



# PIROUETTE®

MACHINE LEARNING AND  
COMPREHENSIVE CHEMOMETRICS  
MODELING SOFTWARE



---

BY INFOMETRIX, INC.

# PIROUETTE

## Machine Learning Chemometrics

Infometrix designed and developed the Pirouette software over three decades ago to address the need for a tool to organize and understand complex data. We focused primarily on analytical instruments, but soon found uses in areas as diverse as clinical applications, environmental monitoring, refinery quality control, and database management. The software balances the necessary algorithmic requirements with a simple user interface and is used for:

- Data mining, visualization, and organization;
- Classification analysis to identify the origin of a sample;



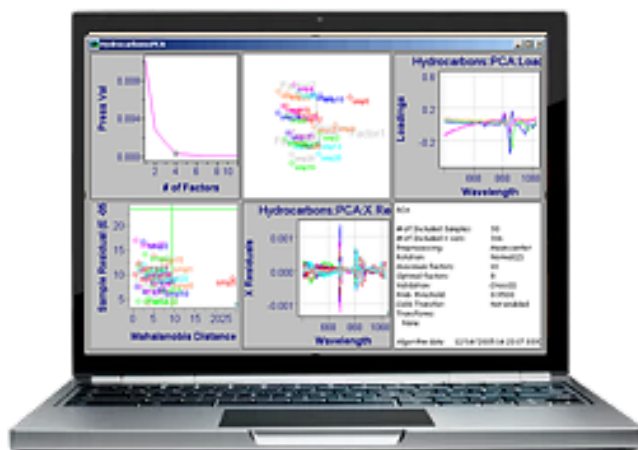
## State-of-the-art interactive software for performing and viewing multivariate analyses

*Pirouette® has its roots in the first broadly available machine learning platform focused on the chemical, bio, food, and consumer products industry. The ARTHUR™ package was released commercially in 1978, evolved into EinSight™ with the introduction of personal computers, leading to Pirouette as the flagship product of Infometrix®.*

- Quantitative analysis to estimate a concentration or other property that is expensive to measure more directly; and
- Unmixing signals to identify and apportion the component parts of a mixture.

Pirouette 5.0, our most comprehensive chemometrics package, is designed specifically for Windows platforms. A simple to use yet very powerful interface facilitates interacting with raw and processed data. Support for many common instrument and data exchange file formats make importing data painless. Thousands of subsets can be created from a single data file, allowing the user to exercise many different what-if scenarios without having to collect and manage multiple data sets. All calculated products are saved in a single file and can be retrieved and manipulated via the Object Manager, a unique data handling system. Transfer of calibration options allow spectra to be adjusted appropriately for prediction with a model from another source. You can even run multiple instances of Pirouette to maximize efficiency.

Even the Pirouette demo has substantial capability. Load any supported file type, including Pirouette native format files, with results present. Use the visualization tools to evaluate your data or data sent to you from a Pirouette user. When you are ready to jump into multivariate analysis, purchase a license to unlock Pirouette's algorithms.





# PRODUCT FOCUS

## *Design and Functions*

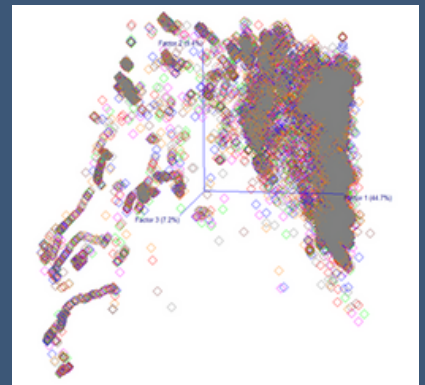
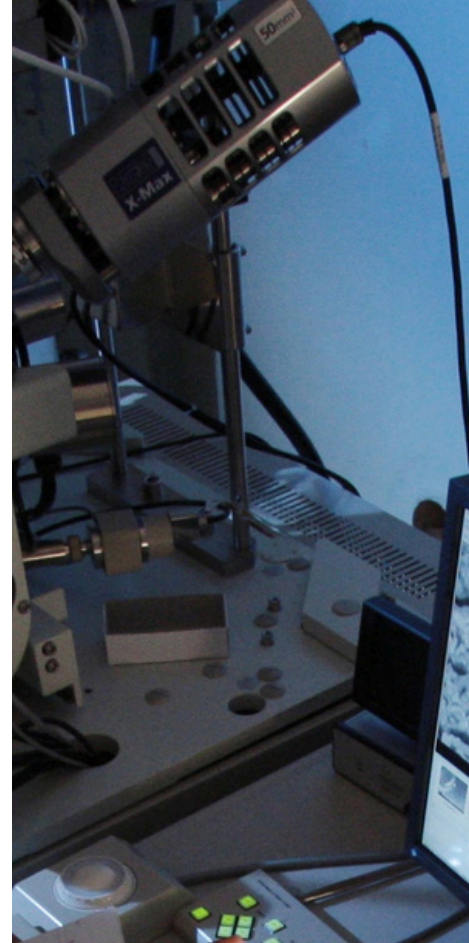
Pirouette was designed with several functions in mind. It is intended as a general guide to visualization techniques, particularly suited to analytical data. From this graphics base, the goals of the analyst fall into one of three broad categories:

