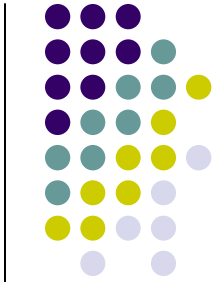
Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS) of Biomass and Coal: A Chemometrics Analysis.

Guilin Jiang,[1] Bonnie J. Tyler,[3]
Larry Baxter,[2] Matthew R. Linford[1]

1. Department of Chemistry and Biochemistry, Brigham Young University.
2. Department of Chemical Engineering, Brigham Young University.
3. Department of Chemical Engineering, University of Utah.

Outline



- Intro to ToF-SIMS
- Intro to Multivariate Statistics
- Principal Component Analysis (PCA) of Coal and Biomass
- Cluster Analysis of Coal and Biomass
- Partial Least Squares (PLS)
- Conclusions

Static ToF-SIMS: A Chemical Microscope

Primary Ion Beam:
Ga^+, In^+, SF_5^+, Au_2^+, Au_3^+ *etc*.
(10-25 KeV)

Mass Analysis of Ions

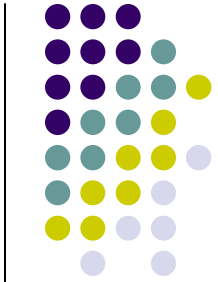
Neutrals.
Secondary Ions.
Electrons.

Substrate

0 – 25Å

Note: 4 eV corresponds to ~ 92 kcal/mol

Samples Analyzed by ToF-SIMS:



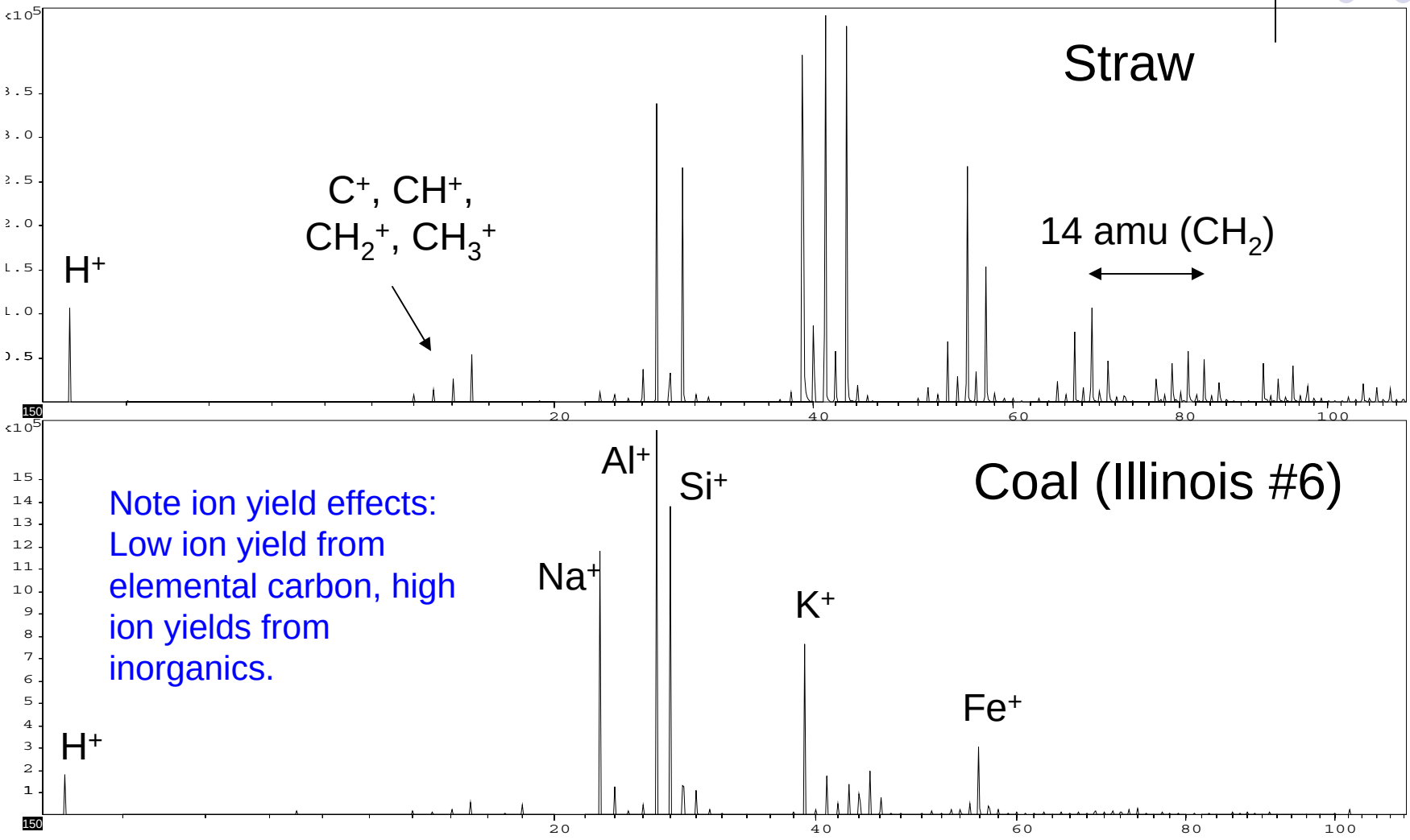
Coal.

- black thunder
- Ill#6
- cjr012
- gat034
- gat035
- gbj010
- glj005
- gpb062
- hgb011
- hna015
- jig014
- jig017
- pct010p
- pct012
- rbl014
- rll012
- rll014

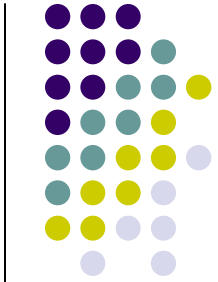
Biomass.

- grainscreening
- sawdust
- sugarbeet pulp
- sheanut shells
- straw
- sunflower shells

Examples of ToF-SIMS Spectra of Biomass and Coal.



Why Use Multivariate Statistics?

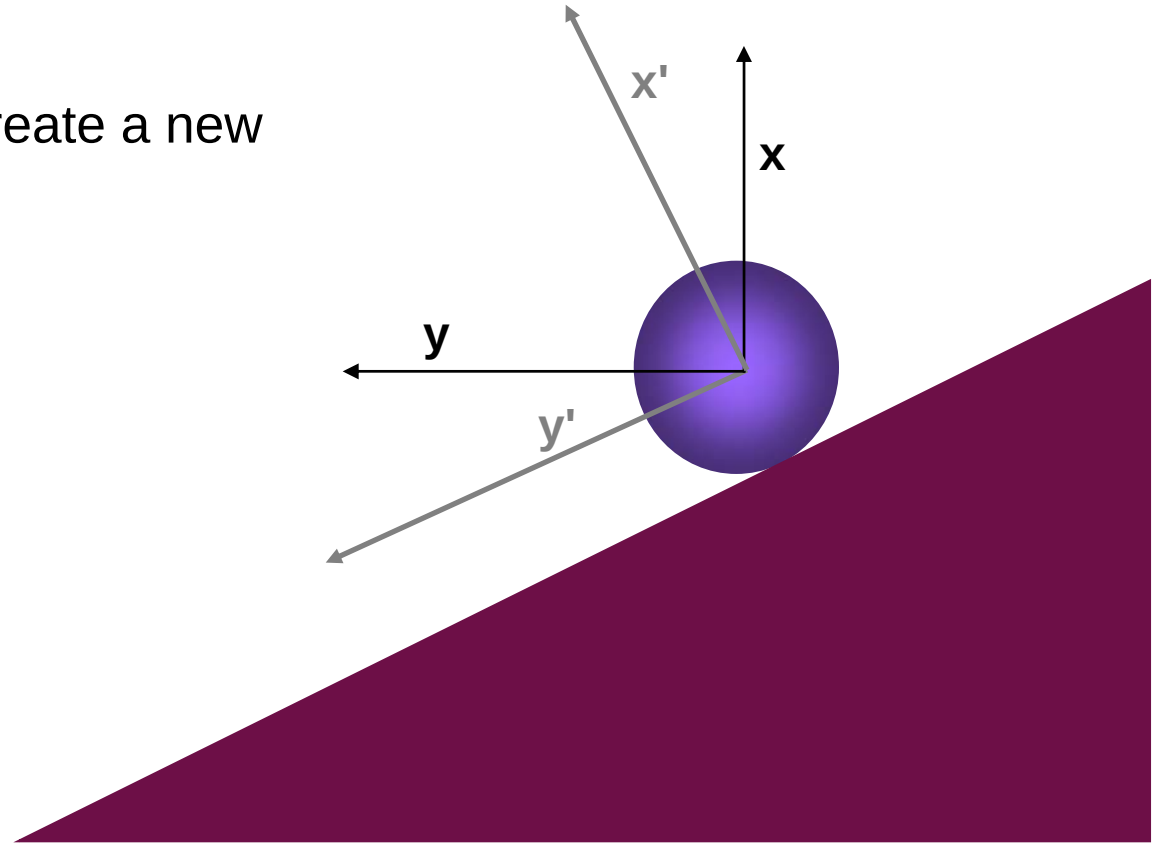


- Quantitative Analysis
- Reduce Size of Huge Data Sets
  - Keep important information
- Remove User Bias
- Efficiently Use the Data
- Chemical Signatures are Multivariate

