ASCA and MLSCA for Process and Instrument Characterization

ASCA = ANOVA Simultaneous Components Analysis
MLSCA = Multi-Level Simultaneous Components Analysis
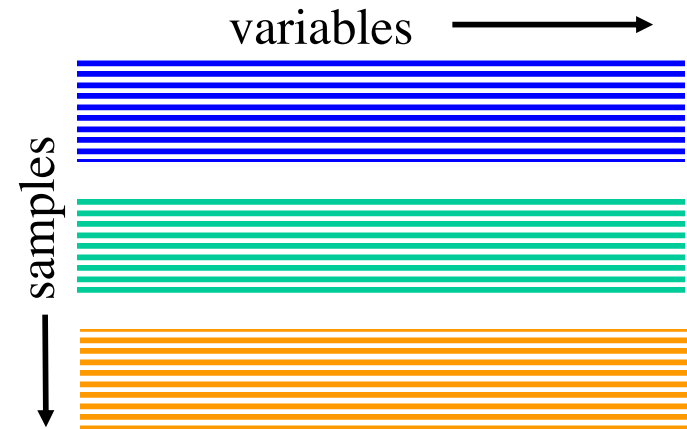
EIGENVECTOR RESEARCH INCORPORATED

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

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

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.
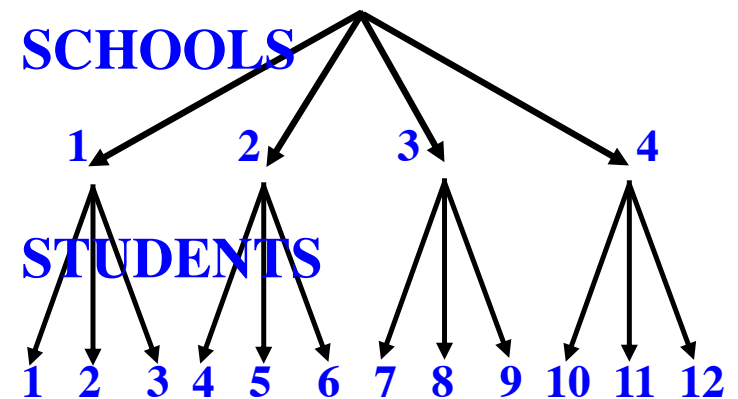
These are both 2-factor designs

EIGENVECTOR
RESEARCH INCORPORATED

4

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

