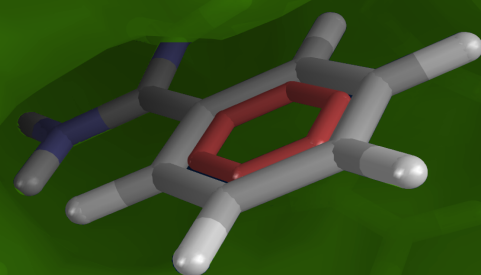
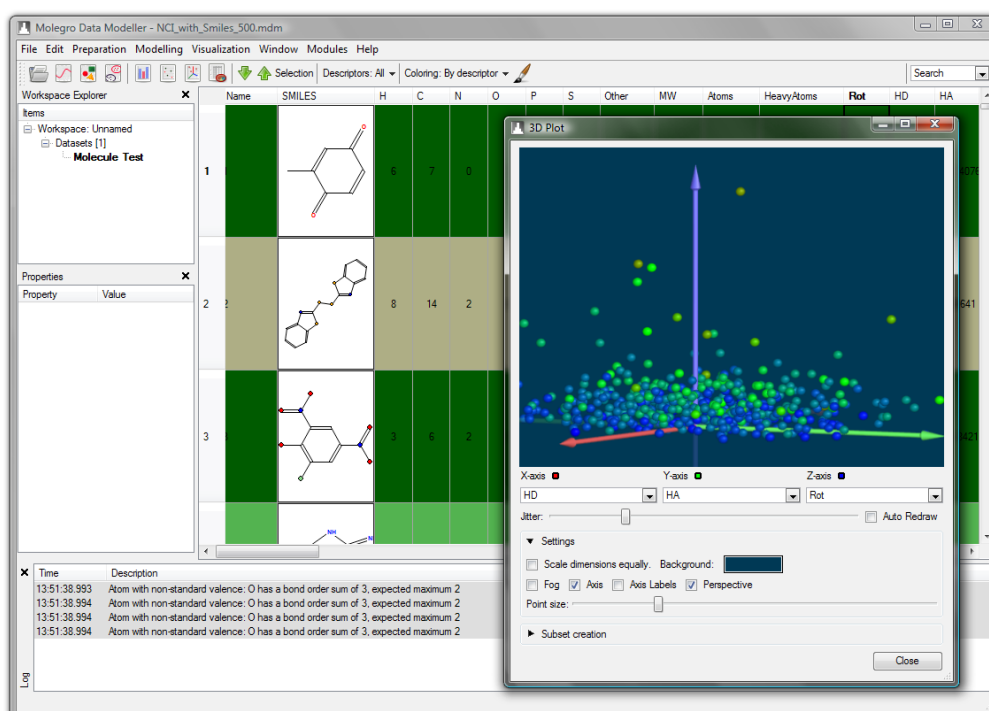


Molegro Data Modeller

Data modelling, exploration, and visualization



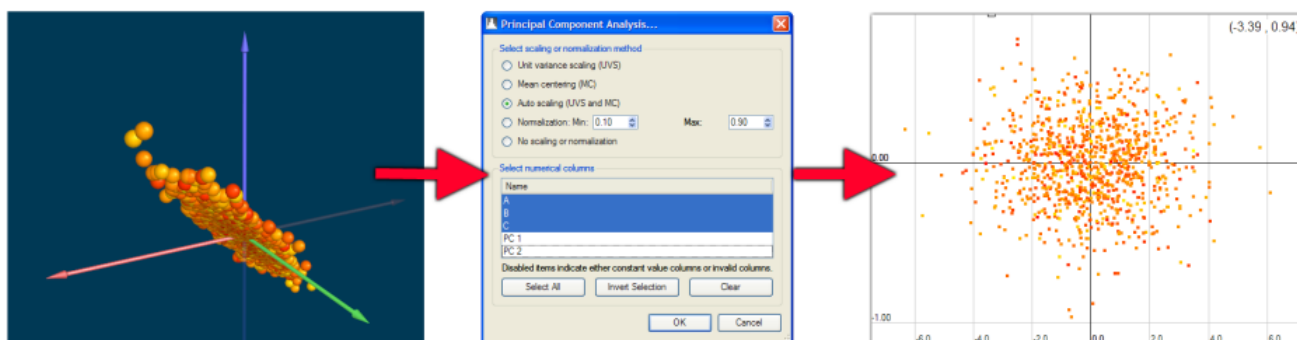
Molegro Data Modeller is a cross-platform application for data mining, data modelling, and data visualization.