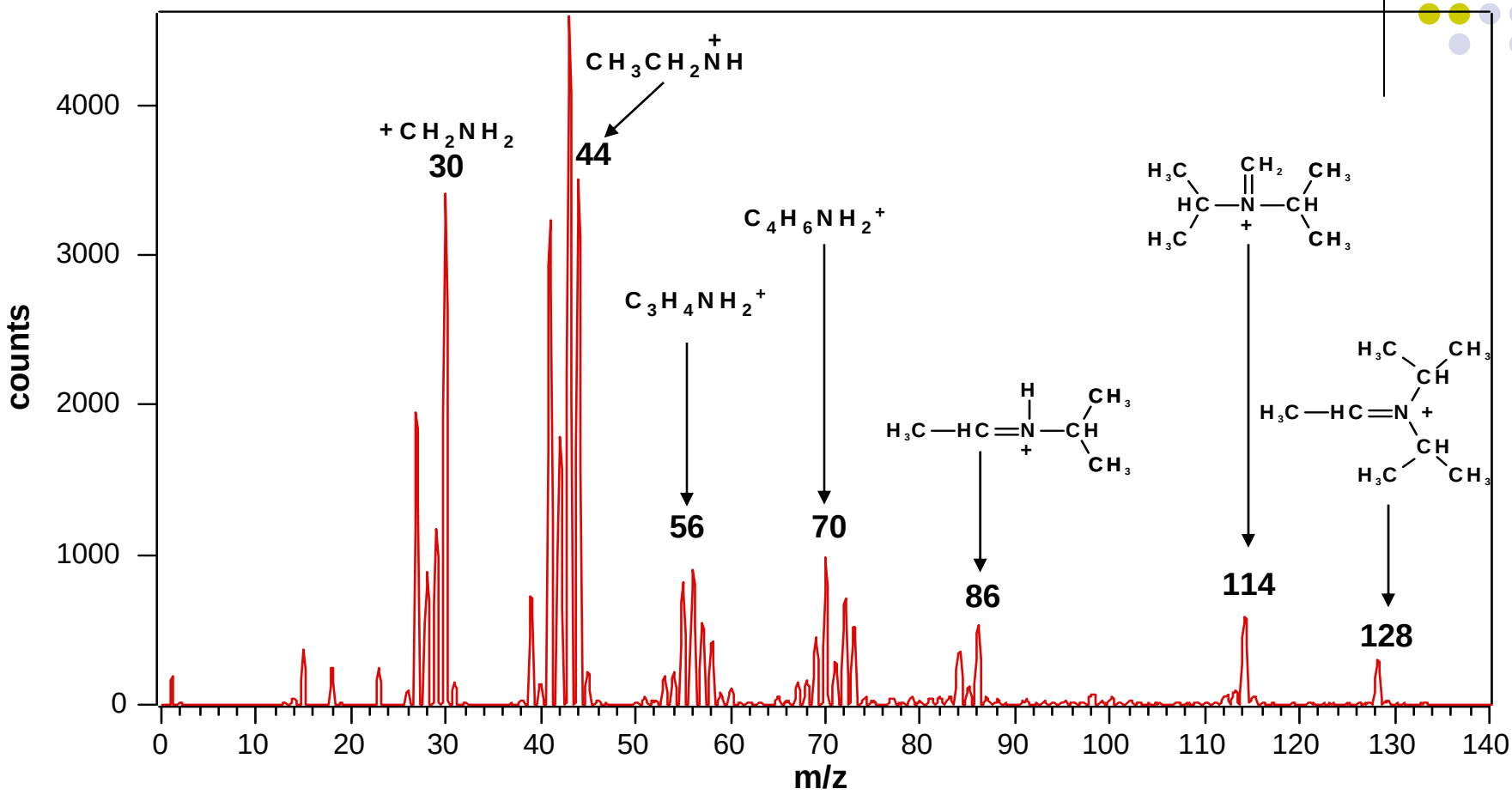
Multivariate Analysis

Reducing the Dimensionality of a Problem

Rotate the axes to create a new coordinate system:



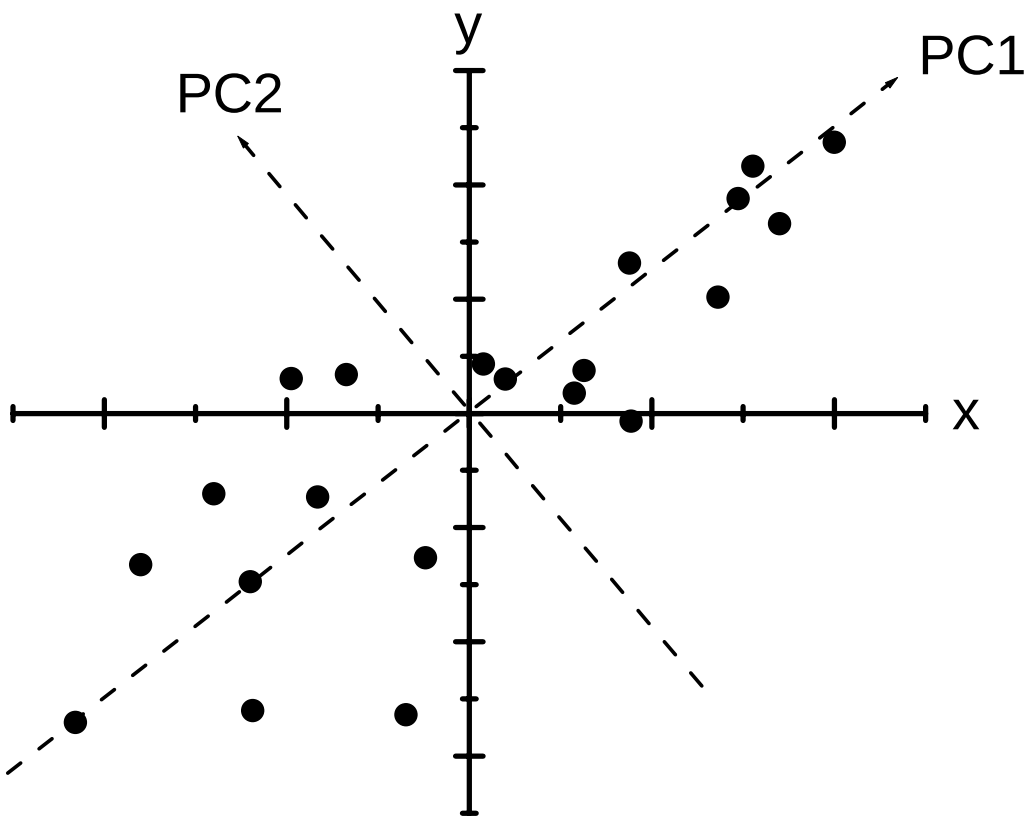
Positive Ion SIMS for Medical Grade Lycra Spandex.



It should be possible to reduce the dimensionality of SIMS data because many peaks can originate from one molecule and so they should co-vary in the data set.

Principal Components Analysis

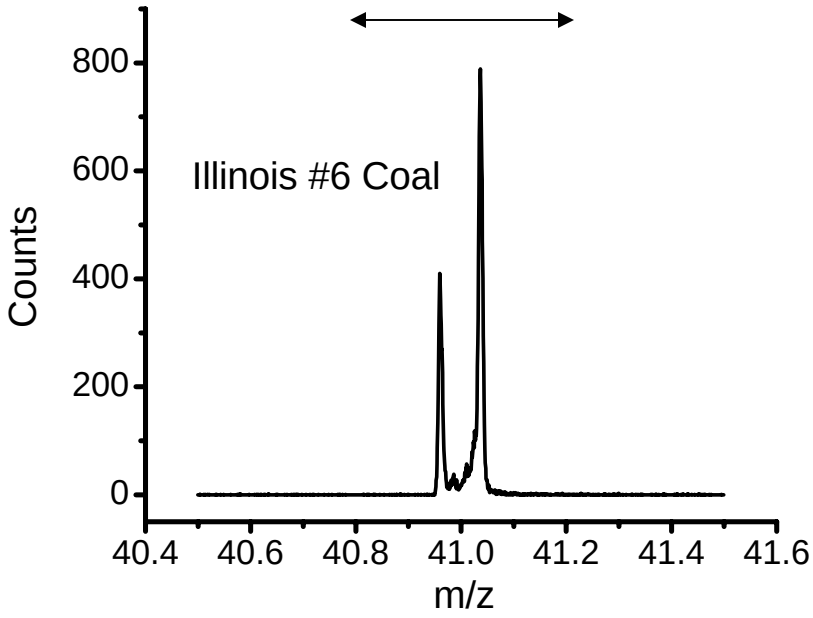
- Rotate the coordinate system.
- The first principal component (eigenvector) is the axis that accounts for the most variation in the data.
- The second principal component is the axis that accounts for the next largest amount of variation in the data, and so forth. PC2 $\perp$ PC1.
- Scores = Projections of the new data on the principal components (new axes).
- Loadings = Contribution of old variables to new variables.
- One can work with many variables/axes (a hyperspace).



Data Preprocessing.

- Select 80 regions (from 1-100 amu).
- Bin data.
- Row scale the data.
- Allows spectra taken under different conditions to be compared.
- Emphasizes large peaks.

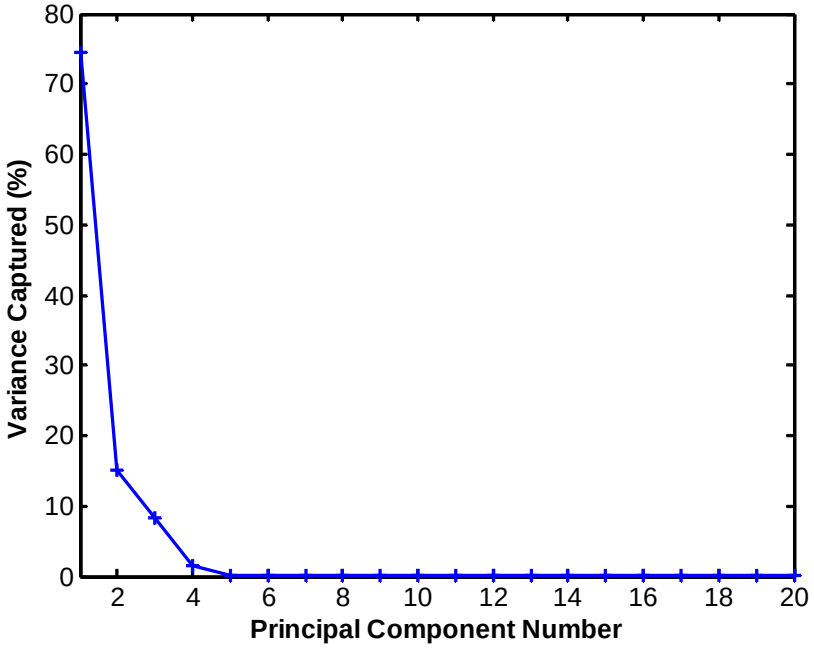
$$ {}^{rs}x_{ij} = \frac{x_{ij}}{\sum_{j=1}^{J} x_{ij}} $$



