Chemometrics

- The application of statistics and mathematical methods of chemistry.
- Recognized as a branch of Analytical Chemistry.

Origins

Basic methods were originally developed in the fields of economics and psychology.

These fields are commonly faced with complex, interrelated data sets.

Examples.

Prediction of economic trends based on various 'indicators.'

Measurement of intelligence.

Why Chemometrics?

Most analytical procedures attempt to make a problem univariate.

- Look at only a single unknown material.
- Mask the presence of other materials.
- Remove potential interference.
- Hold all experimental conditions constant except for analyte.

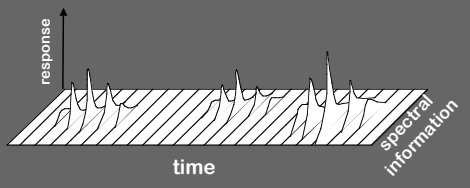
Why Chemometrics?

Some problems are complex by nature.

- The composition of many components contribute to a materials overall properties.
- Some instruments produce a huge number of measurements per sample.

GC/MS example.

- A typical run may produce 2000 spectra with a mass range of 500 m/e.
- This results in 1,000,000 data points for a single sample.
- Too much information to take in at once.



Chemometrics

- Chemometrics has been considered a branch of analytical chemistry since the mid 1970's.
- The introduction of personal computers is one key factor for its increased use.
- Over 12,000 references that contained the word or concept 'chemometrics' were found during a recent SciFinder search.

Chemometrics

It is not a single tool but a range of methods including:

Basic Statistics	Resolution
Signal Processing	Detection
Method Optimization	Pattern Recognition
Factor Analysis	Library Searching
Factorial Design	Neural Networks

Chemometrics

New or modified approaches are introduced very rapidly.

We will attempt to cover most procedures that have found significant use in analytical chemistry.

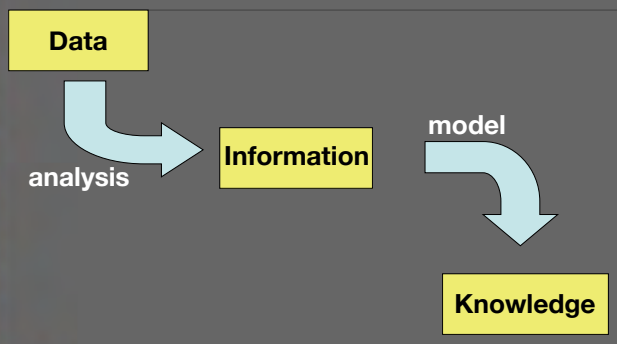
What are data?

Defined as things known or assumed from facts and figures from conclusions.

Broadly defined, data are simply raw information - both qualitative and quantitative.

Raw data are meaningless. We need some form of analysis and model to gain knowledge.

Converting data to knowledge



Kinds of data

Soft

labels, descriptors, category assignments (qualitative) - the water is "hot"

Hard

numerical (quantitative) - the water is 400 K

Avoid numbers that are based on soft assignments.

Types of data

Natural

Arise from natural phenomena.
These are factors we can't control.
Example: it was raining that day or the temperature was 94 ^{O}F when we did a study.

Experimental

Measurement of a property under known, controlled conditions. These are 'laboratory' conditions.

Commonly, both types are involved in a study.

Types of data

Discrete data

Only a finite possible range and interval is possible.

Continuous

Occur over a range but are not discrete. Limits of an instrument may give the appearance of being discrete

Types of data

Classify each as continuous vs. discrete, hard vs. soft and natural vs. experimental.

Score on a multiple choice exam
pH of Lake Erie
Weather conditions
Color of a flower
Current time

Obtaining meaningful data

If your data is bad, nothing can save it.

To collect good data, you must have a plan

The first step should be to ask a few questions.

- What is the desired outcome?
- What is the population?
- What are the parameters?
- What do we already know or can assume?
- What is the basic nature of the problem? research, monitoring, conformance ...

Other factors to consider

Temporal nature of the problem

Long range, short range, one-time

Spacial nature

Global, limited area, local

Other related factors

Most are common sense. However, it requires that you look at the entire problem before starting any work.

Tools we'll use

When dealing with large data sets, modern computer equipment and software is a great asset.

Primary software will be MS Excel (with the data analysis add-on) and XLStat (another Excel add-on). Other tools will be introduced when needed.