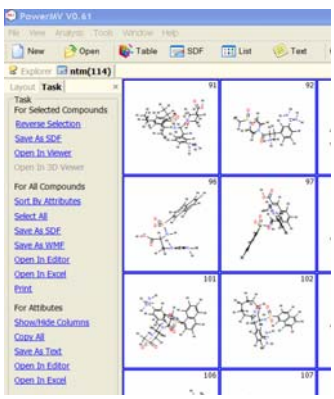
# PowerMV: A Statistical Tool for Chemists

Jun Feng and Jack Liu

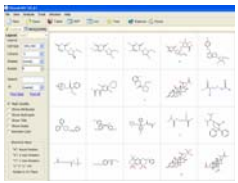
## Basic Functions

- Molecular Viewing
- Chemical Descriptor Generation
- Similarity Search
- Statistical Analysis

## Editing Features



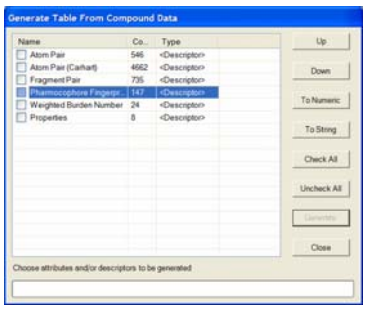
## Molecular Viewing



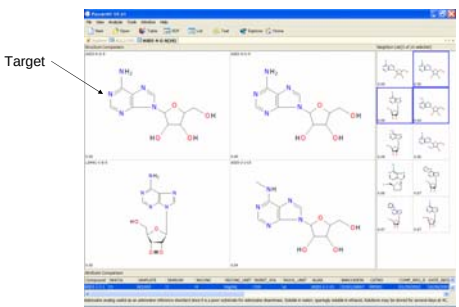
2D

3D

## Chemical Descriptor Generation



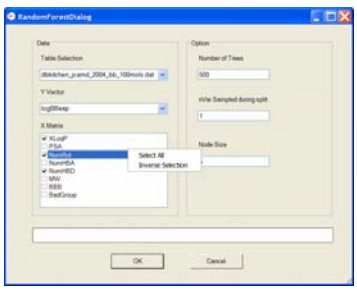

## Similarity Search



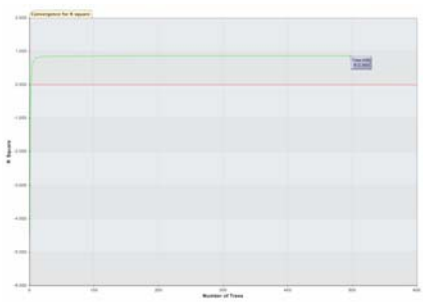
## Statistical Analysis

- Regression/Classification Model with Random Forest ® (through Interface to R package)
- Robust Single Value Decomposition
  Also available in PowerArray ®
- Outlier Detection with Tetrads method.

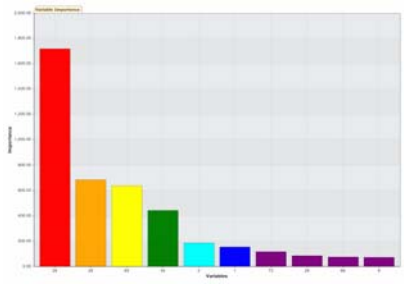
## GUI Interface to Random Forest®: Parameters Input



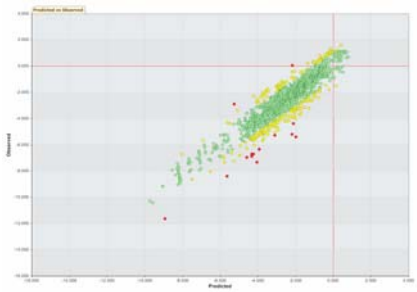

## GUI Interface to Random Forest®: Final Result Display



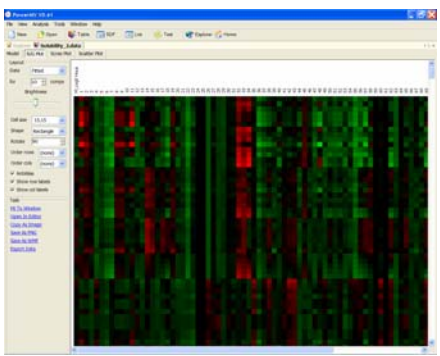
## GUI Interface to Random Forest®: Final Result Display