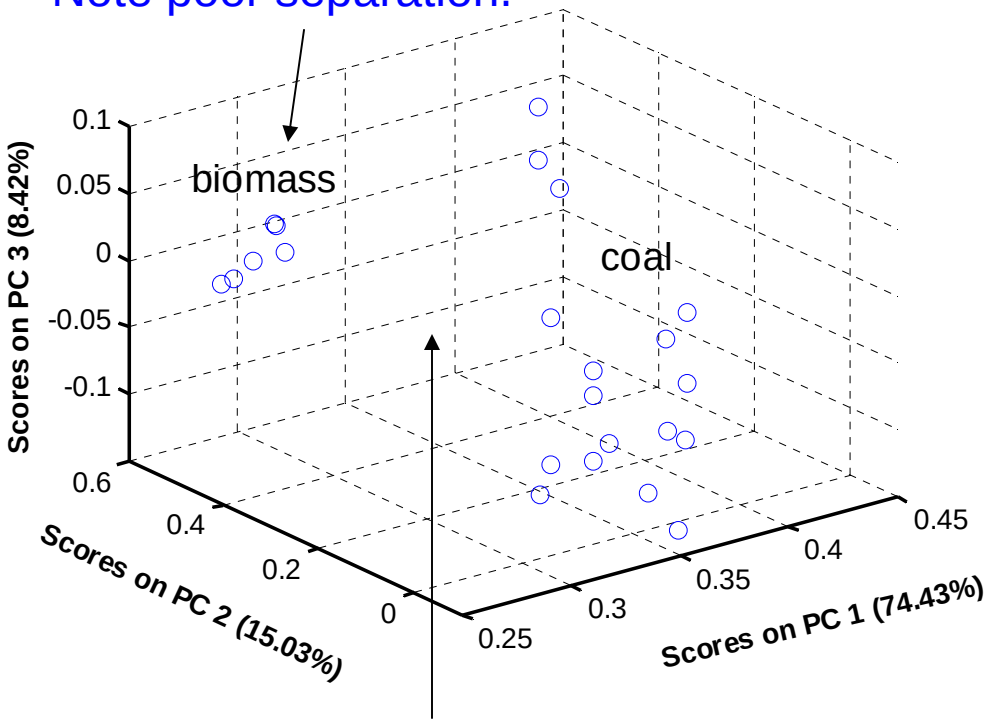
Mock data:

Raw Data:				
	Variables:			
A	1	2	45	1
B	2	6	77	3
C	14	2	1	1
D	10	1	3	2
Row Scaled:				
	Variables:			
A	0.02	0.04	0.92	0.02
B	0.02	0.07	0.88	0.03
C	0.78	0.11	0.06	0.06
D	0.63	0.06	0.19	0.13

PCA: Variance Captured by Row Scaling Alone.



Note poor separation.



PC	Var.	Tot. Var.
1	74.43	74.43
2	15.03	89.46
3	8.42	97.88
4	1.54	99.42

None of the PC's (alone) separate biomass from coal (very well).

Data Preprocessing: Standardization.

- Take the row scaled data.
- Mean Center.
- Divide by the Standard Deviation.
- Puts variables on equal footing.

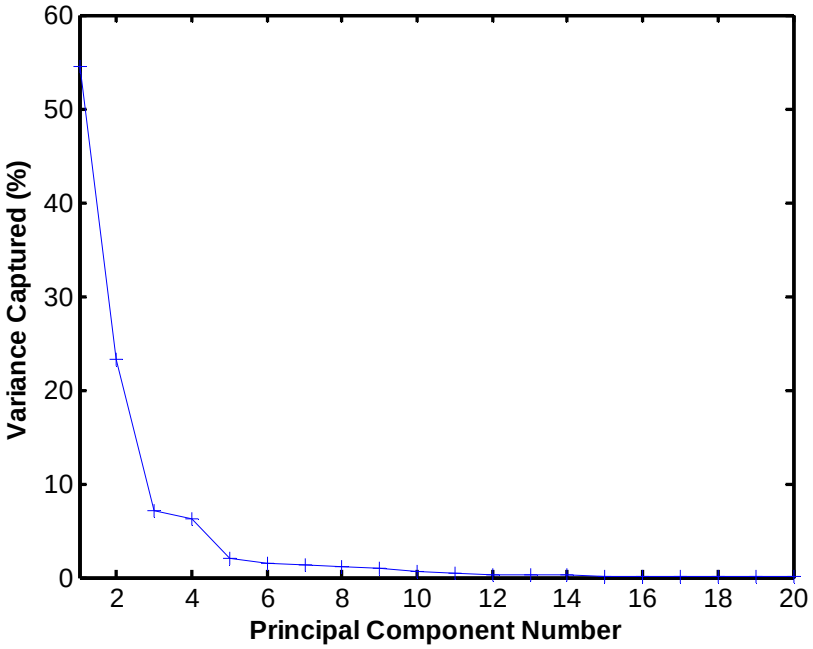
$$ {}^{cen}x_{ij} = x_{ij} - \bar{x}_j $$

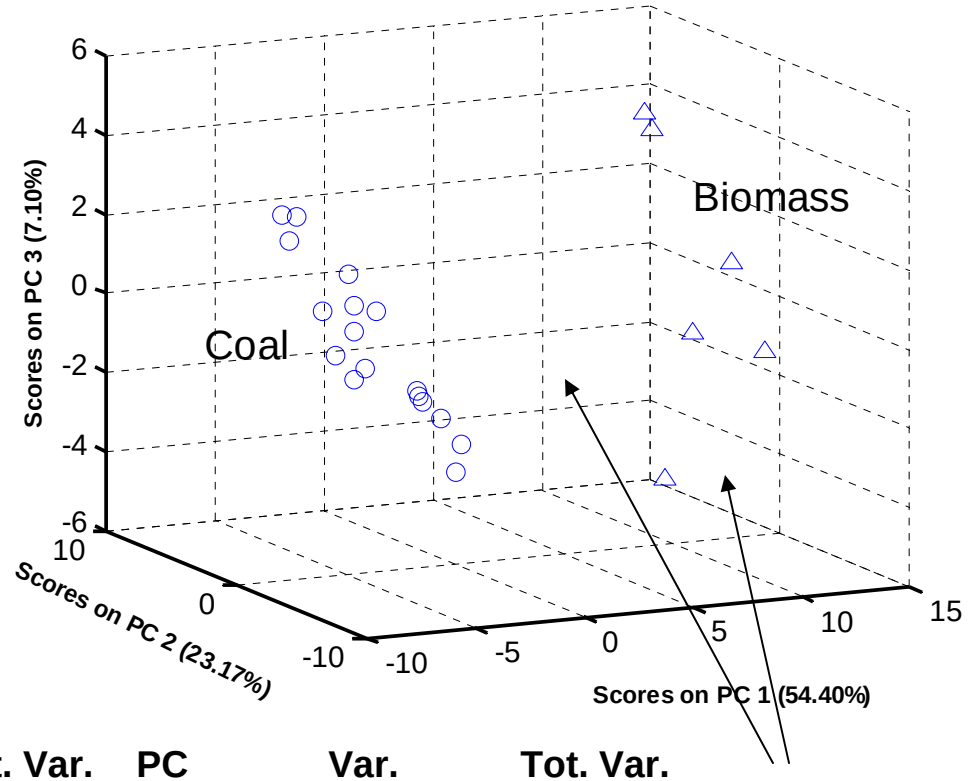
$$ {}^{stn}x_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{\sum_{i=1}^{I} \frac{\left(x_{ij} - \bar{x}_j\right)^2}{I}}} $$

Mock data:

Raw Data:				
	Variables:			
A	1	2	45	1
B	2	6	77	3
C	14	2	1	1
D	10	1	3	2
Row Scaled:				
	Variables:			
A	0.02	0.04	0.92	0.02
B	0.02	0.07	0.88	0.03
C	0.78	0.11	0.06	0.06
D	0.63	0.06	0.19	0.13
Standardized:				
	Variables:			
A	-0.99	-1.17	1.05	-0.95
B	-0.98	-0.10	0.94	-0.61
C	1.21	1.59	-1.16	-0.08
D	0.77	-0.32	-0.82	1.65

PCA: Variance Captured by Row Scaling, (Mean Centering), and Standardization.

