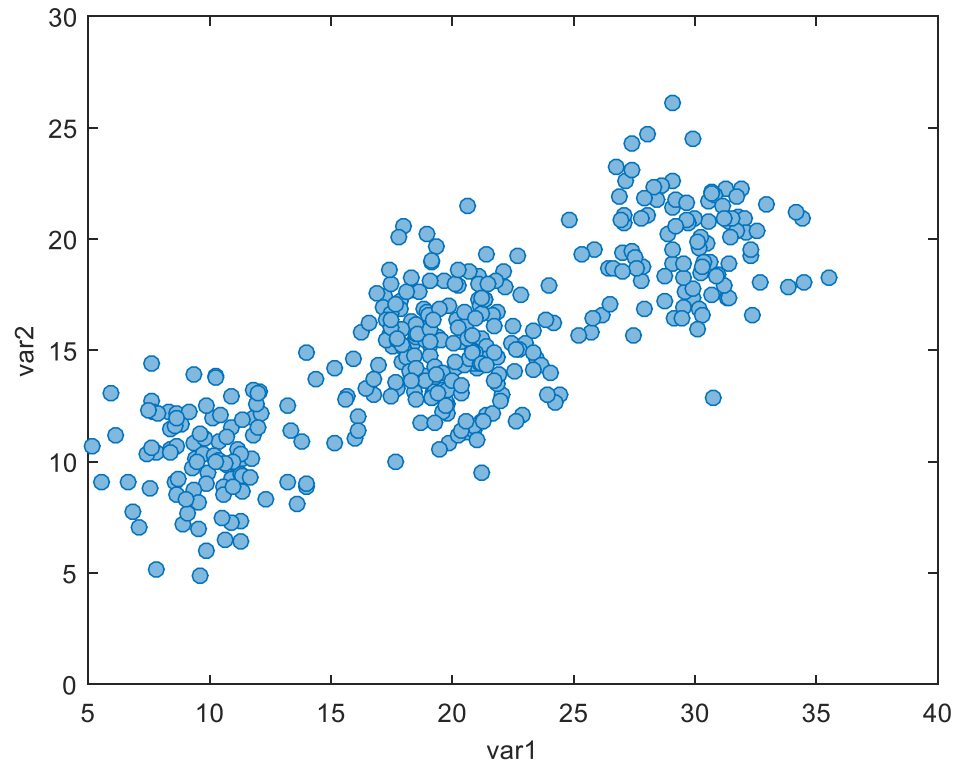
MLSCA is a special case of ASCA applied to data from designed experiments with nested factors.

- 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.

MLSCA: simple example

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

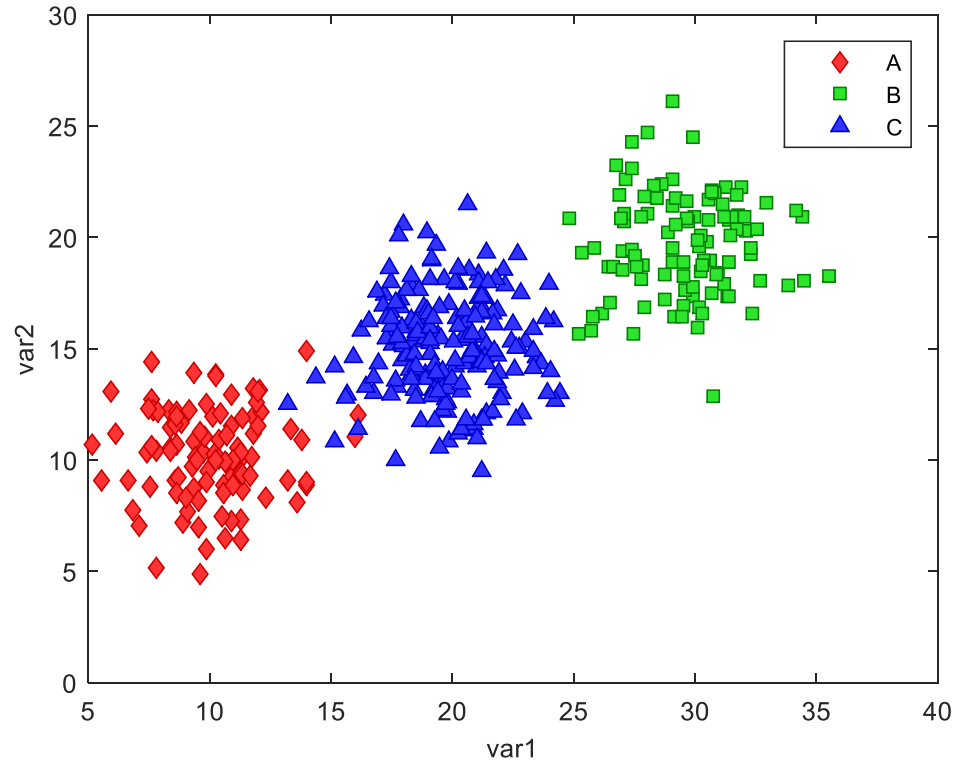
Example: $X: (400,2)$
400 samples from 3
individuals.