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

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

- 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.

EIGENVECTOR
RESEARCH INCORPORATED

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				

EIGENVECTOR
RESEARCH INCORPORATED

ASCA Model

ASCA *Treatment Scores Plot*

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

EIGENVECTOR RESEARCH INCORPORATED
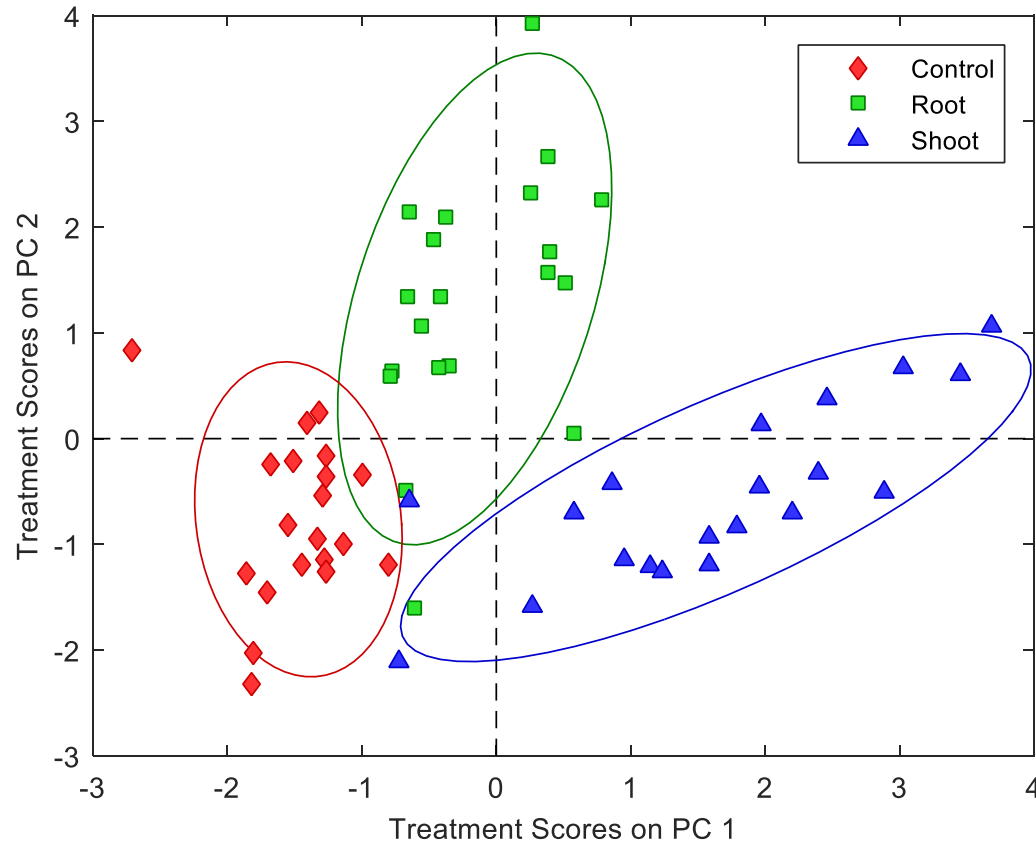
PCA Scores Plot
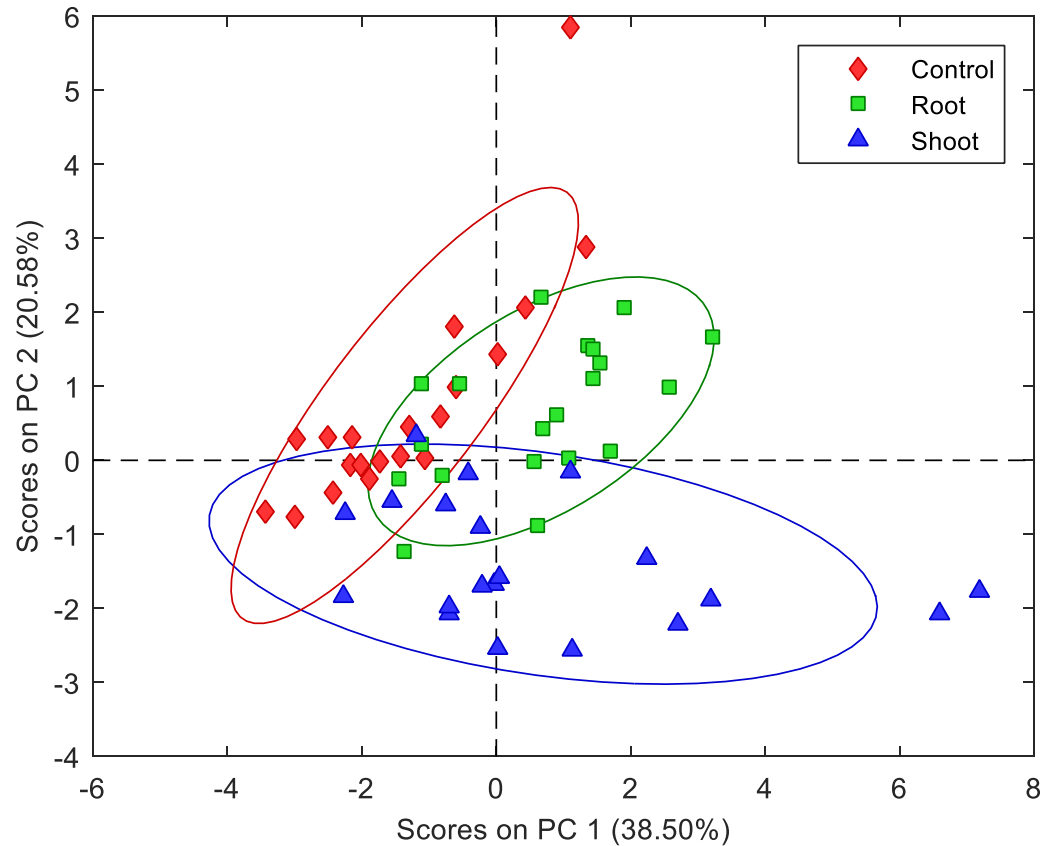
…better than is seen by simply applying PCA to the data.