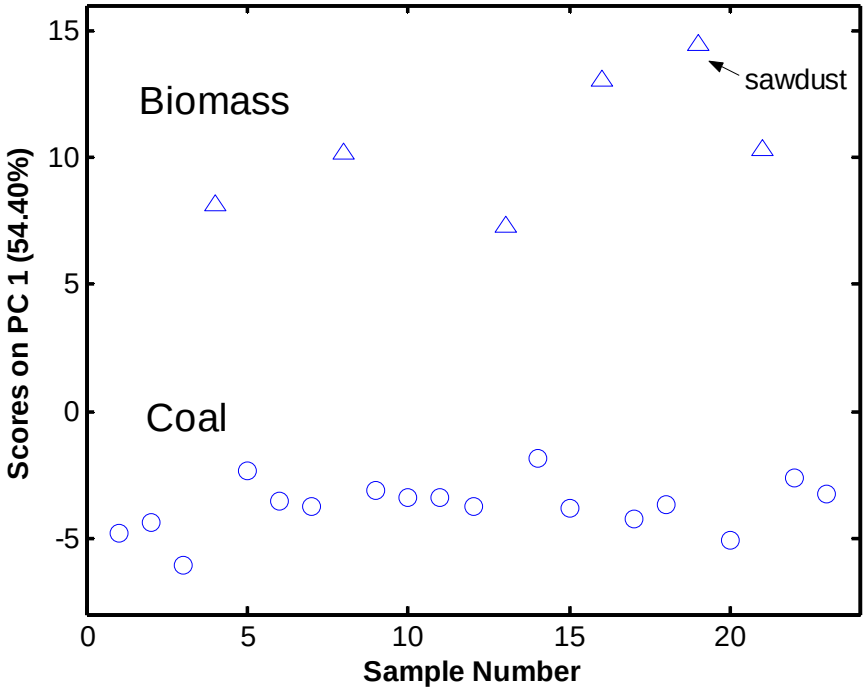


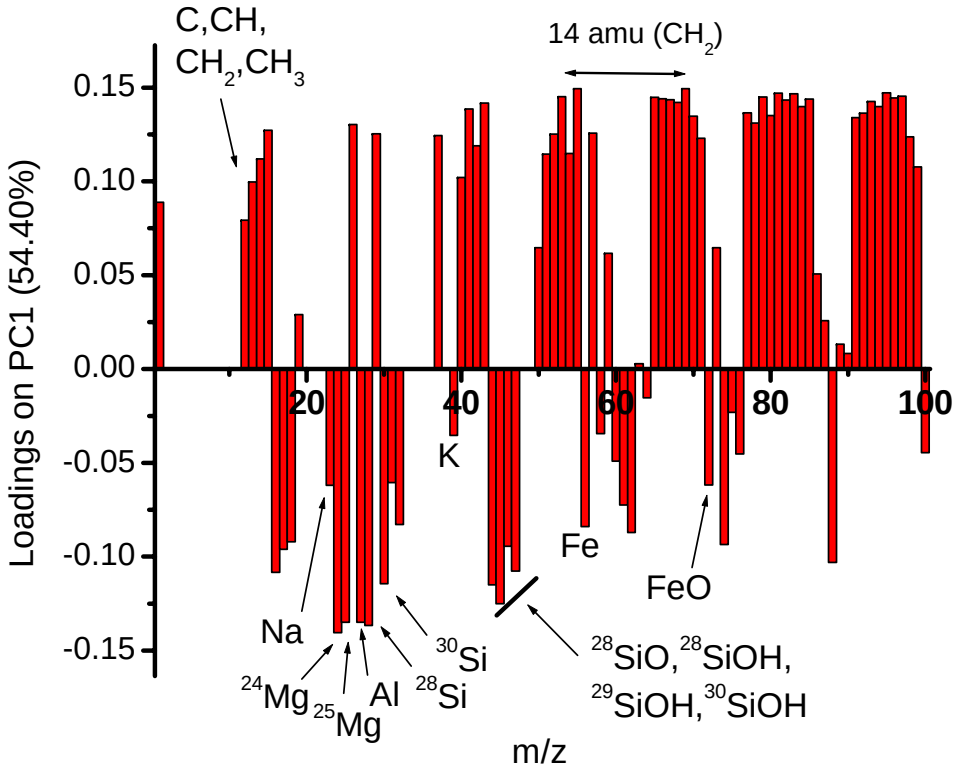


PC	Var.	Tot. Var.	PC	Var.	Tot. Var.
1	54.40	54.40	7	1.32	95.64
2	23.17	77.58	8	1.05	96.70
3	7.10	84.68	9	1.01	97.71
4	6.16	90.84	10	0.68	98.39
5	2.02	92.86	11	0.45	98.84
6	1.46	94.32	12	0.33	99.17

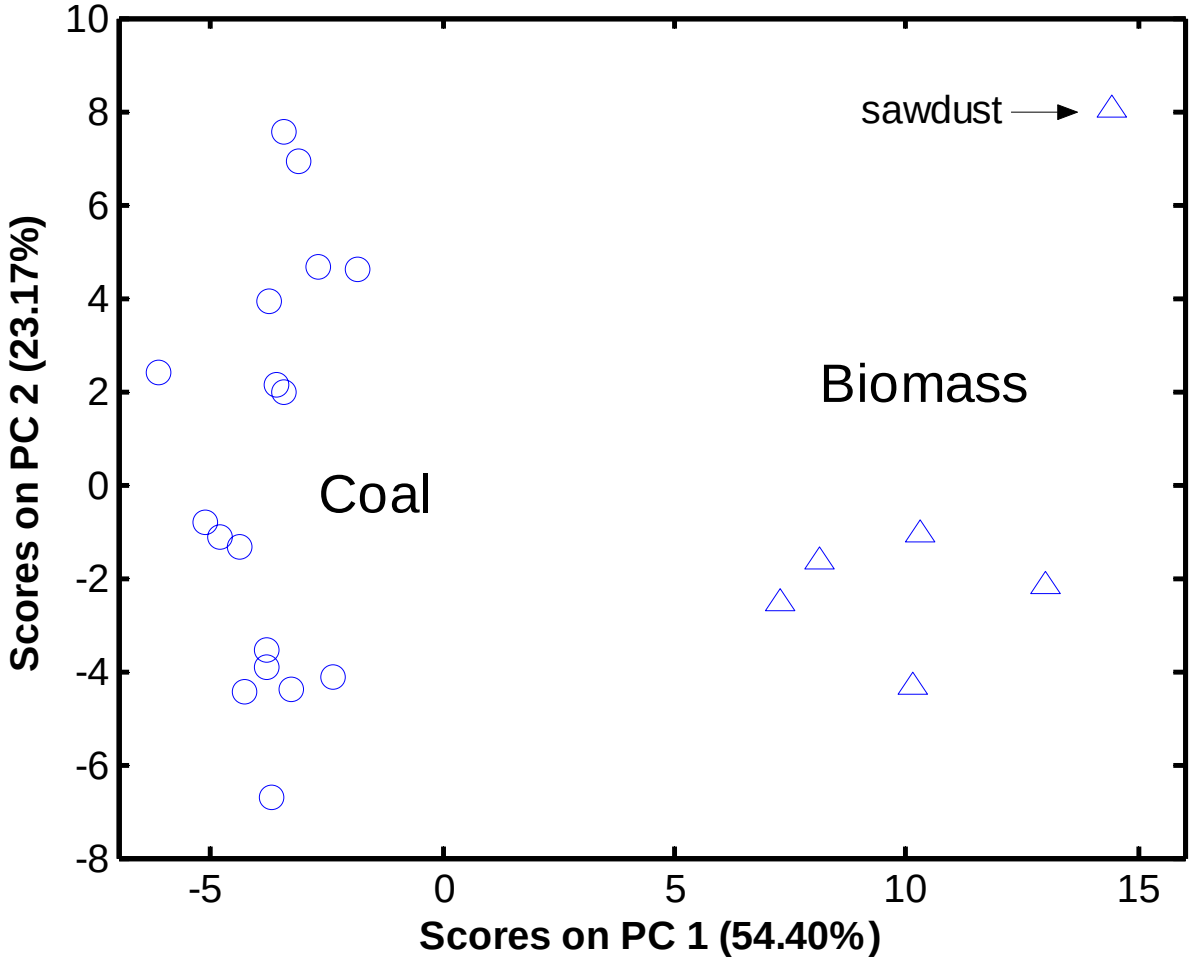
Note good separations.

PC1 Separates Coal from Biomass.

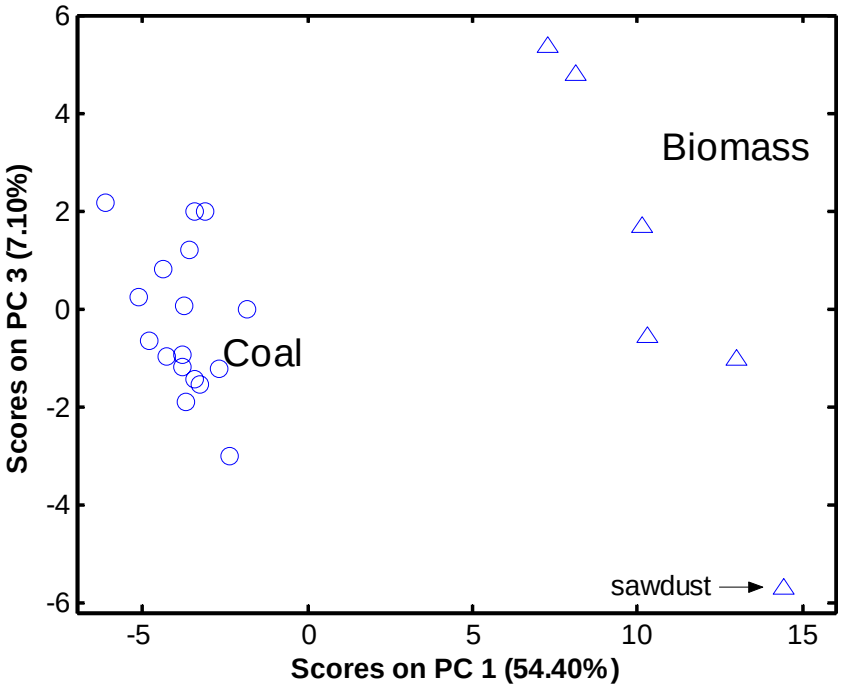


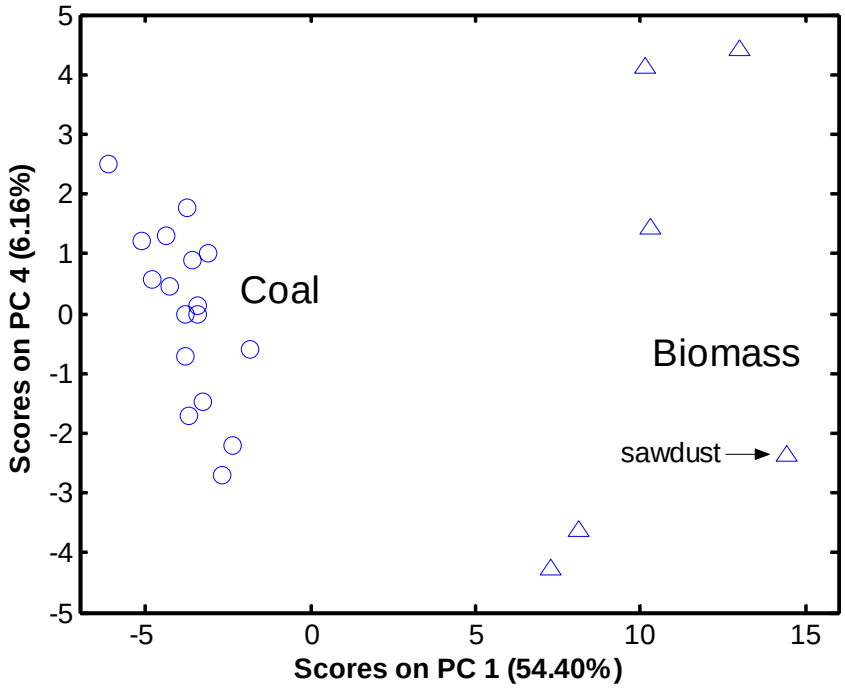


PC2 Separates Coal and Biomass individually (but not from each other).