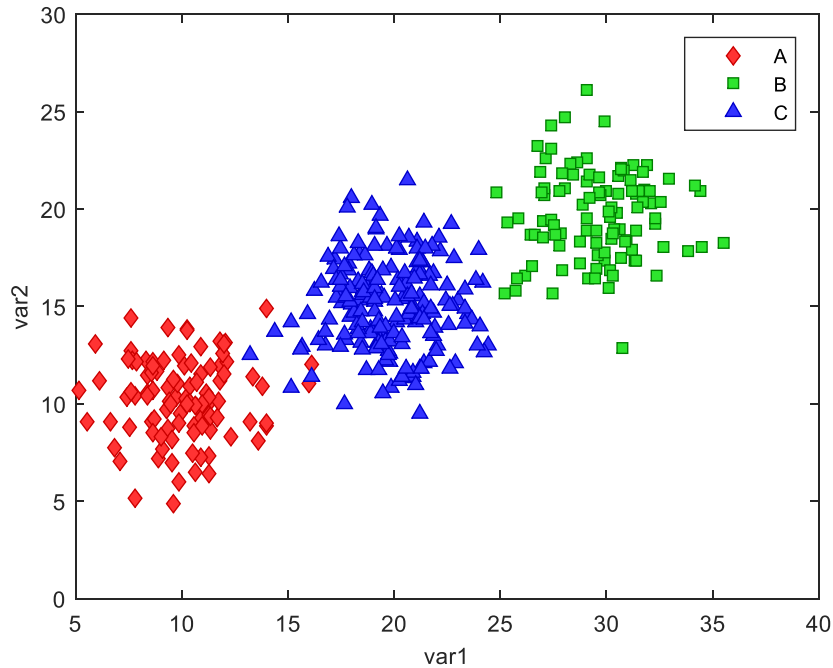


MLSCA: simple example

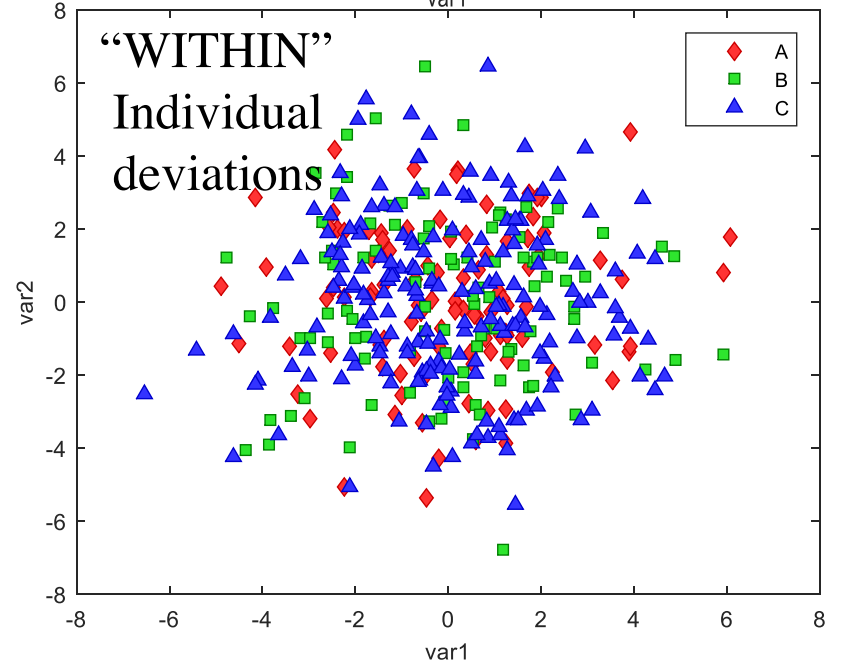
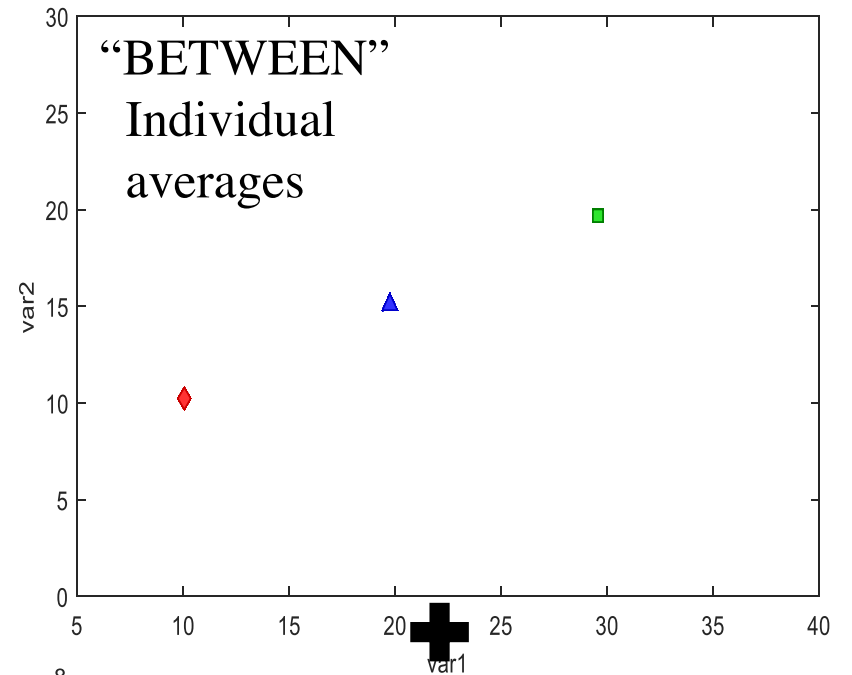
Example: $X: (400, 2)$
400 samples from 3
individuals, A, B, and C.

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





==



$X = \text{average for each individual} + \text{deviations from that}$

Nested Wafer Processing 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
<i>80 REPLI- CATES</i>	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.

MLSCA Method

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

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

X_A : contains factor A level averages,

$X_{B(A)}$: contains factor B level averages for each level A,

E : are the residuals, “within” component.

- Build PCA model for each effect and residual

$$X = X_{avg} + T_A P'_A + T_{B(A)} P'_{B(A)} + T_E P'_E$$

“constant” “between A” “between B” “within”