Create regression and classification models using partial least squares, neural networks, multiple linear regression, or support vector machines. Models are saved together with relevant normalization information making them easy to apply to new datasets. Cross-validation and feature selection can be applied using the built-in wizards. Automatic search for optimal model parameter settings.

Molegro Data Modeller Overview

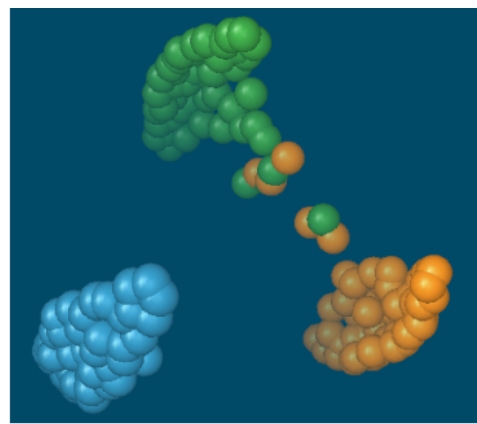
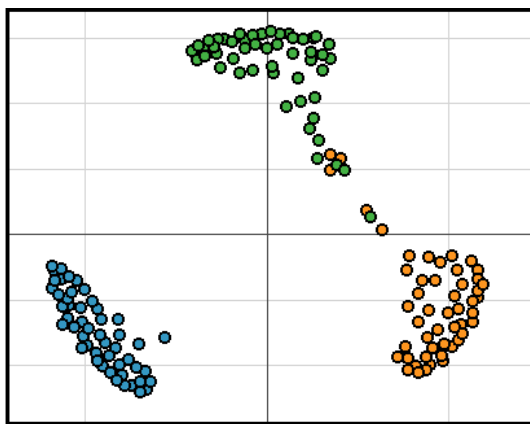
- Data preparation: repair missing values, outlier detection, principal component analysis
- Data modelling: regression, classification, and clustering
- Chemistry: molecule depictions and SMILES/SDF support
- Data transformations
- Data visualization: histograms, 2D/3D plots, project multi-dimensional data onto 2D/3D
- Support for KNIME integration
- Cross-platform: Windows, Mac, and Linux



Data Preparation and Transformation

- Repair missing values
- Outlier detection (distribution or density based)
- Algebraic data transformations
- Chemical data import (SMILES and SDF) and molecule depictions
- Subset creation using grids or random selection
- Principal component analysis
- Data scrambling and randomization
- Correlation matrix: overview zoom, create 2D/3D projections, prune correlated descriptors

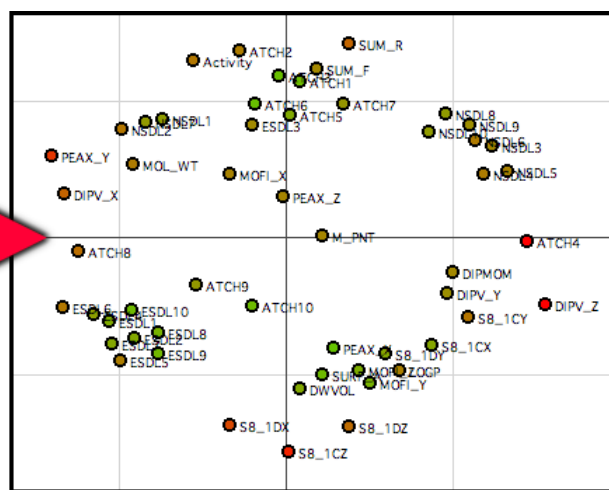
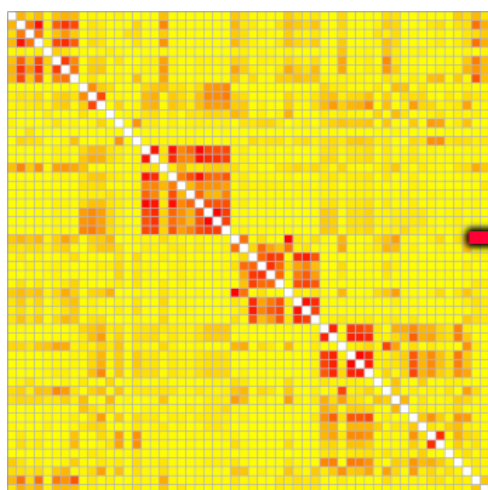
Molecule depictions in Molegro Data Modeller.



Example of Fuzzy Classification. Some data points are located outside the main clusters of the classes.