EIGENVECTOR RESEARCH INCORPORATED

Treatment Loadings Plot

EIGENVECTOR
RESEARCH INCORPORATED
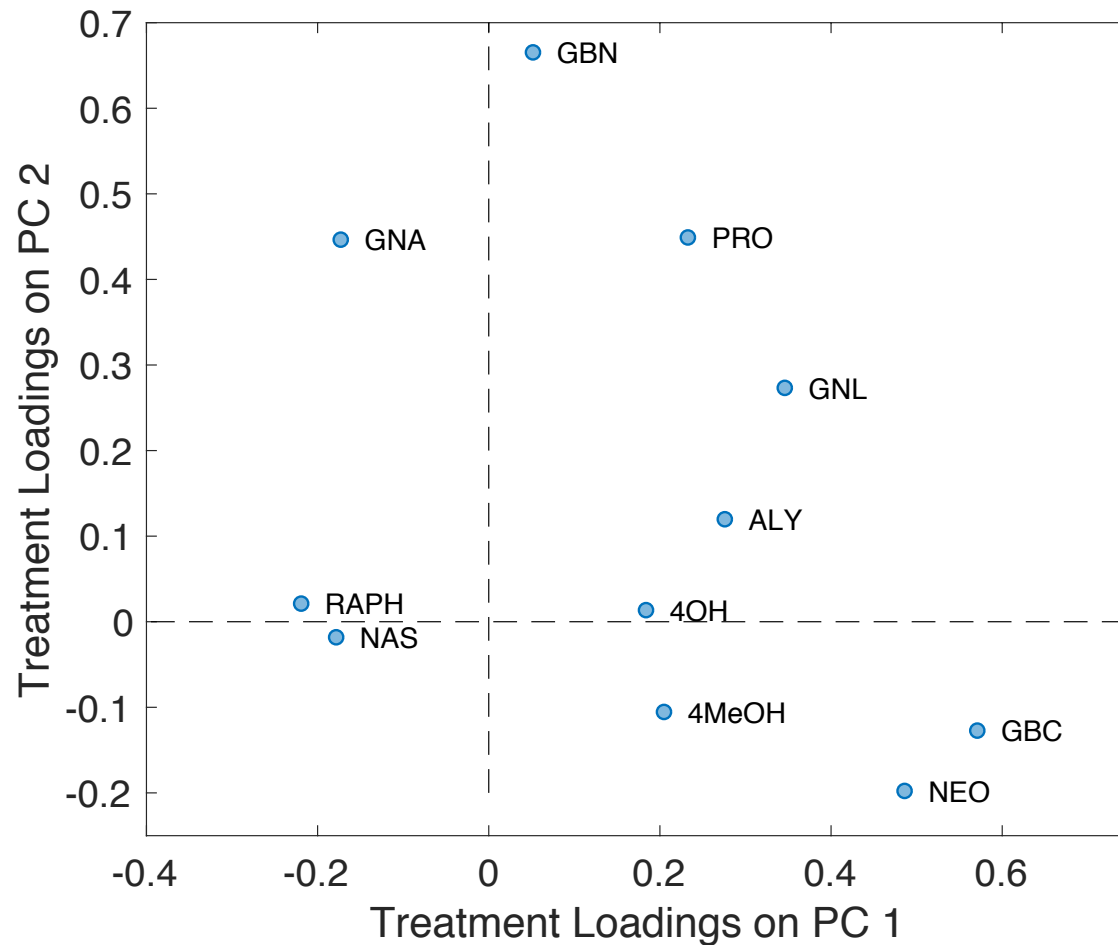
Time Scores and Loadings
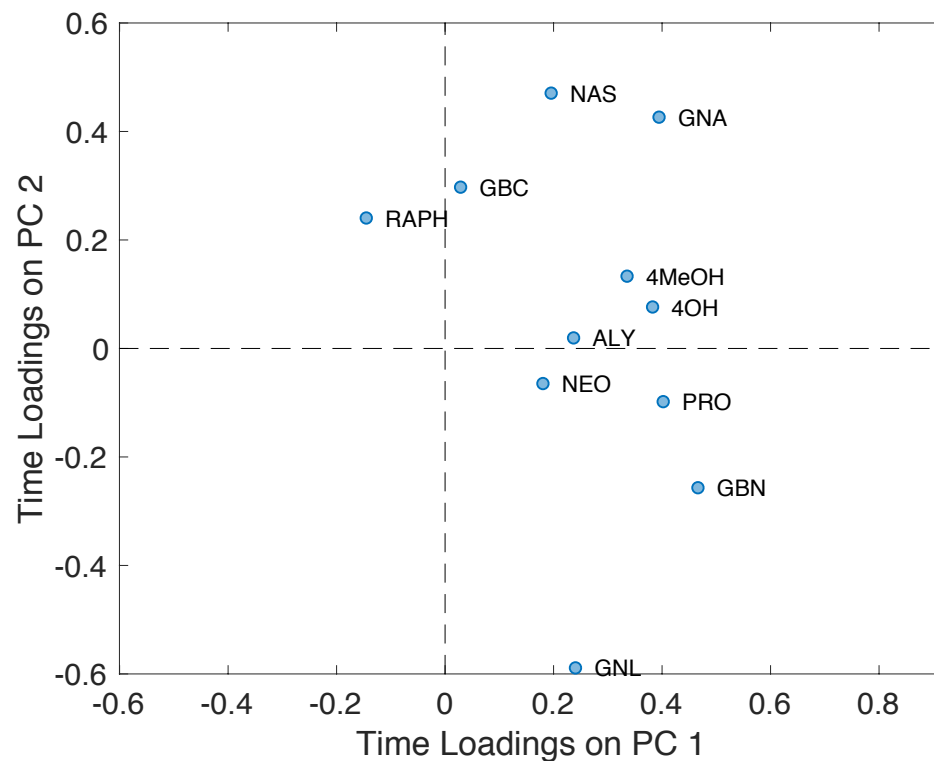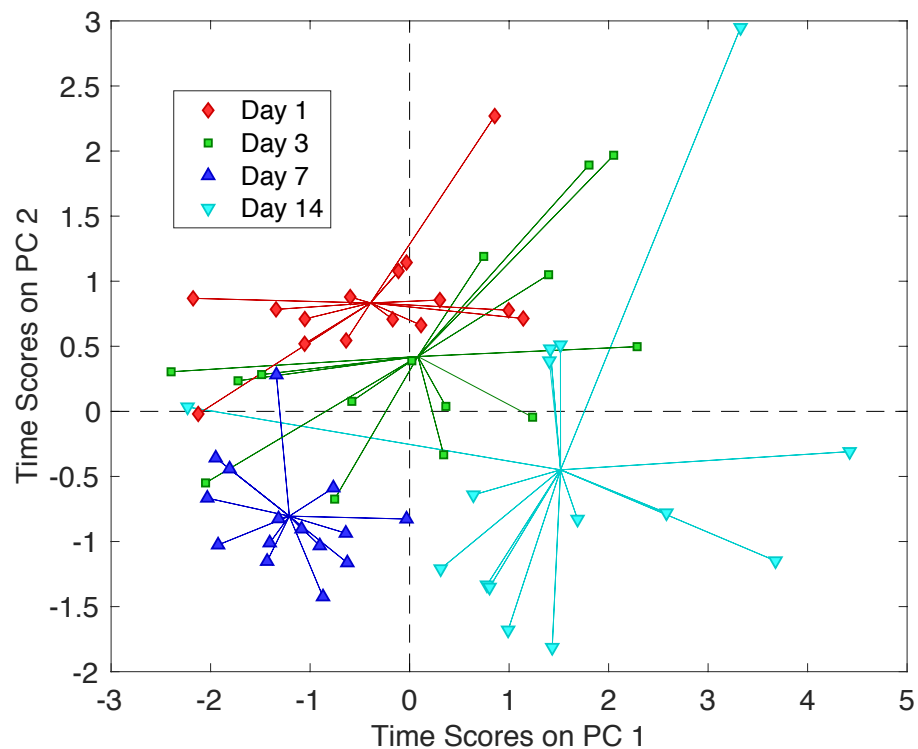
ASCA Conclusions

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.