- 1.Exploratory Data Analysis;
- 2.Qualitative Analysis; or
- 3.Quantitative Analysis.

## Exploratory Data Analysis

In many projects, the first step is to organize the data and look for similarities and trends that enable understanding of a system or process. Pirouette focuses on principal component analysis (PCA) and hierarchical cluster analysis (HCA) as complimentary means of organizing data based on their sample associations. The fact that the graphics representations are linked speeds the pattern recognition process. Understanding may be the goal, or you may look to develop quantitative or qualitative predictive models.

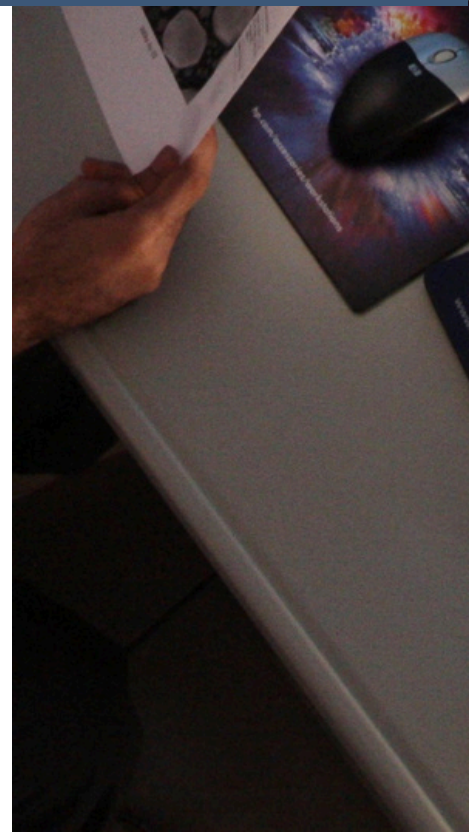
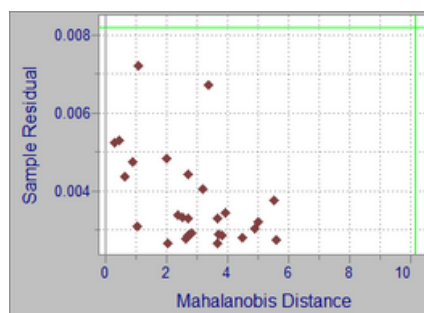
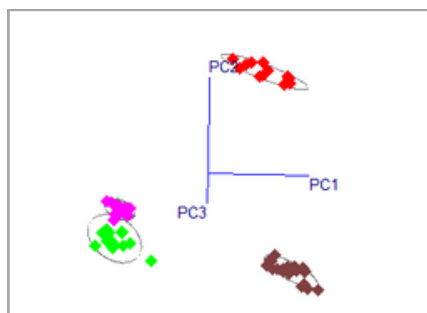
An example of multivariate exploration is shown from the chemical processing industry. Here, 62,000 time points were samples amassing 36 different measurements for each. Shown as a PCA scores plot, the individual points each represent a distinct time and points close to one another means that the pattern of the 36 sensors are similar. This could lead to a model predicting process performance or just remain as a mechanism to understand process stability.



## Qualitative Models

If the purpose of analysis is to identify groupings in a data set, a series of modeling techniques are available in the form of discriminate analysis (PLS-DA), a nearest-neighbor classification (KNN), and a multi-group version of PCA (SIMCA). Use these classification models to identify products that are out of specification and have diagnostics that can explain why they don't match the experience set.

Qualitative models are in wide use for those who want to ensure product quality using all the data and not just a few preselected metrics.