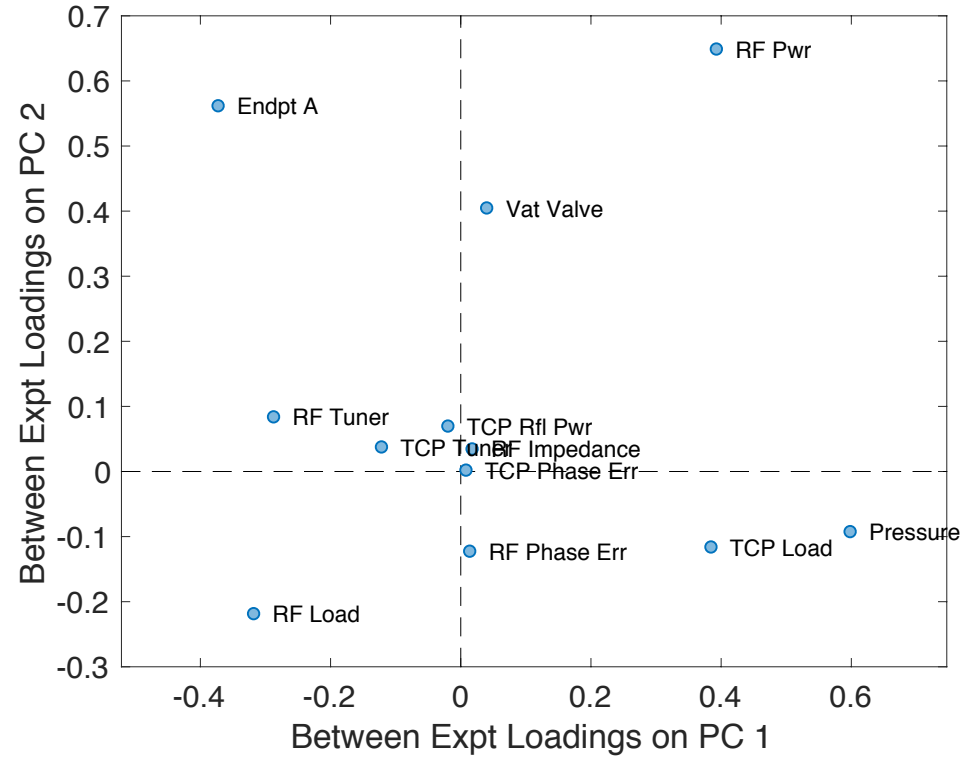
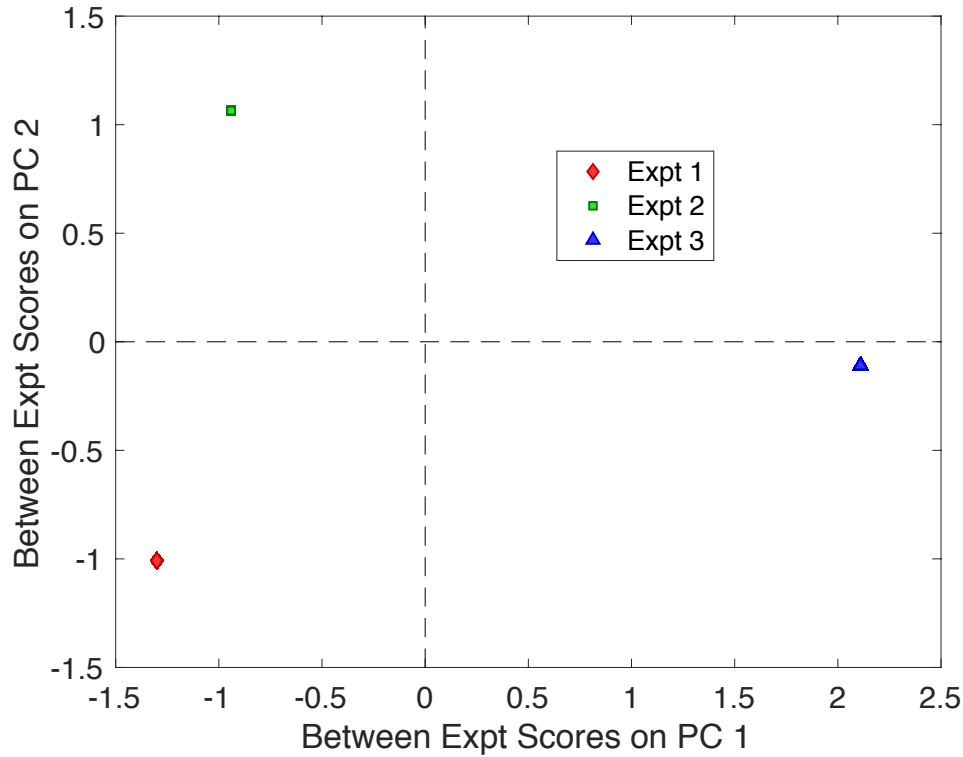
MLSCA Model of Wafer Data