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

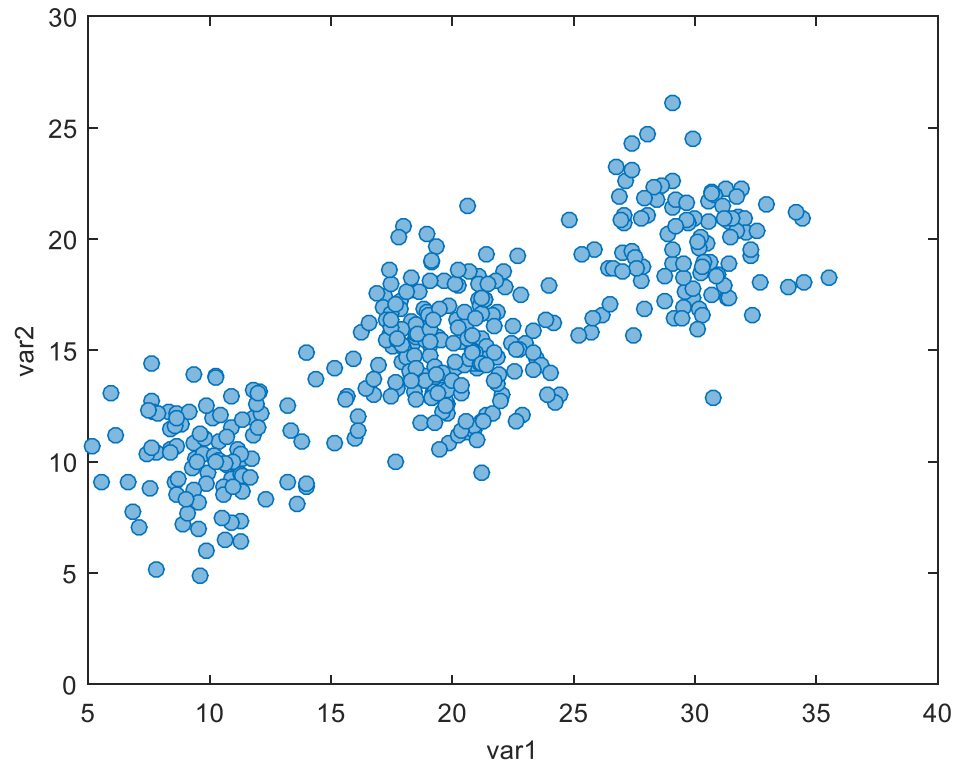# MLSCA
Multi-level Simultaneous Component Analysis

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.

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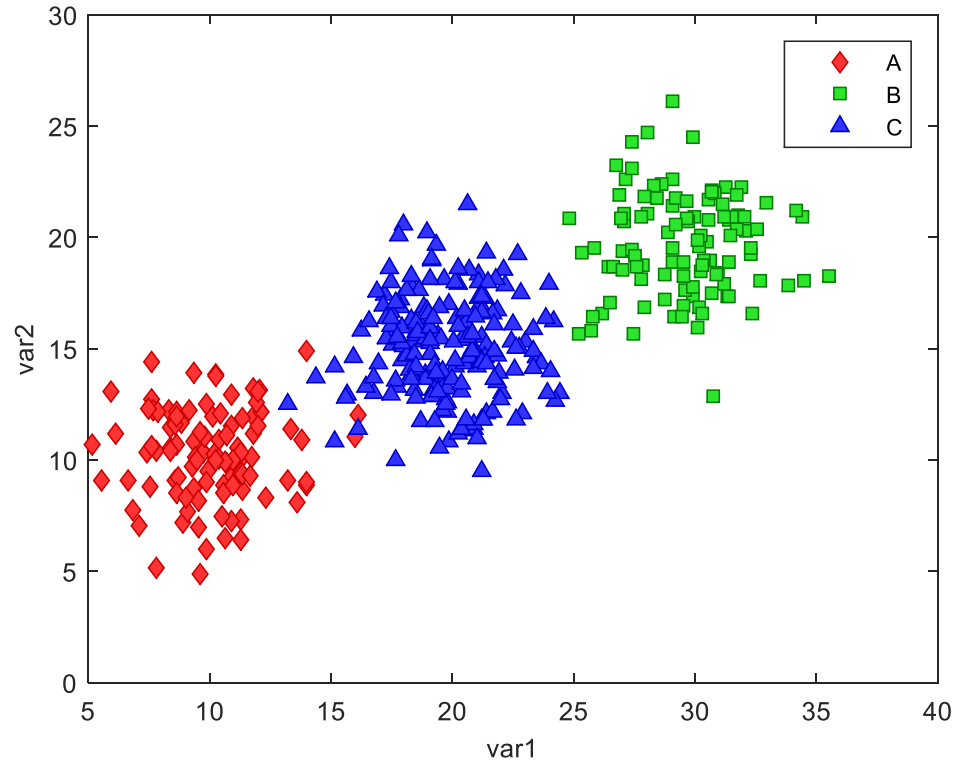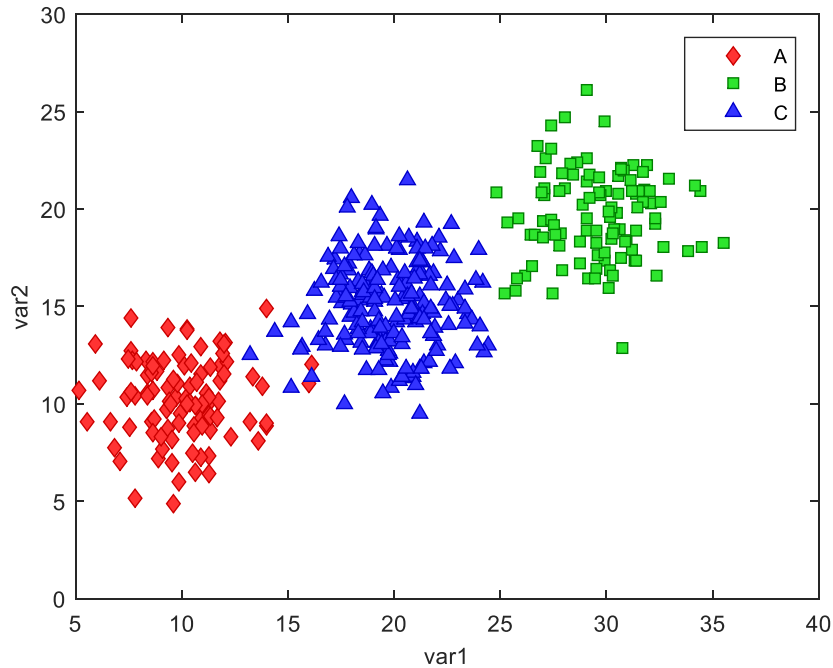
Example: X: (400,2) 400 samples from 3 individuals.