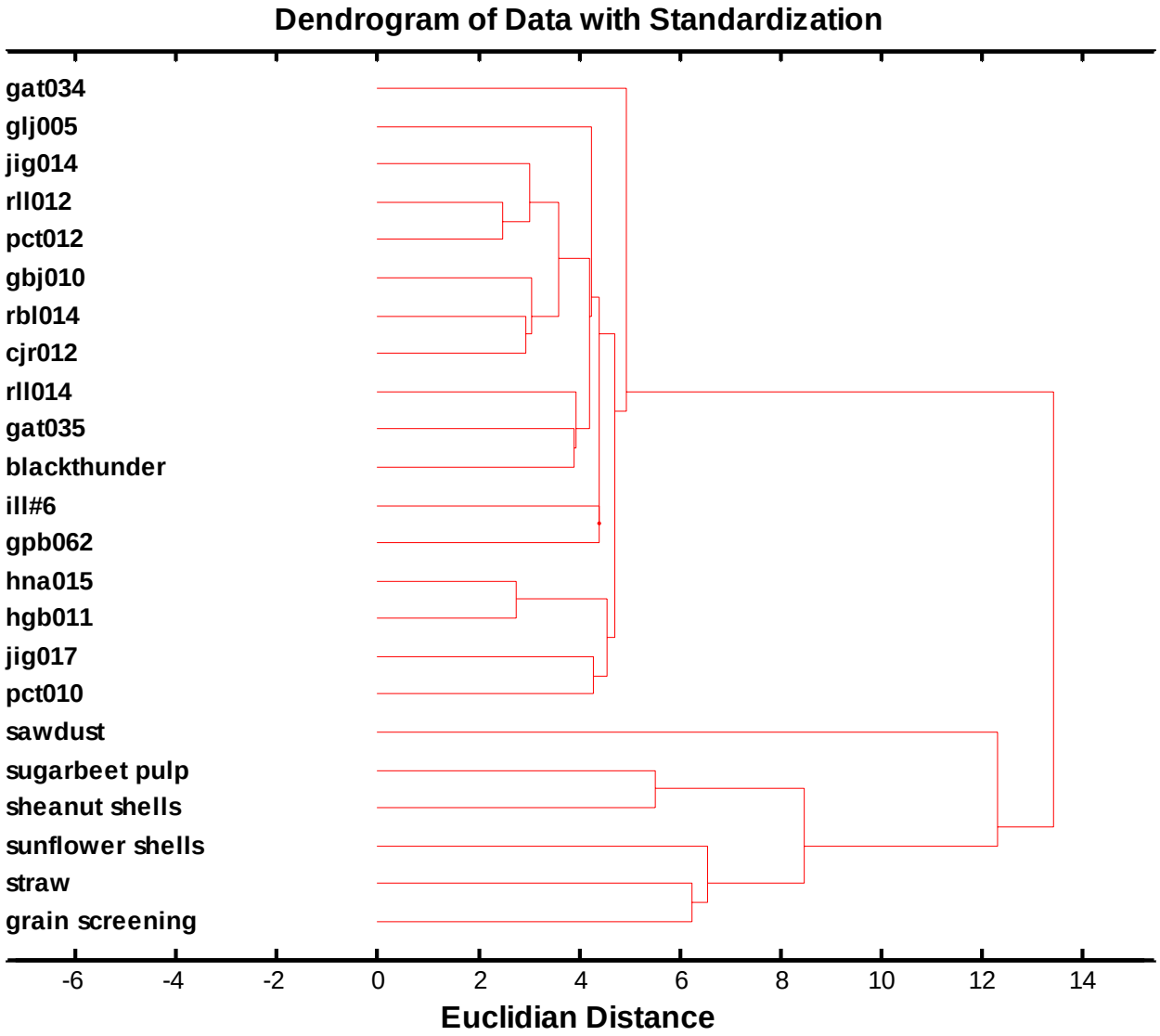


PC3 and PC4 Separate biomass (and coal) individually (but not from each other).





Cluster Analysis Separates Coal and Biomass



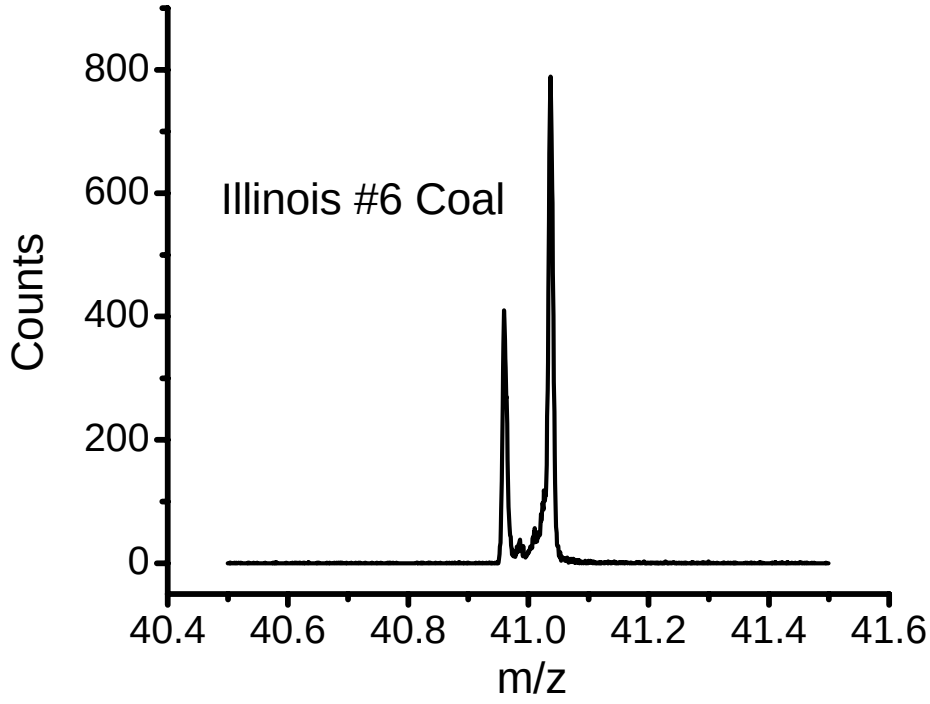
$$d_{xy} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}$$

Improved Preprocessing.
Mass Excesses and Mass Deficits.

mass excess

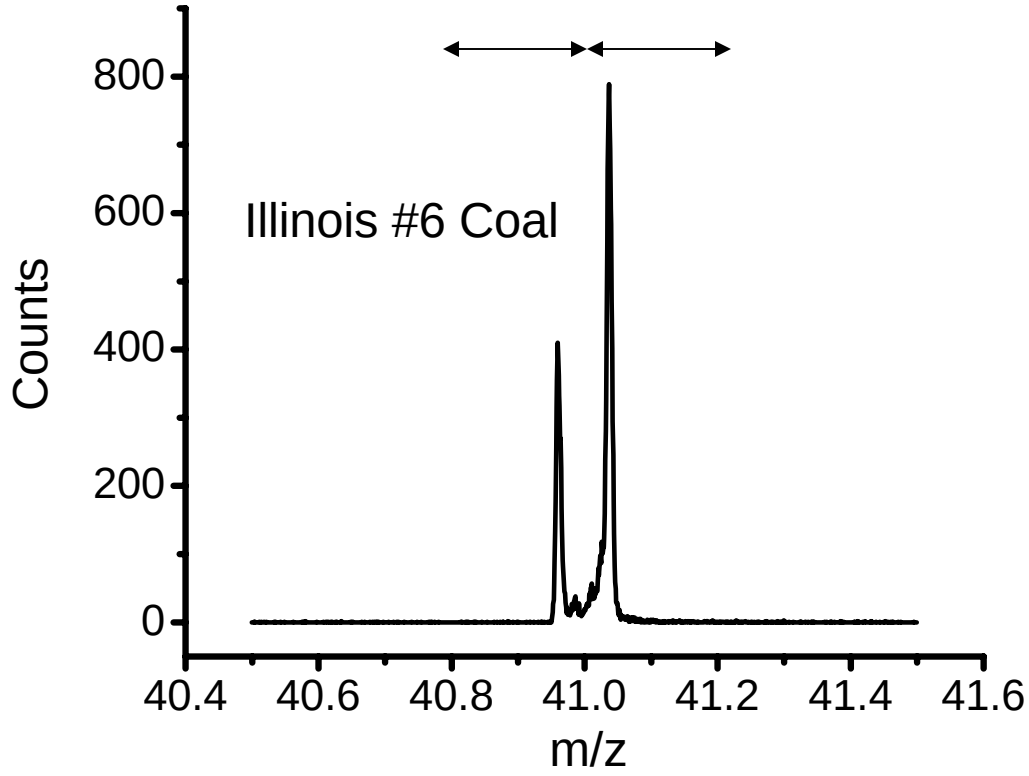
mass deficit

H	1.0078
He	4.0026
Li	7.0160
Be	9.0122
B	11.0093
C	12.0000
N	14.0031
O	15.9949
F	18.9984
Ne	19.9924
Na	22.9898
Mg	23.9850
Al	26.9815
Si	27.9769
K	38.9637
Ca	39.9626
Fe	55.9349