The screenshot shows the MLSCA software interface. At the top, the window title is "Analysis - MLSCA - x, F". The menu bar includes "File", "Edit", "Preprocess", "Analysis", "Refine", "Tools", "Help", and "FigBrowser". Below the menu bar is a toolbar with various icons. The main workspace contains a model equation: $X = 1m + \text{Expt} + \text{Wafer} + E$. Below the equation is a flow diagram with "Response" and "DOE" boxes on the left, two "P" nodes in the middle, and a "Model" box on the right. A "Clutter" button is above the "Model" box, and a "Calibrate" button is below it. Below the flow diagram are "View:" tabs for "SSQ Table" and "MLSCA Settings". Below the tabs is a "Number PCs:" field with an "Auto Select" button. Below that is a "Model Statistics (* = suggested)" section with a table:

	Term	PCs	Cum Eigen Val	Effect
1	Expt	2	3.08	25.69
2	Wafer	4	0.29	2.41
3	Within	2	8.63	71.89
4	Mean	-	-	0.00

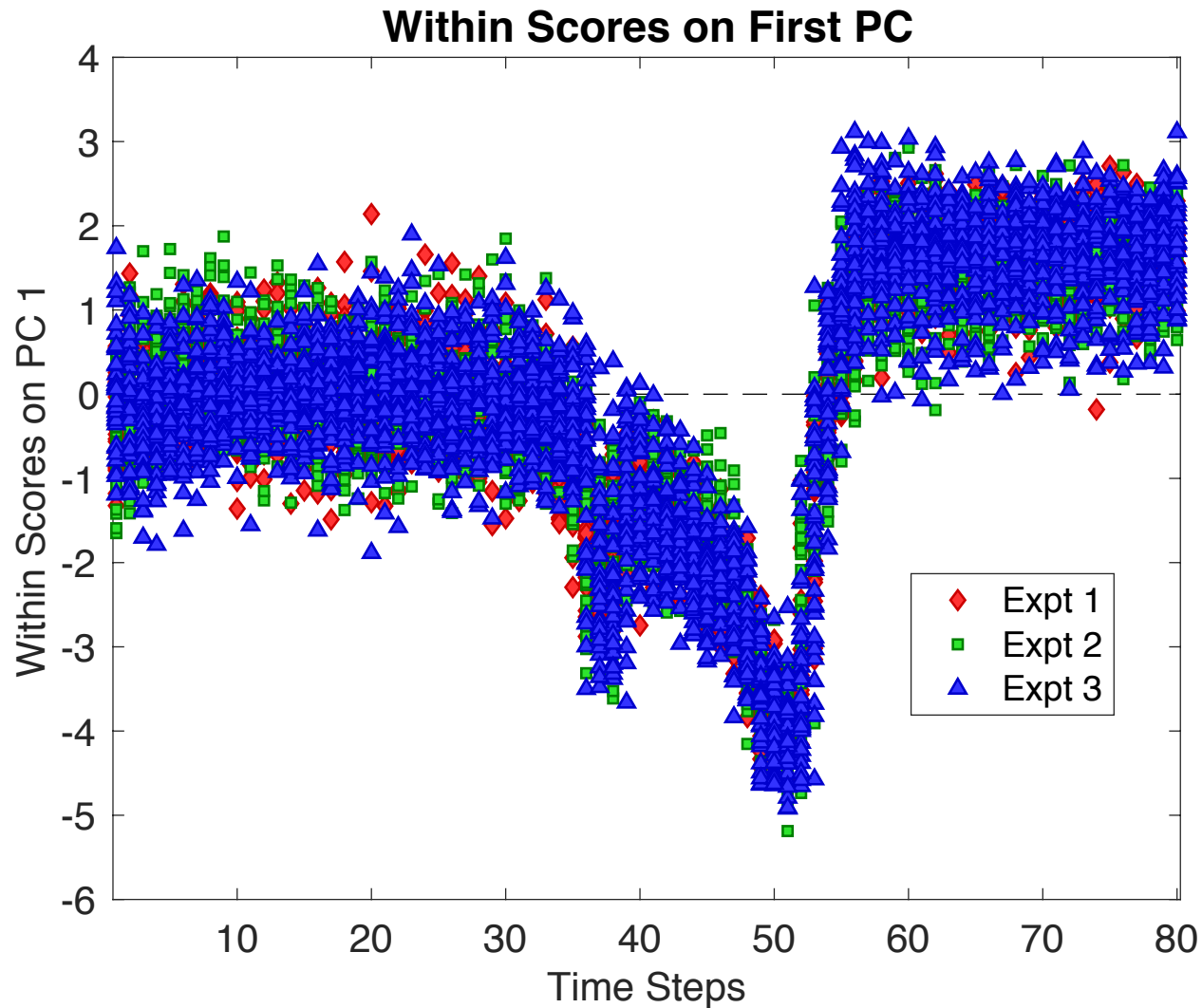
At the bottom of the interface, a text box contains the following message: "A model has been calibrated from the data. Review the model using the toolbar button(s), save the model (File menu), or load test (validation) data (File menu). The number of components, preprocessing options, and other settings can also be modified to adjust the model. The data can be viewed and edited from the Edit menu."

