Chemometrics Program

by John Nocerino

1 Goals & Approach

Goals: Research and Development to enhance or improve our understanding of experiment and measurement processes to *reduce data uncertainty* through existing or developed chemometric techniques.

Approach: Demonstrate the effectiveness of existing chemometric tools with real environmental projects.

Develop new chemometric techniques and software as needed. Educate scientists to use chemometric tools through video tapes, invited lecturers, and developed in-house courses.

Transfer technology through reports, journal articles, presentations at professional meetings, and sponsored symposia.

2 Chemometric Program Future Plans

Gy Sampling Theory for the Representative Sampling of Heterogeneous Particulates (e.g., Soils)

- Run experiments to test Gy theory and its limitations for various matrix and contaminant combinations.
- Develop and test "correct" sampling devices based on Gy theory.
- Continue with laboratory subsampling studies and develop guidance.
- Simulate field subsampling at the field facility and develop guidance.

Robust Statistics to Reduce Data Uncertainty (Develop Scout Modules)

- Study the behavior of the Mahalanobis distance, or other distance measures, for other underlying distributions (*e.g.*, log-normal).
- Develop and study other influence functions.
- Develop robust outlier tests for parallel axes.
- Develop robust censored (truncated) data analysis.
- Develop robust principal component analysis.
- Develop robust regression analysis.
- Develop robust semi-variograms and kriging.
- Develop robust regression.

Figure 1

Some Chemometric Tools Available or Developed to Reduce Data Uncertainty in Experiment or Measurement Processes

Experiment/Measurement Process	Chemometric Solutions
Sampling	Classical, Geostatistics, Pierre Gy, Bayesian
Measurement/Experiment	Experimental Design and Optimization
Measurement QC	Regression Analysis, SQC, Multivariate Cusum charts, Optimal Frequencies, Multivariate Calibration
Data Generation	Exploratory Data Analysis, Outlier Tests
Remediation	LPD, Ruggedness Tests, Validation Tests

Figure 2

Sampling - Problems

❶ Largest variability ❸ Least defined

❷ Least amount of expertise devoted ❹ What is represented?

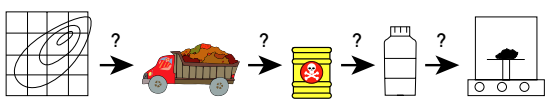


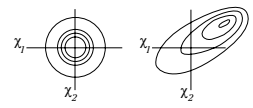
Figure 3

Measurement/Experiment - Problems

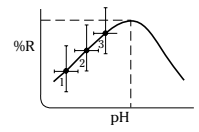
❶ Need training

❷ Need better up-front statistical design to:

ⓐ avoid the one-at-a-time approach;



ⓑ be at the optimal response;



ⓒ look at the important variables.

$y = f(x_1, x_2, x_3, \ldots, x_n)$

3 Scout Statistical Software

File Management

- Merge Files that have Different Variables but Possibly Common Observation Labels
- Appends Files that have Different Observations but the Same (or same number of) Variables

Data Management

- Spreadsheet Edit Features
- Summary Statistics
- Histograms
- Multinormality Tests (Kolmogorov-Smirnov; Anderson-Darling)
- Transform Features
- Prints Data

Outliers (Classical Methods)

- Custom Select Variables
- Generalized Distance Multivariate Outlier Test
- Mardia's Kurtosis Multivariate Outlier Test
- Causal Variables for the Outliers
- Associated (Group) Causes for the Outliers
- Flags Outliers

Robust Methods

- Custom Select Variables
- Univariate Classical and Robust (Huber or PROP Influence Functions, MVT) Statistics
- Robust Analysis (Q-Q Plots, Outliers, Multinormality, Intervals, Contour Ellipses, Discriminant Analysis, PCA, Add or Subtract Specified Means to Data)
- Causal Variables for the Outliers

Principal Component Analysis

- Custom Select Variables
- Based on Covariance or Correlation Matrix
- Displays Matrix
- Computes Components with or without Outliers
- Displays PC Eigenvalues and Variable Loadings
- Transforms Data into Component Scores

Graphics

- Custom Features
- Two-Dimensional Graphics:
  - Correlation Structure Map
  - Bivariate Scatter (Correlation) Plots
- Three-Dimensional Graphics

System

- User's Guide
- System Information
- DOS Shell
- Printer Specifications

017CMB02DB-Nocerino