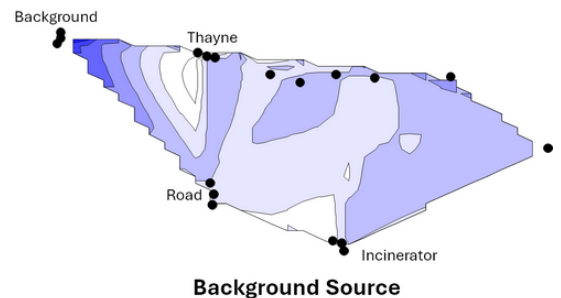
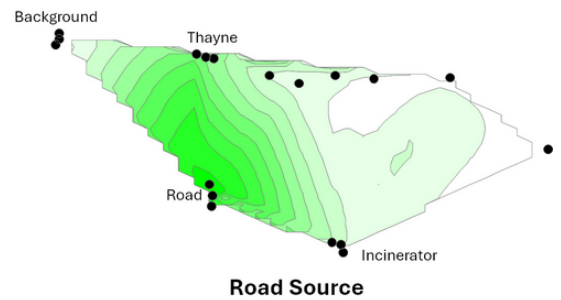
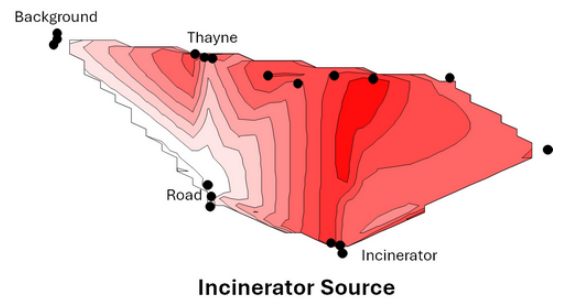


## Quantitative Models

Measuring the concentration of an analyte in the presence of interferences or inferring a performance characteristic of samples is the most common industrial application for both hard modeling techniques like classical least squares (CLS) or soft modeling approaches like principal component regression (PCR) or partial least squares (PLS). These techniques are broadly applicable and their ability to avoid noise in data makes them essential workhorses.

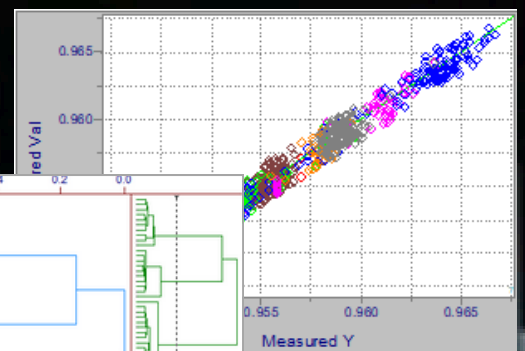
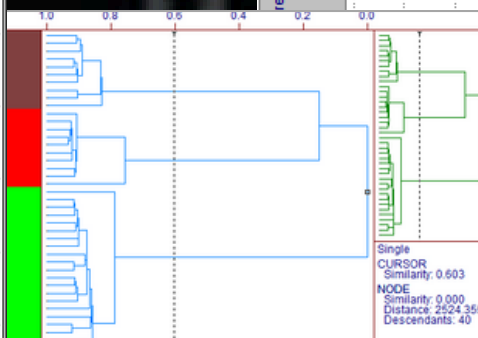
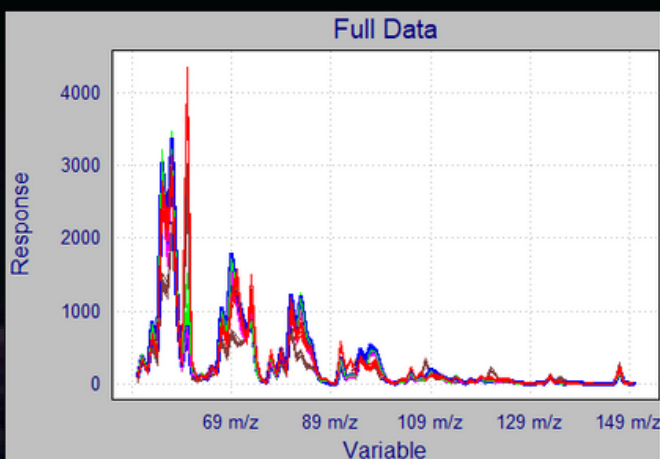
Some quantitative analyses are tied to a bounded set of concentrations; we expect them to add to 100%. In these cases, multivariate curve resolution approaches are appropriate. Pirouette focuses on two model building technologies: multivariate curve resolution (MCR) and alternating least squares (ALS). In chemical production or environmental analysis, these can be used to unmix the influences of multiple sources.