Experiment Scores & Loadings



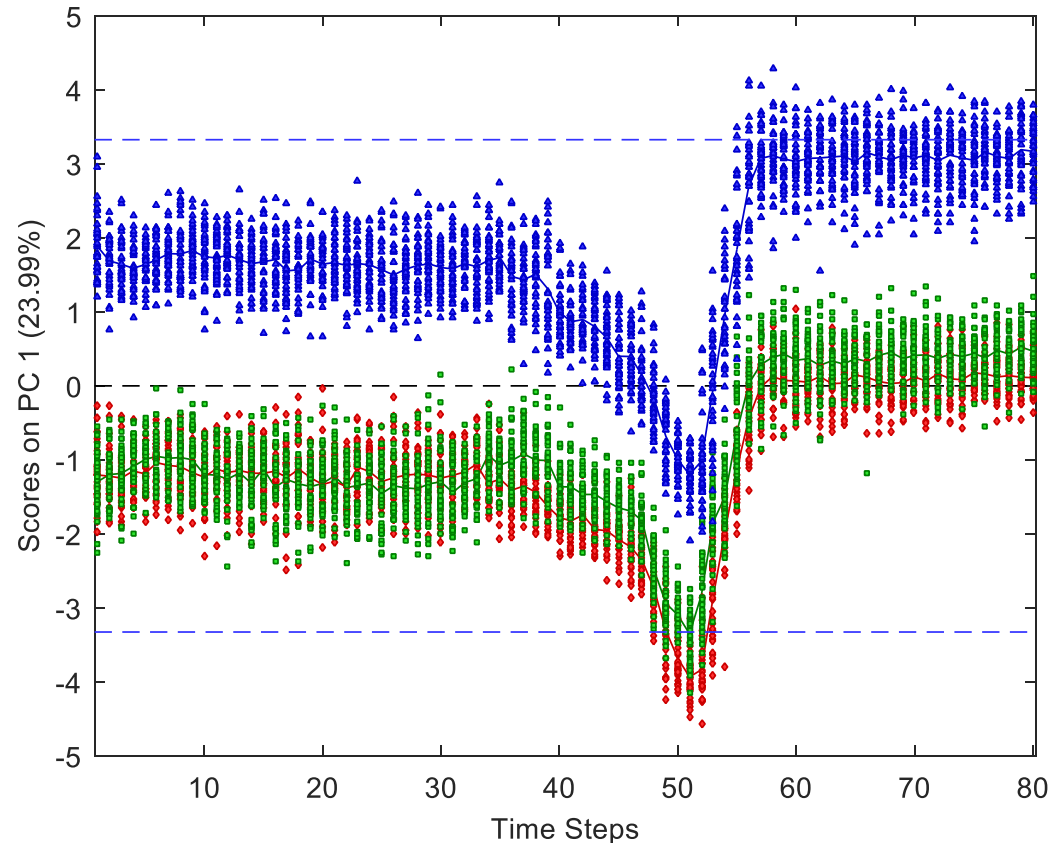
MLSCA Scores Plot

“Within” Residual sub-model, PC 1 vs. time



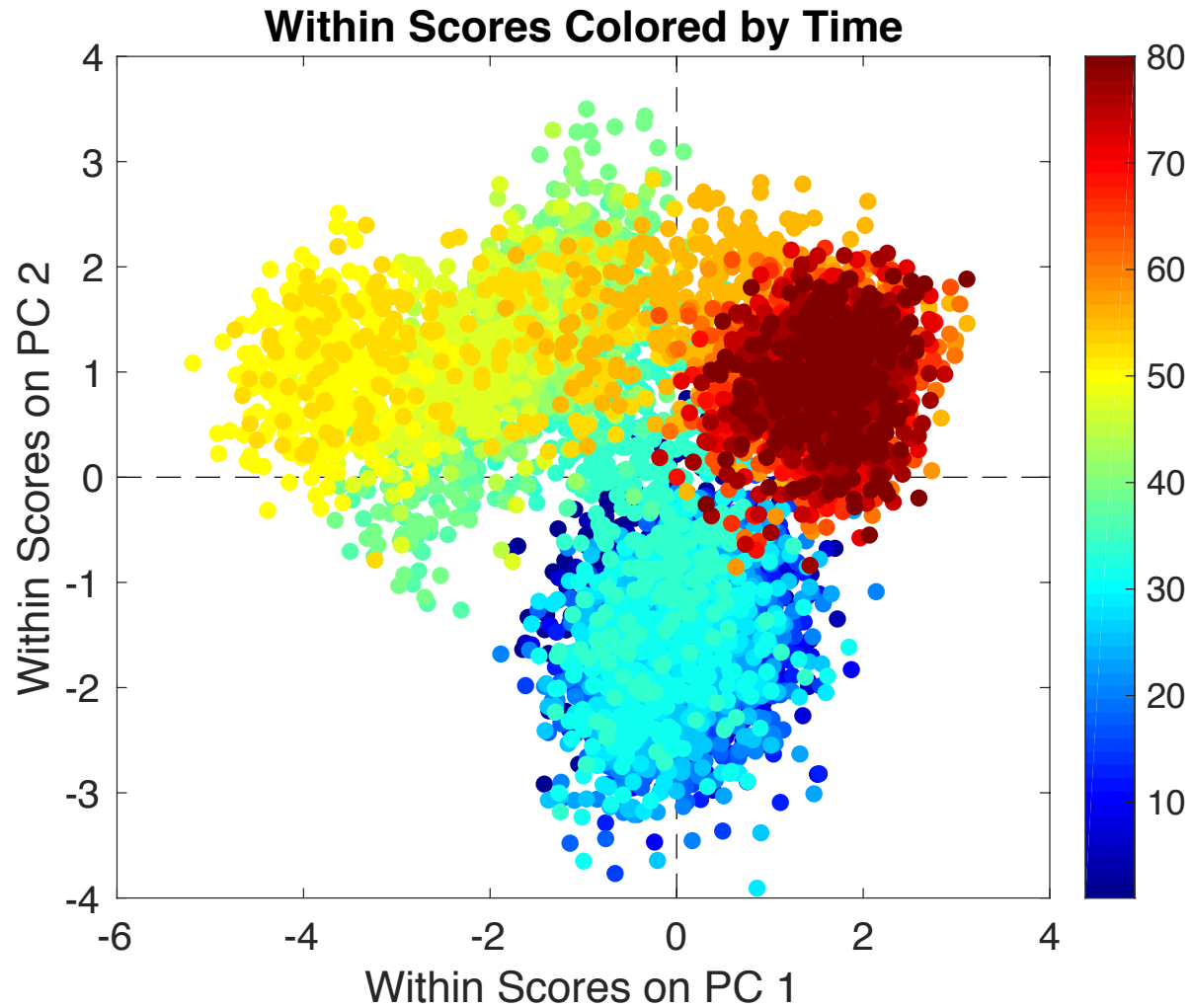
PCA Scores Plot

PC 1 vs. time, Colored by Experiment class