Data Modelling and Analysis

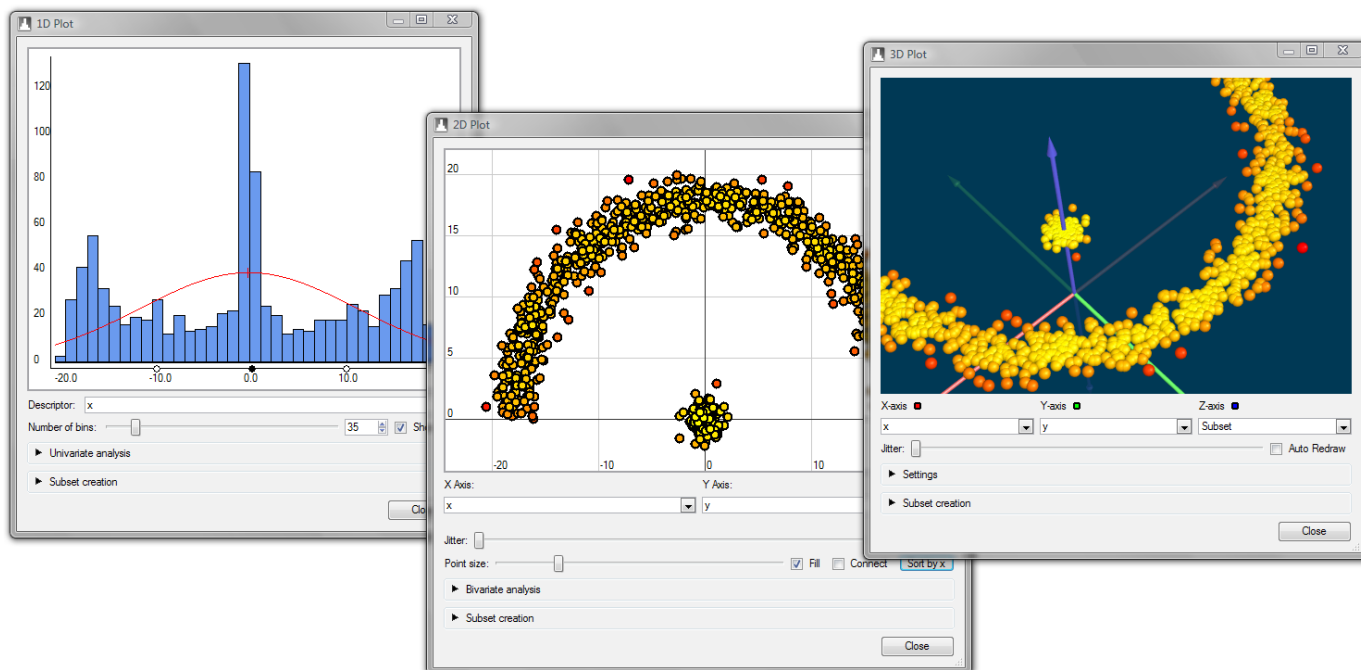
- Regression: partial least squares, neural networks, multiple linear regression, and support vector machines
- Classification: K-nearest neighbors and support vector machine classification
- Clustering: K-means, density-, and threshold-based clustering
- Cross-validation and leave-one-out validation is easy to apply.
- Statistical measures: mean, median, standard deviation, skewness, kurtosis, Pearson and Spearman rank correlation, correlation matrix, confusion matrix, F-measure, Bayesian information criterion (for model selection)
- Feature selection is easy to set up and automate. Multiple feature selection schemes and model selection measures are available.
- Automatic search for optimal parameter settings (for SVM, PLS, KNN, and NN)



A correlation matrix projected onto two dimensions using high-dimensional visualization: points that are close to each other correspond to descriptors that are highly correlated.

Data Visualization and Exploration

- Histograms
- 2D scatter plots
- Selections in the plots are synchronized with selections in the spreadsheet
- Hardware accelerated 3D plots
- High-dimensional visualization: project data onto 2D or 3D using a spring-mass map model
- Gnuplot and bitmap export
- Jitter and custom coloring
- Molecule depictions in scatter plots



Supported Platforms

- Windows 7, Vista, XP, and 2003
- Mac OS X 10.5 or later (Intel and PowerPC)
- Linux: Most major distributions including Fedora Core, Red Hat, and Ubuntu (both 32 and 64 bit)

Evaluation Copies

- To get a free 30 day trial license visit: www.molegro.com/trial.php

Additional Information

- www.molegro.com
- info@molegro.com

molegro

computational drug discovery

Molegro develops cheminformatics software for drug discovery and data mining.

We focus on combining state-of-the-art algorithms with an intuitive graphical user interface experience.

Through collaboration with both academia and industry we ensure our products are continuously evolving.

C. F. Møllers Allé 8
Building 1110
DK-8000 Århus C
Denmark

Phone: (+45) 87 15 55 69
Fax: (+45) 87 15 41 02
www.molegro.com