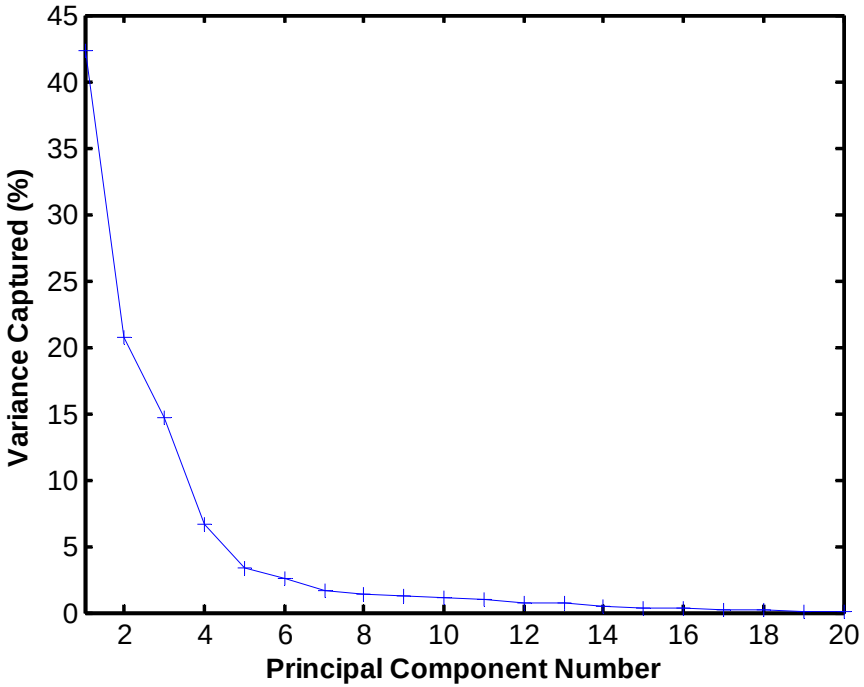


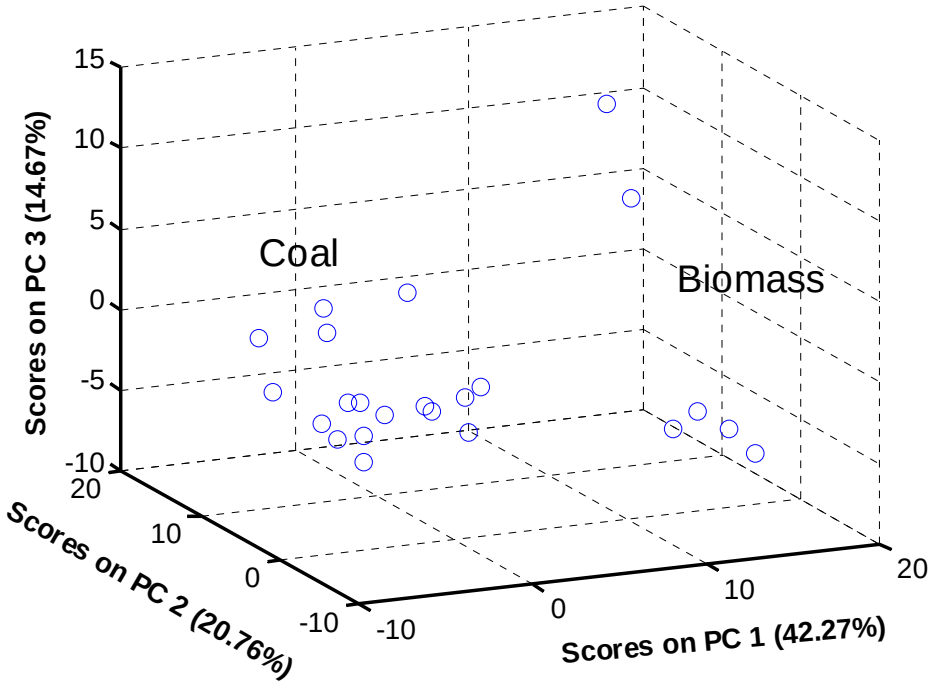
Data Preprocessing: Improved Binning.

- The difference between organic and inorganic components in coal and biomass is a significant source of variation in the data.
- Separate data into organic and inorganic components.
- Increase the number of peaks from 80 to 165.



PCA: Row Scaling, Standardization, Binning the Inorganic and Organic Regions of each Mass.

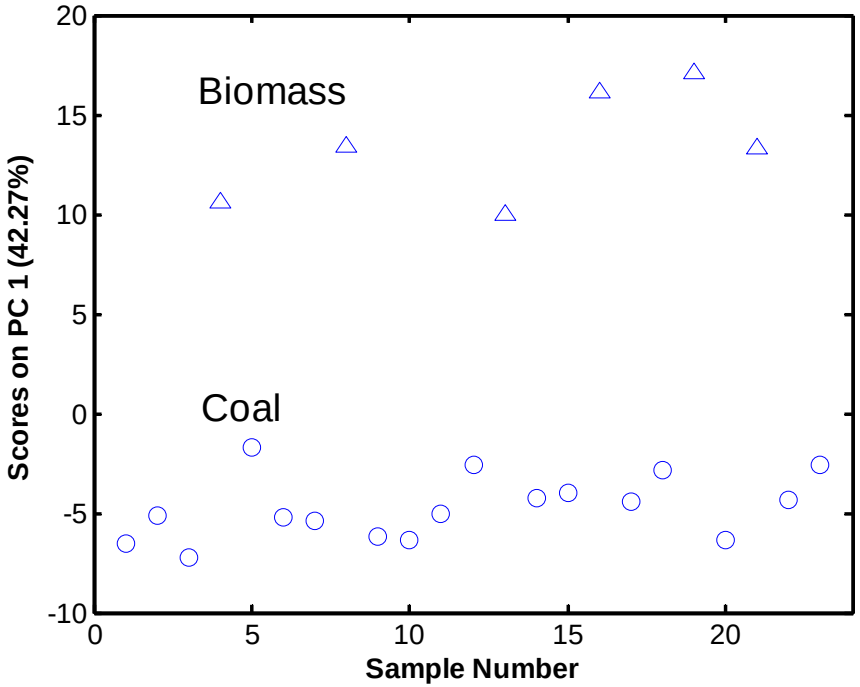


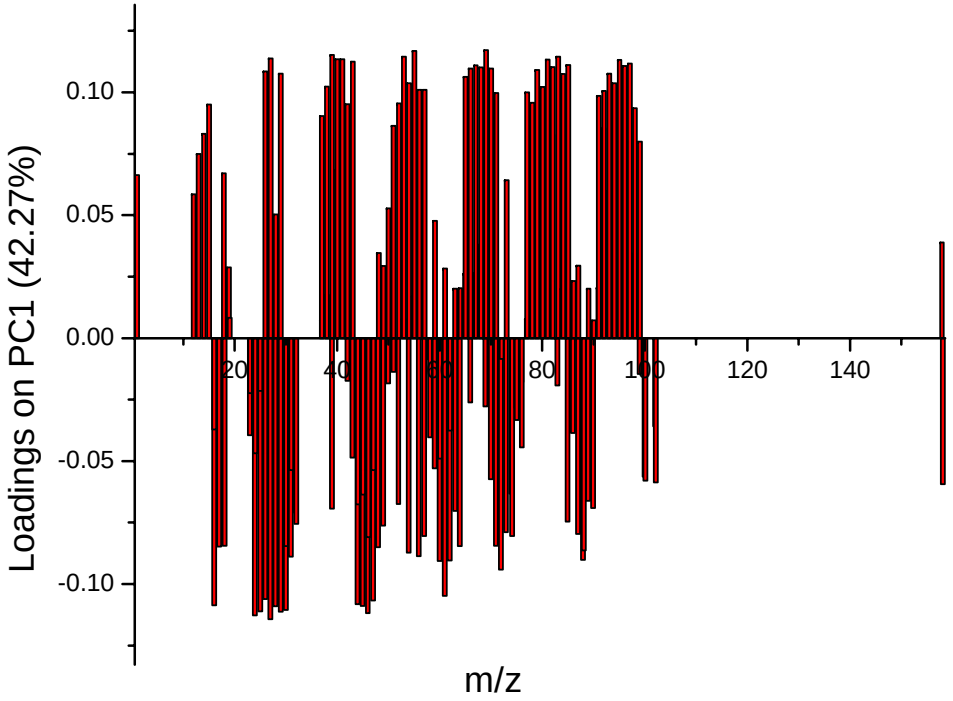


PC	Var.	Tot. Var.
1	42.27	42.27
2	20.76	63.03
3	14.67	77.70
4	6.59	84.29

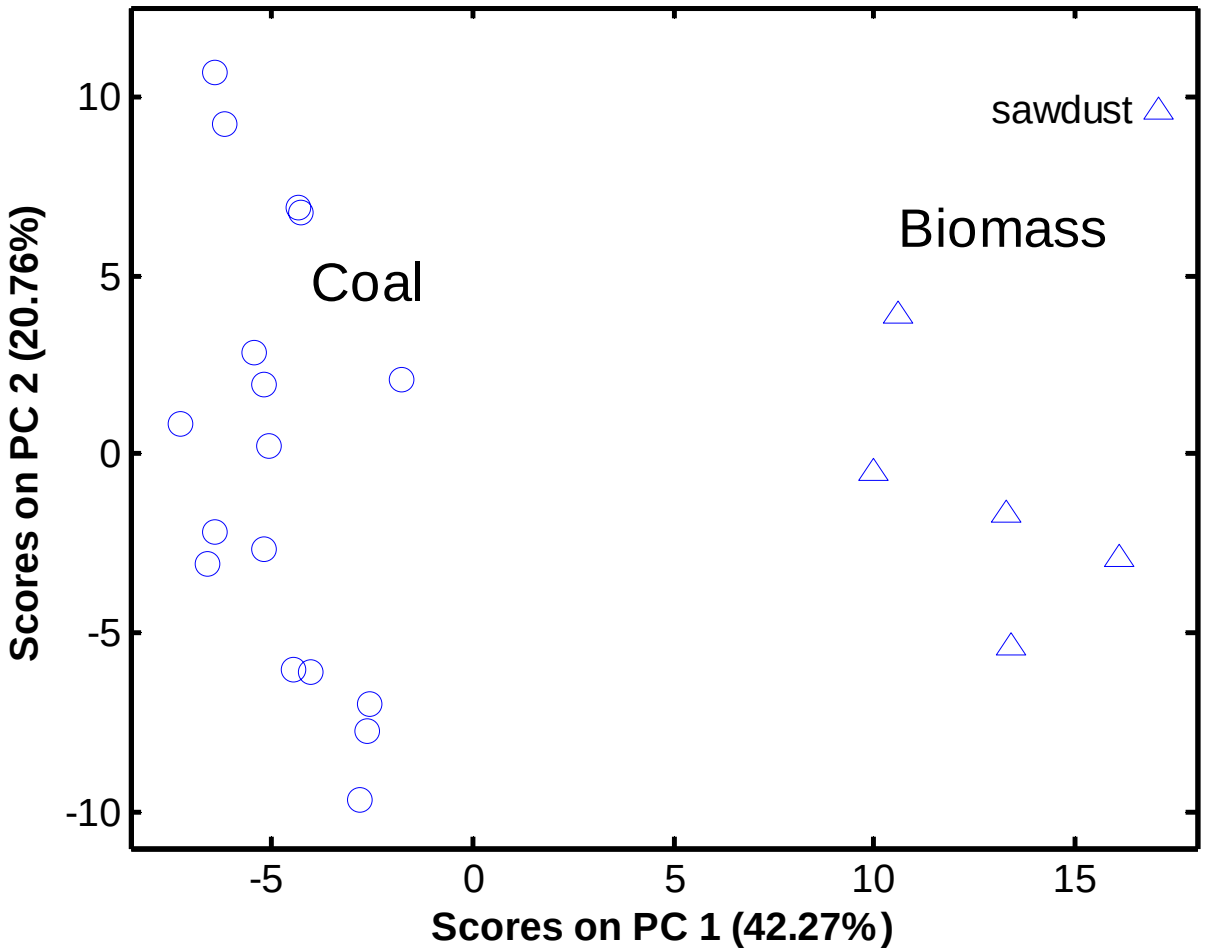
It takes 16 PC's to capture 99% of the variance in the data.