EIGENVECTOR
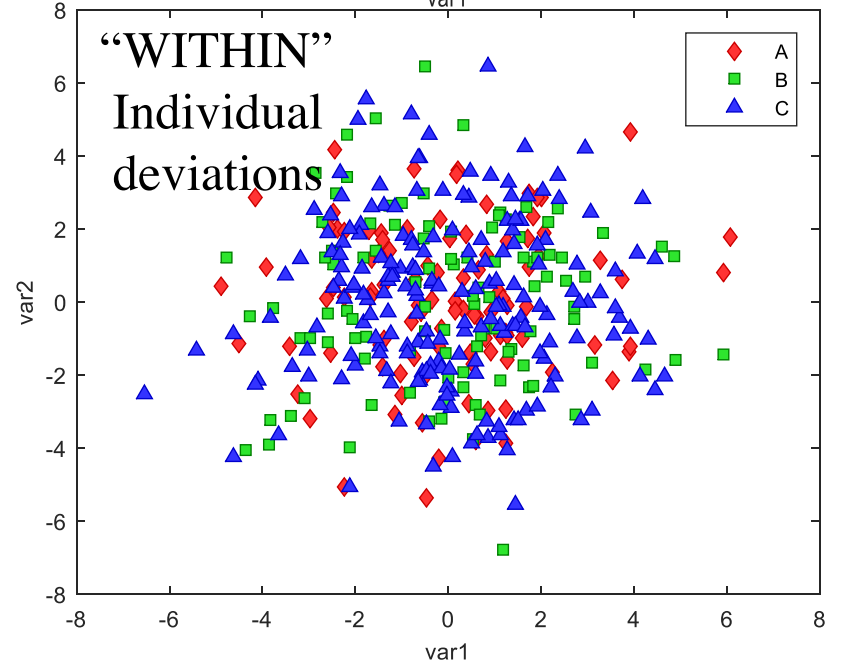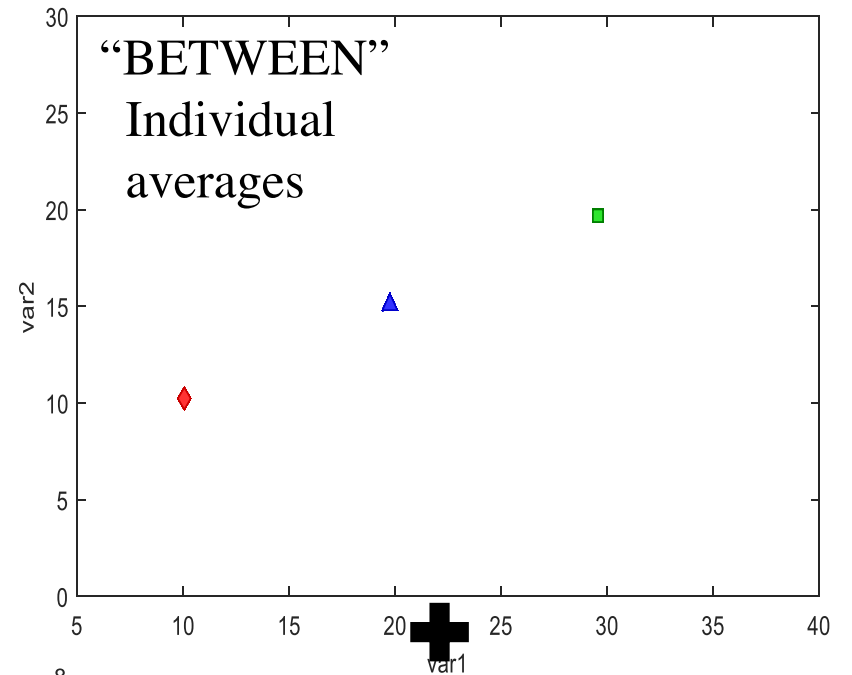**RESEARCH INCORPORATED**

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.

EIGENVECTOR
RESEARCH INCORPORATED

X = average for each individual
+ 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
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.