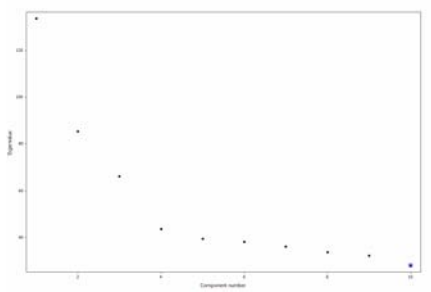
## GUI Interface to Random Forest®: Final Result Display



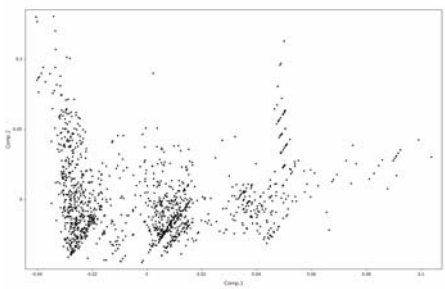
## Robust Single Value Decomposition



## Robust Single Value Decomposition: Scree Plot



## Robust Single Value Decomposition: Scatter Plot

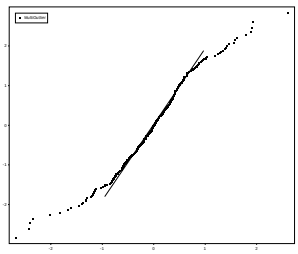


## Outlier Detection: Detect outliers in two-way tables.

| Labels | 1-Octanol | Ether | Chloroform | Benzene | Carbon Tetrac | Hexane |
|---|---|---|---|---|---|---|
| METHANOL | -0.77 | -1.15 | -1.26 | -1.85 | -2.1 | -2.8 |
| ETHANOL | -0.31 | -0.57 | -0.85 | -1.62 | -1.4 | -2.1 |
| PROPANOL | 0.25 | -0.02 | -0.4 | -0.7 | -0.82 | -1.52 |
| BUTANOL | 0.88 | 0.89 | 0.45 | -0.12 | -0.4 | -0.7 |
| PENTANOL | 1.56 | 1.2 | 1.05 | 0.62 | 0.4 | -0.4 |
| HEXANOL | 2.03 | 1.8 | 1.69 | 1.3 | 0.99 | 0.46 |
| HEPTANOL | 2.41 | 2.4 | 2.41 | 1.91 | 1.67 | 1.01 |
| ACETIC_AC | -0.17 | -0.34 | -1.6 | -2.26 | -2.45 | -3.06 |
| PROPIONIC | 0.33 | 0.27 | -0.96 | -1.35 | -1.6 | -2.14 |
| BUTYRICA | 0.79 | 0.61 | -0.27 | -0.96 | -0.97 | -1.76 |
| HEXANOIC | 1.92 | 1.95 | 1.15 | 0.3 | 0.37 | -0.46 |
| PENTANOI | 1.39 | 1 | 0.28 | -0.1 | -0.42 | -1 |
| TRICHLOR | 1.33 | 1.21 | -0.69 | -1.3 | -1.66 | -2.63 |
| DICHLORO | 0.92 | 1.31 | -0.89 | -1.4 | -2.31 | -2.72 |
| CHLOROAC | 0.22 | 0.37 | -1.92 | -1.6 | -2.56 | -3.14 |
| METHYLAC | 0.18 | 0.43 | 1.16 | 0.53 | 0.32 | -0.26 |
| ETHYLACE | 0.73 | 0.91 | 1.8 | 1.01 | 0.95 | 0.29 |
| ACETONE | -0.24 | -0.21 | 0.24 | -0.05 | -0.3 | -0.91 |
| ETHYLAMI | -0.3 | -1.18 | -0.35 | -1.3 | -1.27 | -1.77 |
| PROPYLAM | -0.28 | -0.54 | 0.26 | -0.52 | -0.59 | -1 |
| TRIMETHY | 0.27 | -0.26 | 0.54 | -0.29 | -0.09 | -0.48 |
| N-BUTYLA | 0.74 | 0.11 | 0.56 | -0.08 | -0.04 | -0.62 |
| DIETHYLA | 0.57 | -0.07 | 0.81 | -0.05 | 0.03 | -0.48 |
| PYRIDINE | 0.65 | 0.08 | 1.42 | 0.41 | 0.23 | -0.21 |
| ANILINE | 0.9 | 0.85 | 1.42 | 1 | 0.6 | -0.3 |

Bradu-Hawkins, Technometrics 1983?

## Outlier Detection: Detect outliers in two-way tables.



## Further Direction

- To improve similarity search methods.
- To include more statistical methods.
- To calculate more chemical descriptors.

## Acknowledgement


- Support from HighQ Foundation
- Dr.  Kejun Li
- Andrew Wong
- Dr. Stan Young
- Dr. Alan Karr