PC1 Separates Coal from Biomass.

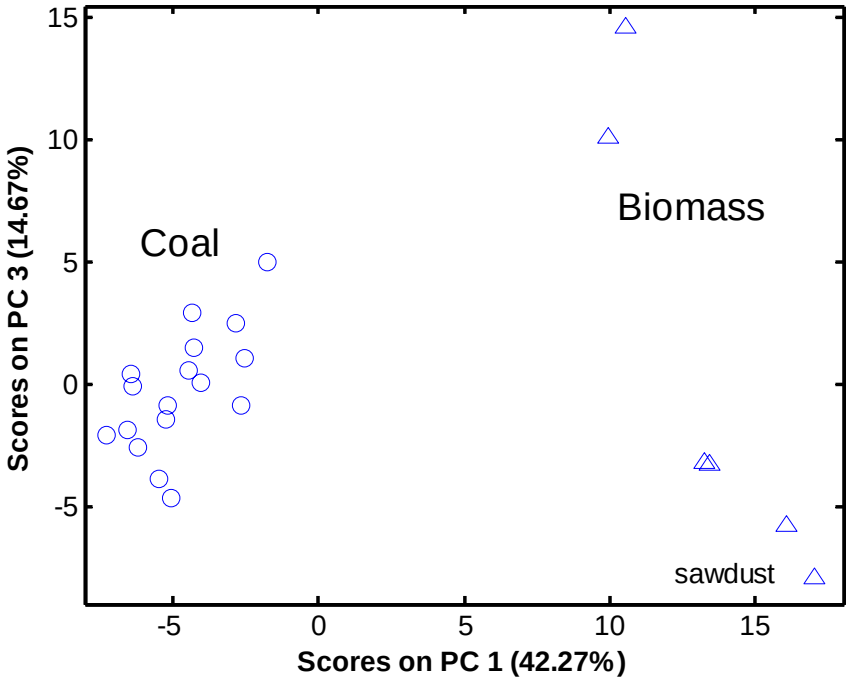


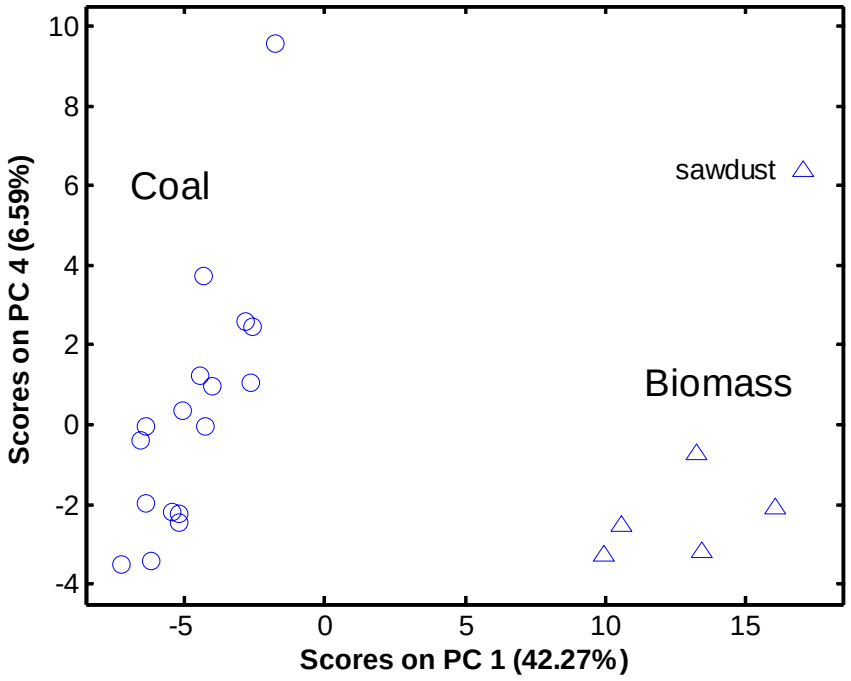


PC2 Separates Coal and Biomass Individually.

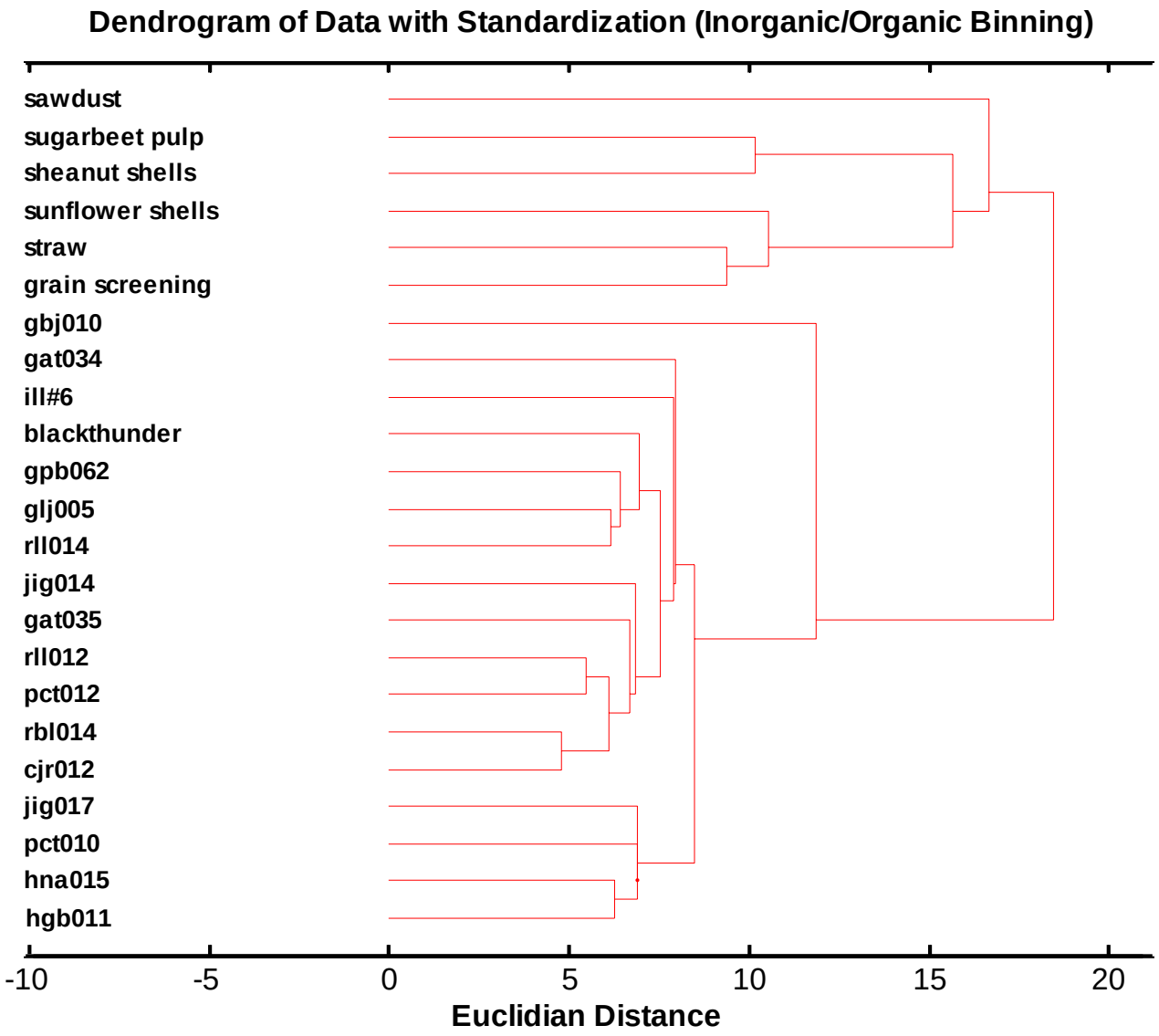


PC3 and PC4 Separate biomass (and coal) individually (but not from each other).