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

$$X = X_{avg} + T_A P'_A + T_{B(A)} P'_{B(A)} + T_E P'_E$$
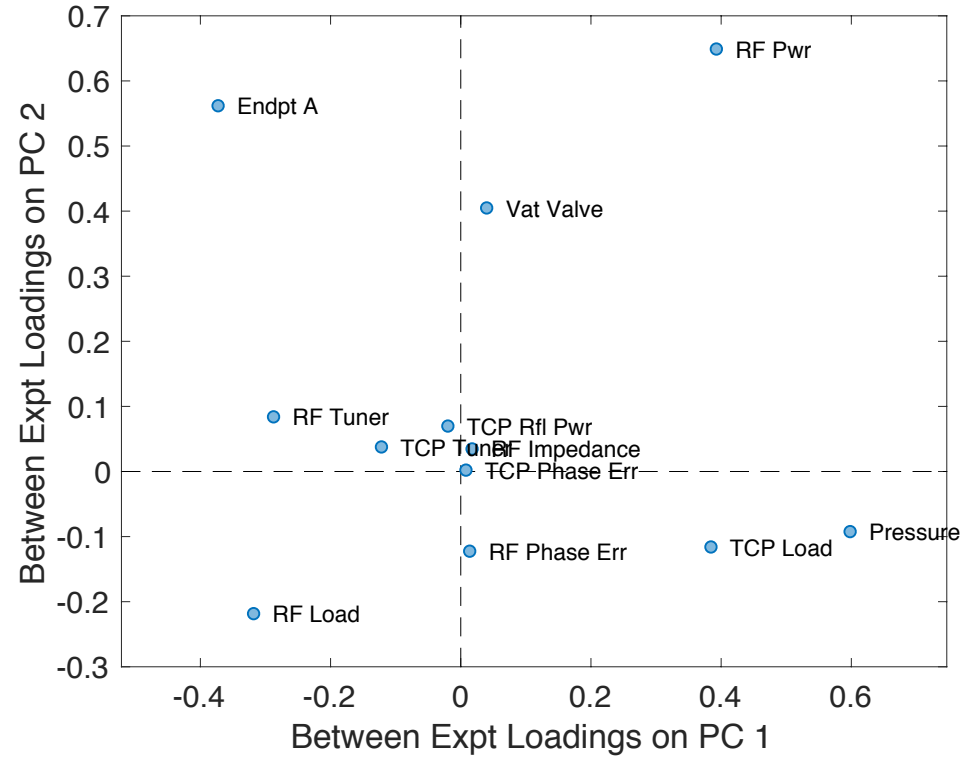
"constant"   *"between A"*   *"between B"*   *"within"*
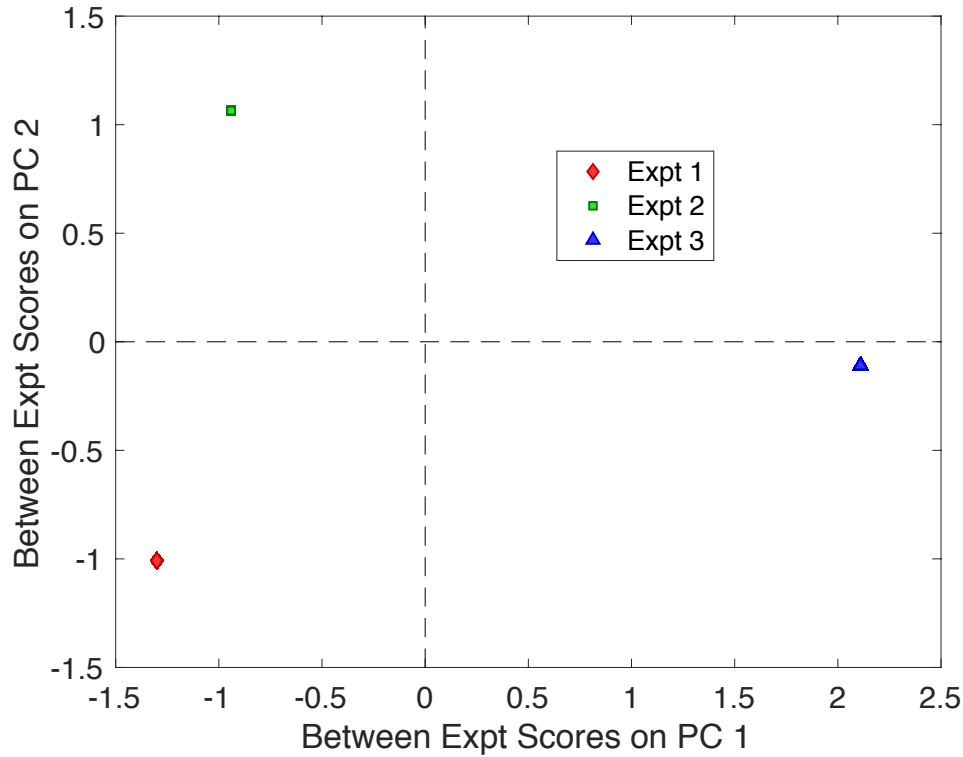
EIGENVECTOR
RESEARCH INCORPORATED