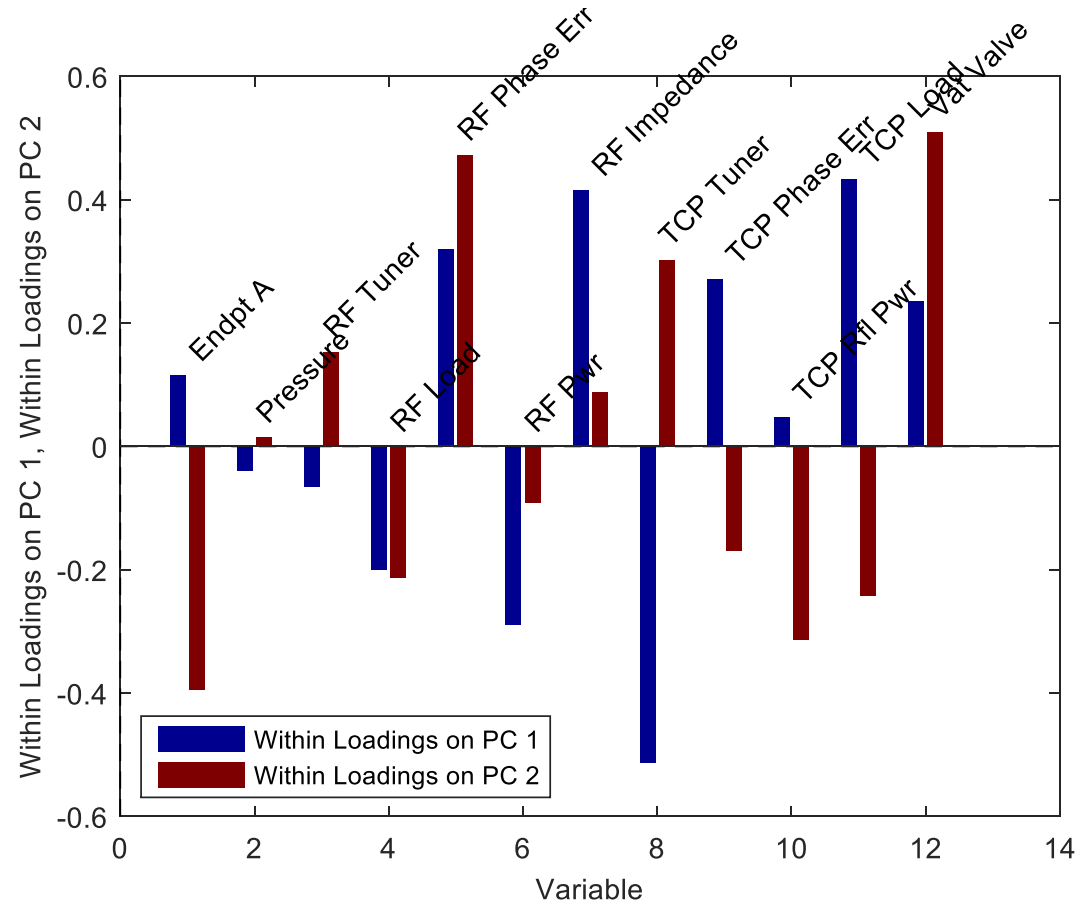
The spike at time step 47-48 is not seen in PC 1.
It shows up in PC 2 because the offset between experiments dominates PC1 in simple PCA.

MLSCA “Within” Scores Plot



MLSCA Loadings Plot

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



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.

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.

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.