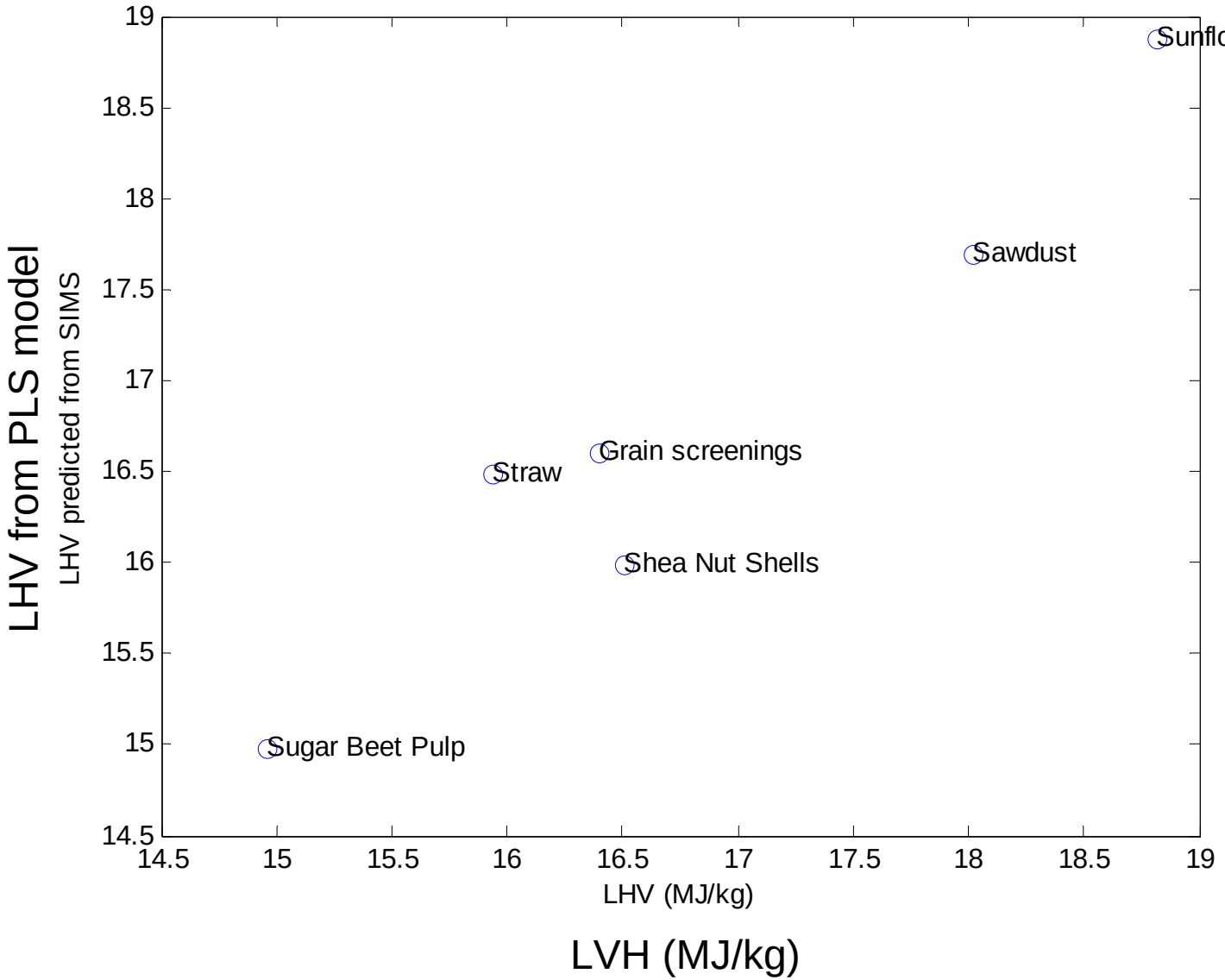


Cluster Analysis Separates Coal and Biomass

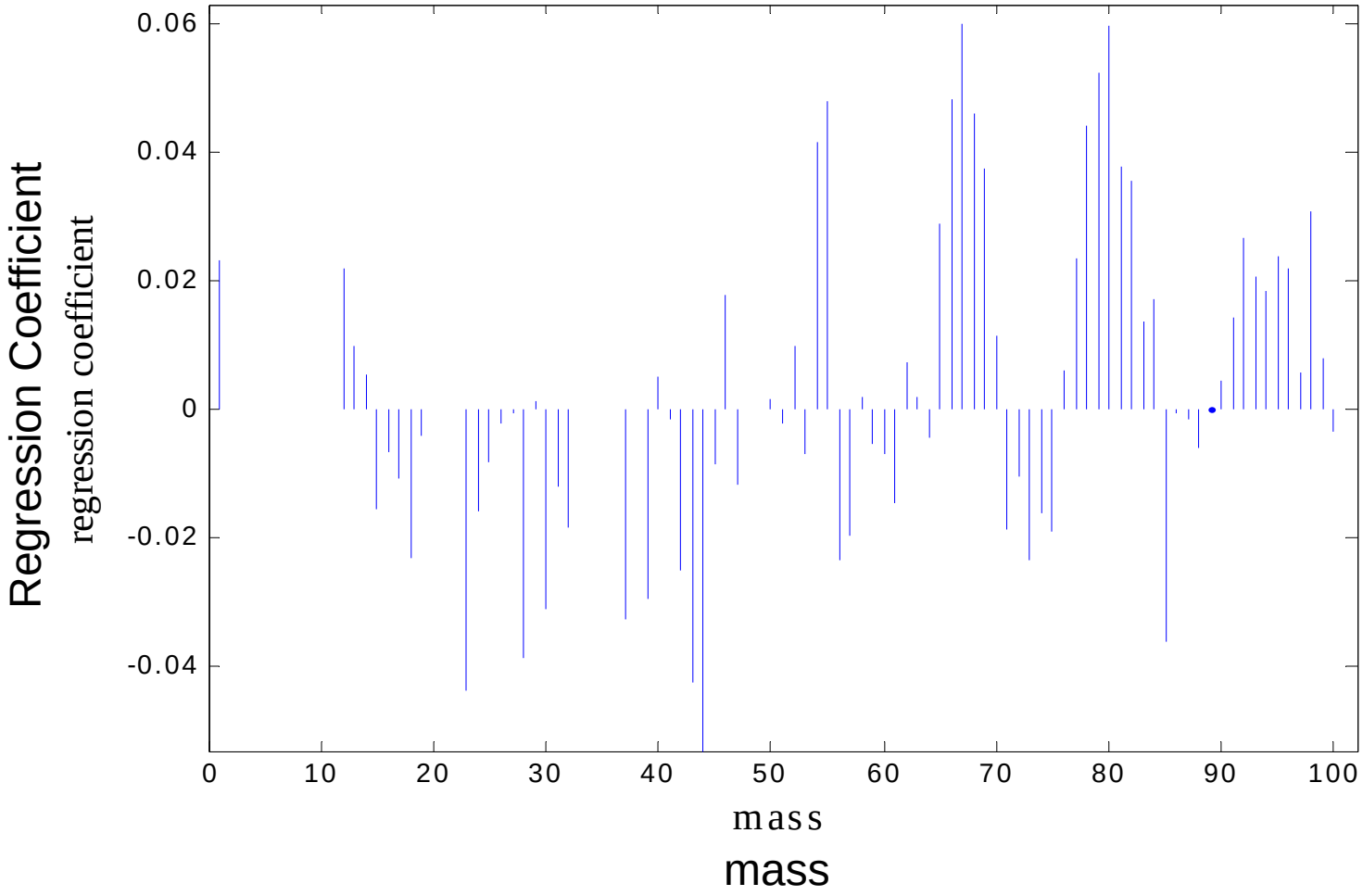


$$d_{xy} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}$$

LHV Predicted using 2 component PLS model of auto scaled SIMS data (80 peaks)

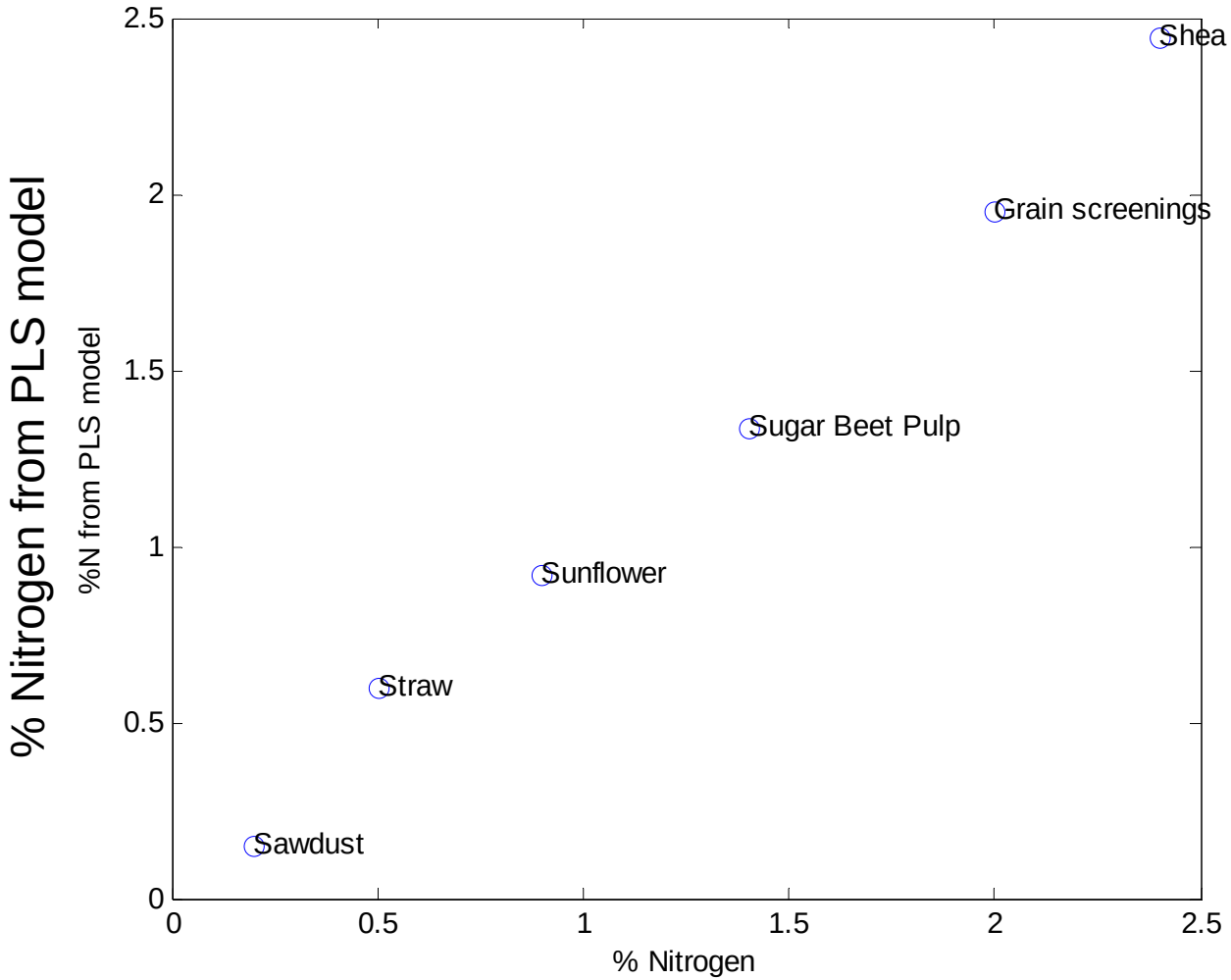


PLS regression coefficients



Peaks with negative coefficients (inorganics and oxygenated organics) lead to lower LHV while peaks with positive coefficients (longer chain hydrocarbons most likely from fatty acids) lead to higher heating values.

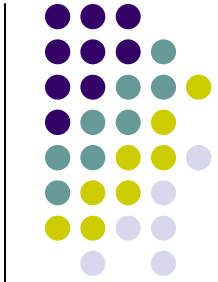
%N Predicted using 3 component PLS model of auto scaled SIMS data (80 peaks)



% Nitrogen

The regression coefficients show a strong weighting for NH_4^+ and for even mass peaks typical for nitrogen containing organic ions

Conclusions



- ToF-SIMS of coal and biomass can be performed.
- Multivariate analyses show that the spectra of coal and biomass are distinctly different.
- More chemical information is extracted from the data by standardization than by row scaling alone.
- Principal Components Analysis and Cluster Analysis help build the case for ToF-SIMS as a useful tool for fuel analysis.
- Future work: continue PLS (partial least squares) on the data to predict coal and biomass properties from ToF-SIMS data.