# Demo Example for Use of Software MassFeatGen

**K. Varmuza, Vienna, Austria**

From section 6 of User Guide, www.lcm.tuwien.ac.at (software)

A set of 10 mass spectra is used and 8 spectral features will be calculated. The spectra selected for this example and the features generated do not claim any spectroscopic relevance but are solely chosen for simplicity in a demo example.

### Mass spectra used

| No | Compound | Brutto formula | Mol.weigth | Class |
|---|---|---|---|---|
| 1 | Hexane | C6 H14 | 86 | 1 |
| 2 | Hexane, 2,3-dimethyl | C8 H18 | 114 | 1 |
| 3 | 1-Heptene | C7 H14 | 98 | 2 |
| 4 | Cyclopropane, pentyl- | C8 H16 | 112 | 2 |
| 5 | 4-Octanone | C8 H16 O | 128 | 3 |
| 6 | 4-Octanone, 7-methyl- | C9 H18 O | 142 | 3 |
| 7 | Chlorobenzene | C6 H5 Cl | 112 | 4 |
| 8 | Phenol, 4-chloro-3-methyl- | C7 H7 O Cl | 142 | 4 |
| 9 | Butylbenzene | C10 H14 | 134 | 5 |
| 10 | Benzene, 3-butenyl- | C10 H12 | 132 | 5 |

The data set contains five substance classes, each with two spectra: class 1 contains alkanes, class 2 hydrocarbons with one double bond equivalent, class 3 aliphatic ketones, class 4 aromatic chloro compounds, and class 5 benzyl compounds.

### Feature computed

1 Square root of intensity at m/z 91
2 Feature "DUST" (relative peak intensities up to m/z 78)
3 Feature "IBAS" (base peak intensity in % of total intensity sum)
4 Feature "EVEN" (relative peak intensities at even mass numbers)
5 Logarithmic intensity ratio of intensities at m/z 91 and m/z 92
6 Intensities of peaks from ions $C_nH_{2n+1}$
7 Intensities of peaks from ions $C_nH_{2n-1}$
8 Measure for isotope peak pattern of one chlorine atom

The created feature file looks as follows

| Feature | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| | 0.00000 | 97.32912 | 17.57160 | 24.35424 | 0.00000 | 70.37371 | 40.82412 | 99.85750 |
| | 3.16228 | 98.48934 | 26.97599 | 25.86998 | 0.00000 | 69.66106 | 36.87509 | 99.76339 |
| | 3.16228 | 96.46806 | 15.35627 | 36.48649 | 0.00000 | 35.34522 | 76.36592 | 99.99683 |
| | 0.00000 | 86.14200 | 17.10864 | 47.08298 | 0.00000 | 46.26309 | 77.42817 | 99.17346 |
| | 0.00000 | 84.50660 | 15.54002 | 20.31080 | 0.00000 | 93.03446 | 40.42821 | 99.80399 |
| | 0.00000 | 87.21874 | 23.19647 | 22.59337 | 0.00000 | 70.61784 | 33.63502 | 99.96016 |
| | 0.00000 | 49.67499 | 34.21143 | 58.80944 | 0.00000 | 5.56341 | 3.16228 | 99.99991 |
| | 0.00000 | 45.30410 | 21.49151 | 37.56716 | 0.00000 | 7.96817 | 5.05525 | 99.97254 |
| | 100.00000 | 28.23735 | 34.90401 | 34.20593 | 13.10063 | 7.81527 | 11.25892 | 99.77883 |
| | 100.00000 | 39.72125 | 34.84321 | 27.87456 | 42.69360 | 0.00000 | 8.16497 | 99.99683 |

The feature file contains 10 rows (for 10 spectra) and 8 columns (for 8 features).

## Discussion of the generated features

An over-interpretation of spectral features should be avoided - the pragmatic way to test them for desired applications is recommended. Only a short discussion of some of the generated features is tried here.

Feature 1 is the scaled intensity at mass 91; as expected it is high for the two benzyl compounds and low for the others.

Features 2 (DUST), 3 (IBAS), and 4 (EVEN) do not show evident information about the used substance classes.

Feature 5 is based on the logarithmic intensity ratio $I(91)/I(92)$; as expected it is high for the two benzyl compounds and low for the others.