*Pirouette is the most powerful and yet easy to use machine learning processing and display program available. The routines included in this version are broadly applicable and often encountered in current literature. The graphical representations of the data, the data management functions, and the interactive windowing environment are useful and unique.*



Comprehensive, yet easy to use solutions for all levels of expertise and disciplines in product QC/QA

## OPTIMAL DATA VISUALIZATION



# PIROUETTE FEATURES

Features plus latest in Version 5.0

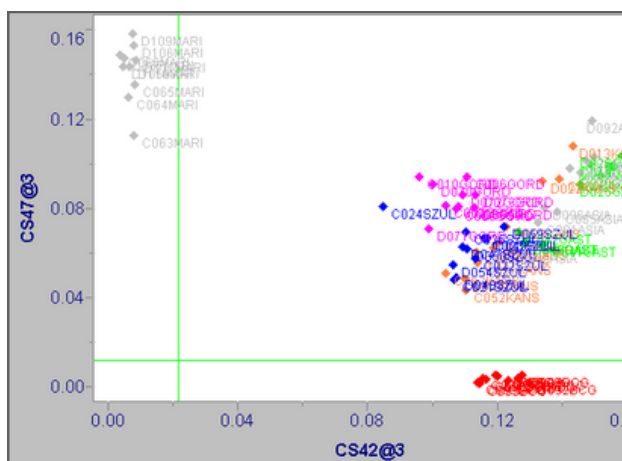
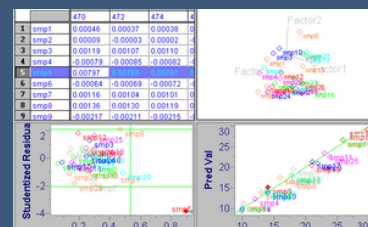
## Rotating 3-Dimensional Scatter Plot Enhances Visualization

Scatter plots are often used to present information about a data set, displaying each sample on a 2-dimensional graphic with one variable on the x-axis and another on the y-axis. When a data set has more than 2 variables, the information from the 3rd variable is not visible in a 2-dimensional plot. A 3-dimensional scatter plot can offer more information in a single view and is the default plot format for most of Pirouette's data viewing. But, a static 3-dimensional view is only part of the picture. Free rotation of the 3-D view will reveal relationships among samples or variables not visible in a static 2-D plot. And, if there are more than 3 variables, Pirouette® makes it easy to swap which variables get plotted.



## Dynamic Linking Highlights Samples Across Views

Visualization is a key component in multivariate analysis, and Pirouette makes this easy. In particular, with large data sets, it is critical to know how samples appear in relation to others. A simple way to do this is by “highlighting” a sample to make it stand out (also applicable to groups of samples). In Pirouette, we have taken this to the next step that those in dynamic graphics call linking. A highlighted sample will appear highlighted in all sample-oriented views, whether presented as a table or a graphic. Combined with cloaking, this is a powerful tool for investigating relationships in your data.



## Decision Diagrams Facilitate Understanding

Most Pirouette algorithms offer outlier diagnostics to help you optimize the training set and corresponding model. You can view these diagnostics as a table or graphically. Plotting one diagnostic against another allows quick evaluation of outliers by comparing a sample point to the thresholds for each metric. SIMCA also presents the Class Distances as a decision diagram in which the distances of the samples to a class model are shown for 2 classes at a time. In each of these cases, the multiple thresholds divide the plot space into subregions of membership and not.