MLSCA Model of Wafer Data

Experiment Scores & Loadings

EIGENVECTOR RESEARCH INCORPORATED

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

Within Scores on First PC

24
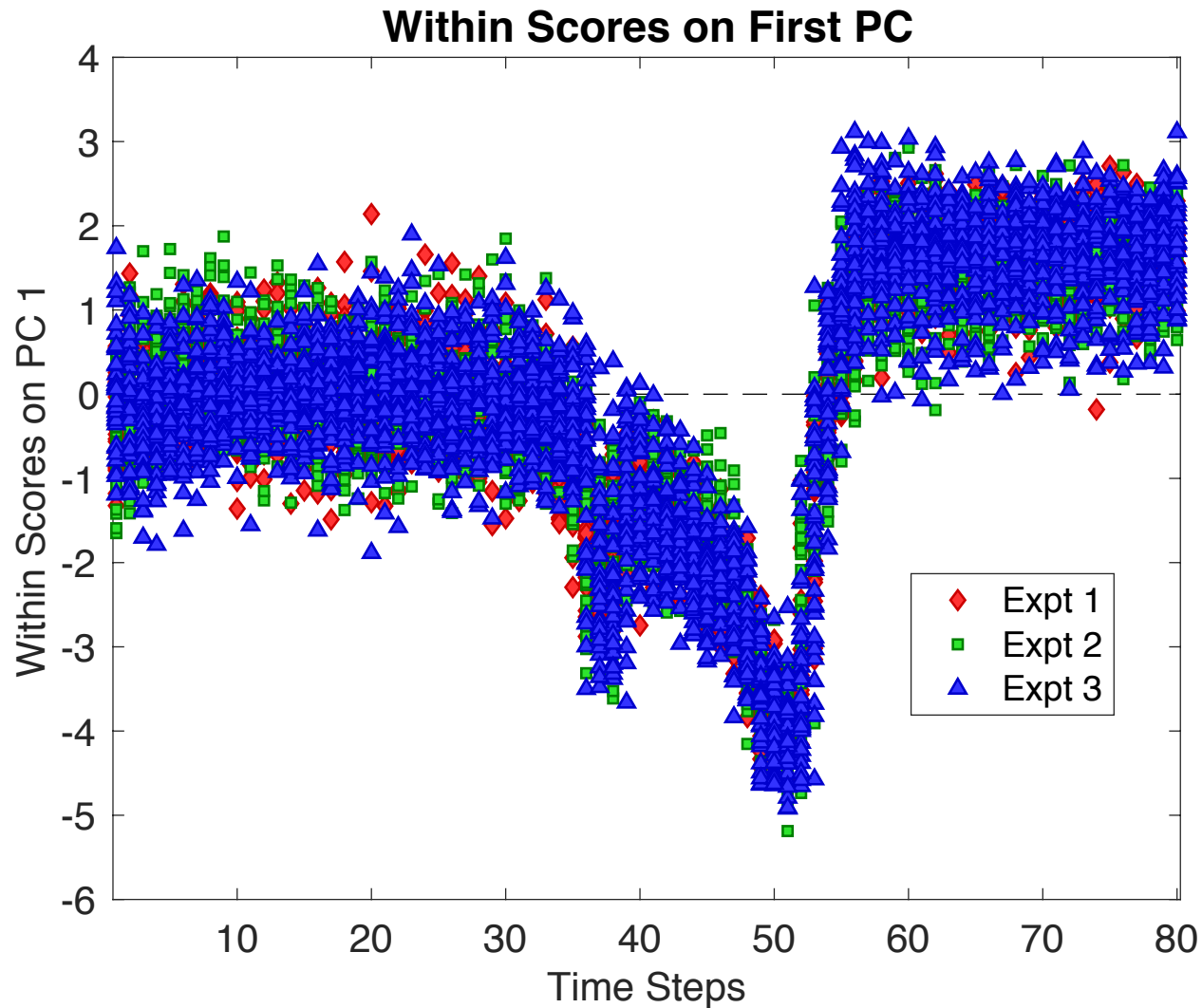
PCA Scores Plot
PC 1 vs. time, Colored by Experiment class
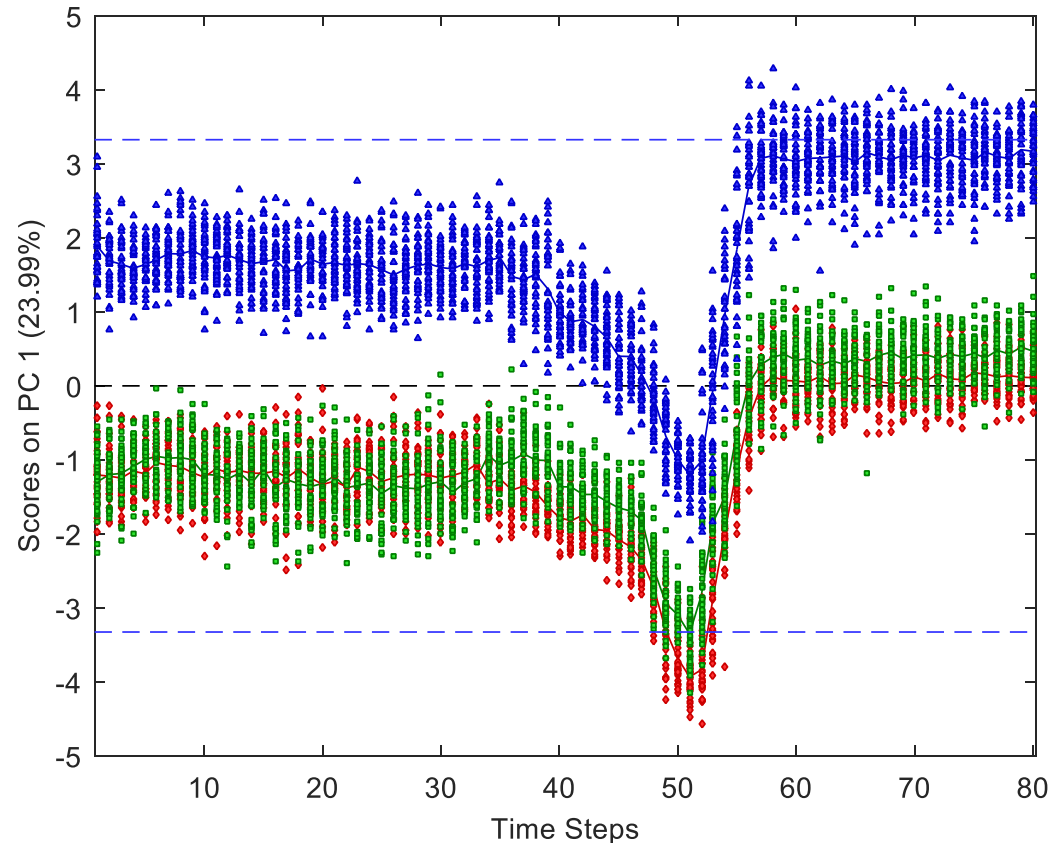
The spike at time step 47-48 is not seen in PC 1.
It shows up in PC 2 because the offset between
experiments dominates PC1 in simple PCA.