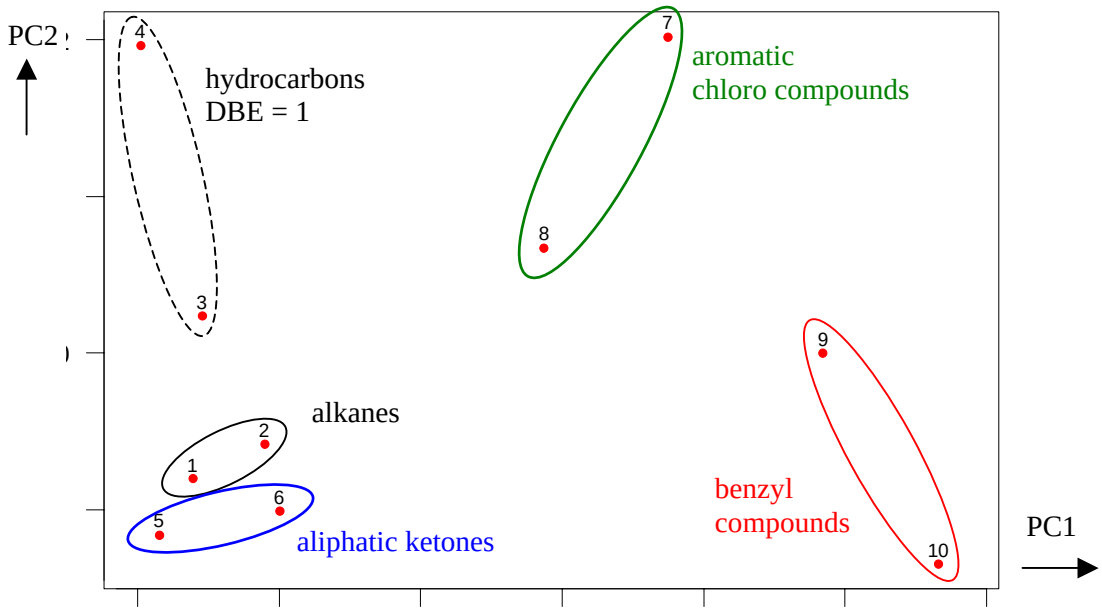
Feature 6 reflects the presence of $C_nH_{2n+1}$ ions; highest values appear with alkanes and aliphatic ketones. Feature 7 reflects the presence of $C_nH_{2n-1}$ ions; highest values appear with alkenes.

Feature 8 is considered to be sensitive for the isotope peak pattern of $Cl_1$. However, feature values show that all 10 spectra contain a peak group similar to the chlorine isotope peaks.

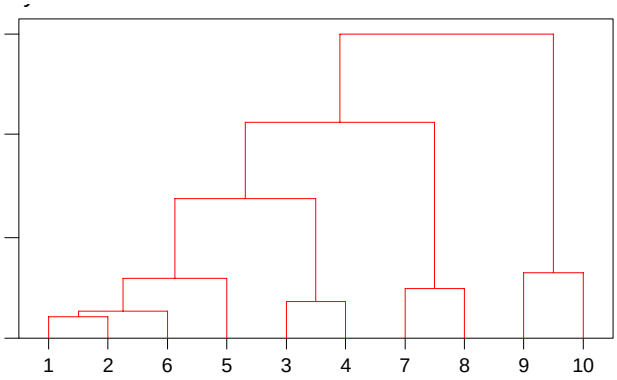
## Multivariate data analysis with the generated features

The features constitute a 10x8 matrix. Principal component analysis (PCA) is a widely used method in chemometrics to visualize such data. The score plot shows a tendency of forming clusters for the five substance classes (software for multivariate data analysis is not contained in MassFeatGen). Other features may give another clustering.

An alternative method to PCA in this example is hierarchical cluster analysis. A dendrogram obtained with this method shows a similar clustering as obtained by PCA.




PCA score plot of 10 mass spectra represented by 8 spectral features (auto-scaled). Variances preserved by first (PC1) and second (PC2) principal component are 55.4 and 18.9 % of total variance, respectively. Each point corresponds to a compound; clustering of substance classes has been manually indicated by ellipses.




Dendrogram from hierarchical cluster analysis (Euclidean distance of original features, agglomerative and complete linkage) of 10 mass spectra represented by 8 spectral features.

## More information

Kurt Varmuza
Laboratory for Chemometrics
c/o Institute of Chemical Engineering, Vienna University of Technology,
Getreidemarkt 9/166-2, A-1060 Vienna, Austria
kvarmuza@email.tuwien.ac.at, www.lcm.tuwien.ac.at
Fax: +431-58801-16091, voice: +431-58801-16060