## Pirouette Specifications Version 5.0

### Minimum System Requirements

3 GHz PC, 12 GB RAM  
100+ MB free on hard disk  
Pointing device  
Windows 7, 8, 10, 11

### Help

Full documentation as cross-referenced PDF  
Link to Adobe® Acrobat® Reader

### Data and Data Treatments

No file size limitations  
Binary, ASCII, Lotus®, Excel®  
Common Instrument Formats, AIA, JCAMP, Galactic®, ChemStation®  
Write: Files, Subsets, Calculated objects in Binary, ASCII, Excel, SPC, and AIA formats  
Models: Binary, ASCII, Galactic, Guided Wave  
Merge:  
Single or Multiple files, drag and drop, by Sample, by Variable  
Subsets  
Unlimited number  
By exclusion or inclusion  
Saved with file, separate results maintained  
Sample and Variable selection  
Kennard & Stone, PCA Hypergrid, Leverage  
Fisher or Variance weights, StDev rank  
Output  
Printers, via Print Manager  
Clipboard, of graphics, data or results  
Edit functions  
Cell contents and ranges  
Columns and/or Rows  
Cut, Copy, Paste, & Clear; Insert & Delete  
Spreadsheet  
X-block, Y-block, & Category-block  
Go To; Sort, by value or by name  
Find and Fill missing values  
Fill by Value, Mean, Median, Interpolation, Zero, PCA fill  
Pretreatments  
Smooth, 1st & 2nd Derivative (5 - 95 points)  
Log10, Multiply, Normalize, Subtract  
Divide by (2-norm, 1-norm, max, range, value)  
Baseline (linear, quadratic, cubic, sample)  
Multiplicative Scatter Correction  
Standard Normal Variate  
Mean-centering, Variance scale  
Autoscale, Pareto, Range scale

### User Interaction

ObjectManager  
Data and Results tree  
Drag and drop into chart windows  
Data object history

Plots  
2D Scatter, 3D Rotatable Scatter  
Line Plot, Multiple 2D Arrays  
Point labels, Cloaking  
Interaction  
Point/Range Selection, Color by category  
3D Spinning, Magnify, Point Labeling  
Linking selections across views  
Preferences  
User defined preference sets  
Custom interface colors, graphics, fonts  
Custom plot symbol size, window size  
English, Spanish, German, Japanese, Portuguese, French, Italian  
Exploratory Data Analysis  
Hierarchical Cluster Analysis  
Single Link, Centroid, Complete, Median, Incremental, Group Average, Flexible  
Orientation by Sample or by Variable  
Sample or Variable Dendrogram  
Principal Components Analysis  
Model Probability Control  
Projection Model  
Cross, Step Validation  
Varimax Raw, Normal, & Weighted  
Scores, Rotated Scores  
Loadings, Rotated Loadings  
Eigenvalues, Rotated Eigenvalues  
Errors, Outlier Diagnostics, Contributions  
Modeling Power  
X Residuals & Reconstructed  
Dynamic factor selector in Prediction  
Projected Scores

### Multivariate Classification

K Nearest Neighbors  
Unlimited number of neighbors or classes  
Classification Model  
Votes Matrix, Misses Vector  
Misclassification Matrix  
Predicted Class, Class fit  
Soft Independent Modeling of Class Analogy  
Same as above plus  
Model & Prediction Probability control

Scores, Loadings, Eigenvalues, Residuals  
Interclass Residual and Distance  
Discrimination Power & Modeling Power  
Class Projections, Distances, Probabilities  
Best & Next Best Predicted Class  
PLS Discriminant Analysis  
Regression Tool Control  
Class Predicted, Misclassifications

### Multivariate Regression

Principal Components Regression, Partial Least Squares Regression, Classical Least Squares, Locally Weighted Regression  
Unlimited number of dependent variables  
Cross, Step, by Category Validation  
Any number of left out samples  
Orthogonal Signal Correction  
Scores, Loadings, Eigenvalues  
Errors (PRESS, SEC, SEV, r)  
Y Fit, Outlier Diagnostics, Contributions  
X Residuals, X Reconstructed  
Correlation spectrum, Regression Vector  
Dynamic prediction factor selector  
Predicted properties, Prediction scores  
Outlier diagnostics, Contributions  
Pures and uncertainty bounds (CLS)  
Mixture Analysis  
Multivariate Curve Resolution, Alternating Least Squares  
Eigenvalues, Scores, Loadings  
Solution Select, Feasible Region, Source Amounts & Profiles, X Residuals  
Prediction Model  
Feasible Regions & Source Amounts

### Calibration Transfer

Algorithms supported:  
KNN, SIMCA, PLS, PCR  
Direct standardization, Piecewise direct, Additive, Multiplicative



**THE PREMIER CHEMOMETRICS COMPANY**

## **CONTACT US**

**For evaluation, quotation, presentation or additional information**

---

Infometrix, Inc.  
19119 North Creek Pkwy  
Suite 100  
Bothell, WA 98011

T: 425-402-1450  
[www.infometrix.com](http://www.infometrix.com)  
[info@infometrix.com](mailto:info@infometrix.com)