EIGENVECTOR
RESEARCH INCORPORATED

MLSCA "Within" Scores Plot

Within Scores Colored by Time

EIGENVECTOR
RESEARCH INCORPORATED

MLSCA Loadings Plot
"Within" Residual sub-model, PC 1 and 2

MLSCA Conclusions

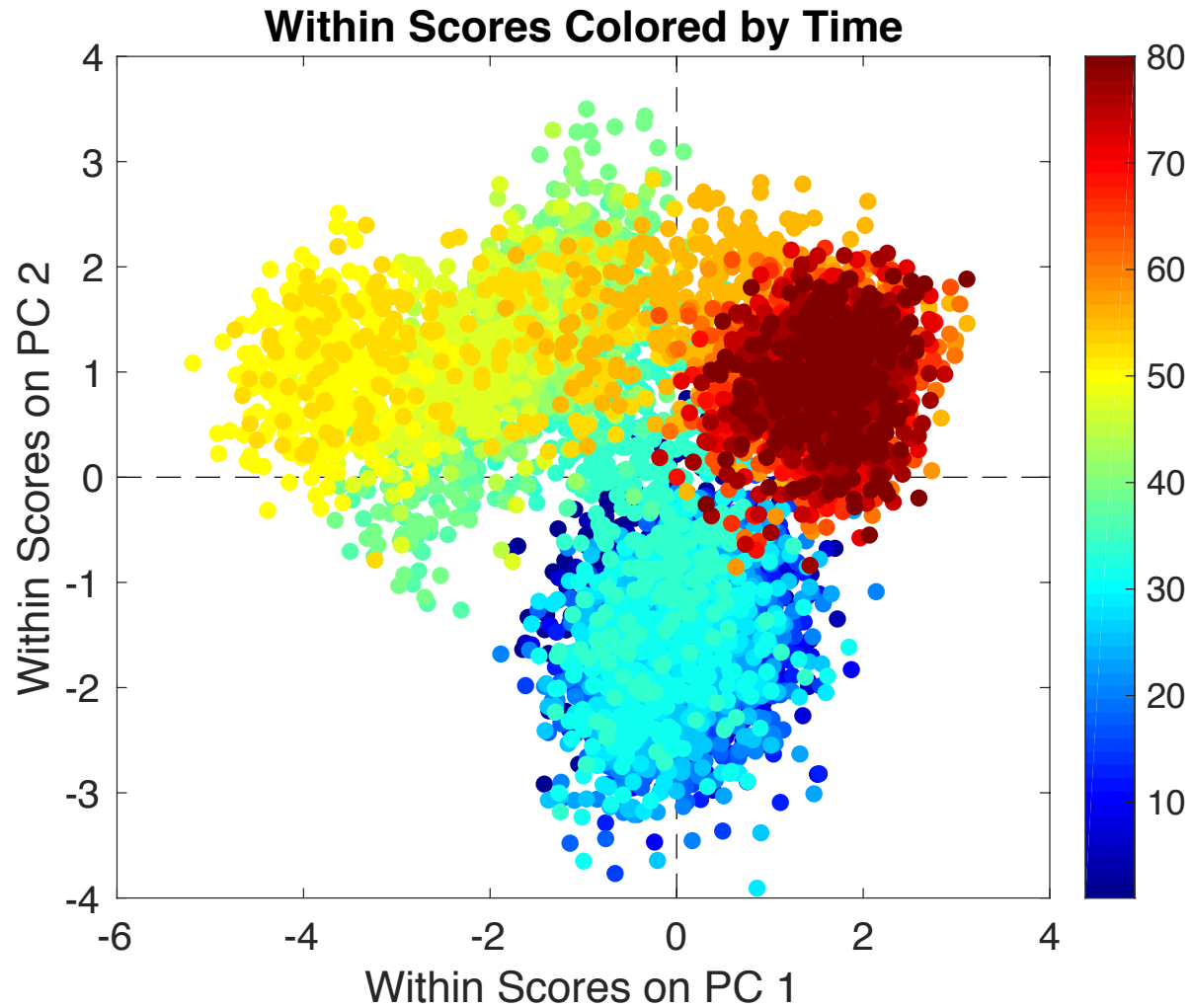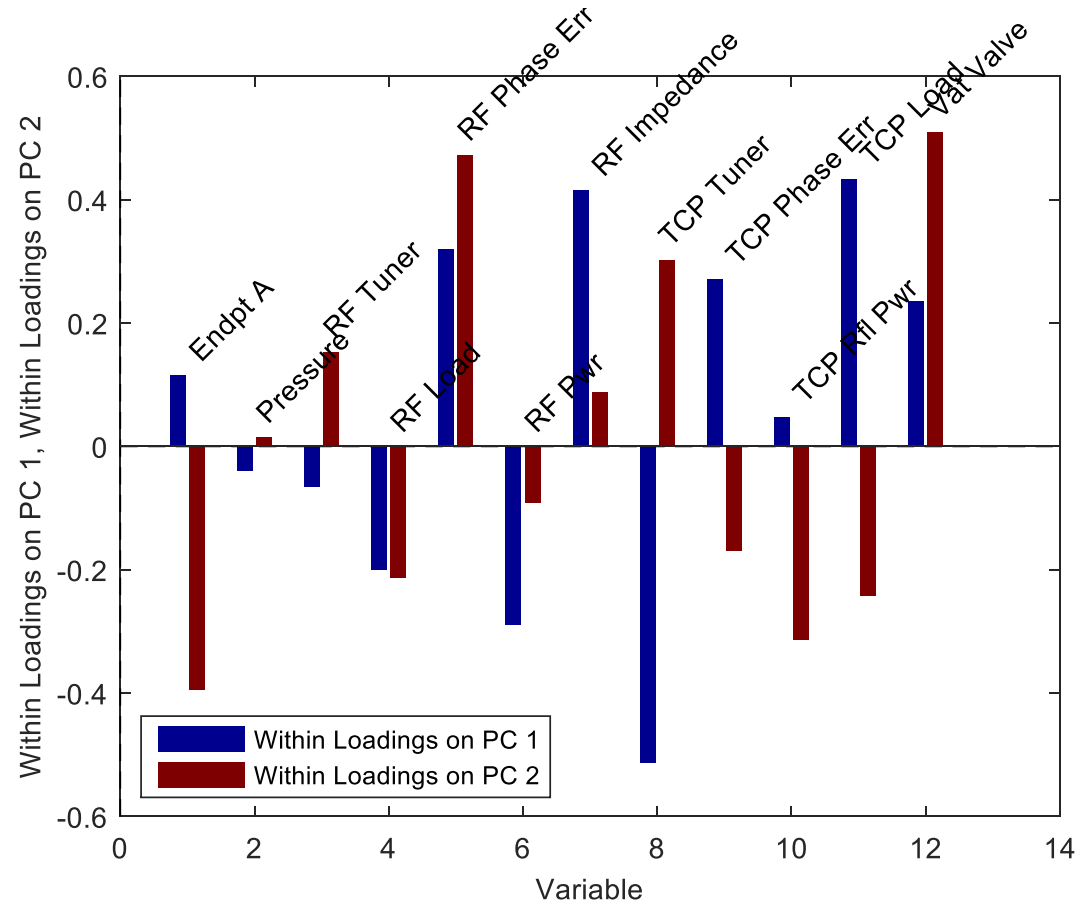MLSCA allows the variation associated with each nested factor to be resolved, and to see the main variables involved.

- 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.

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.

31