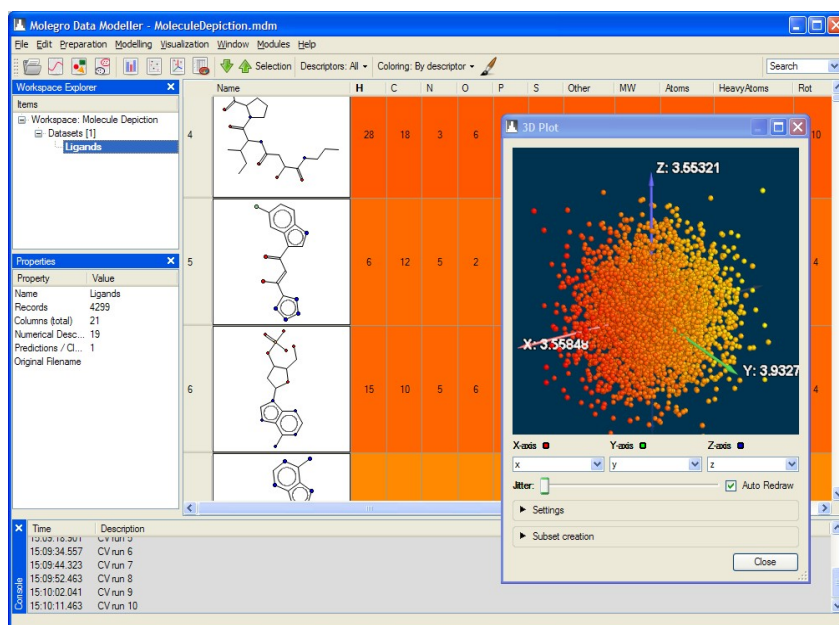


Molegro Data Modeller

Data modelling, exploration, and visualization

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

The highly interactive user interface is ideal for fast and intuitive data exploration, as opposed to complex workflow based solutions or command-driven statistical products.

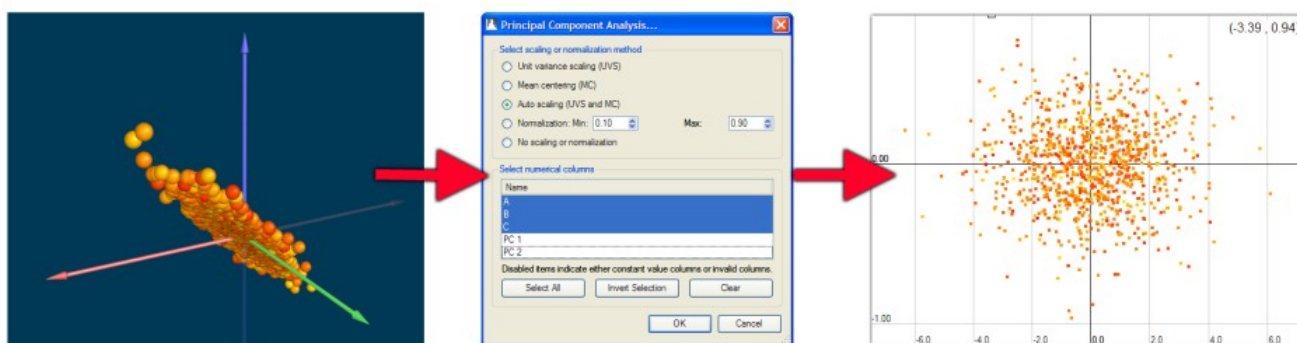


The main user interface of Molegro Data Modeller.

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

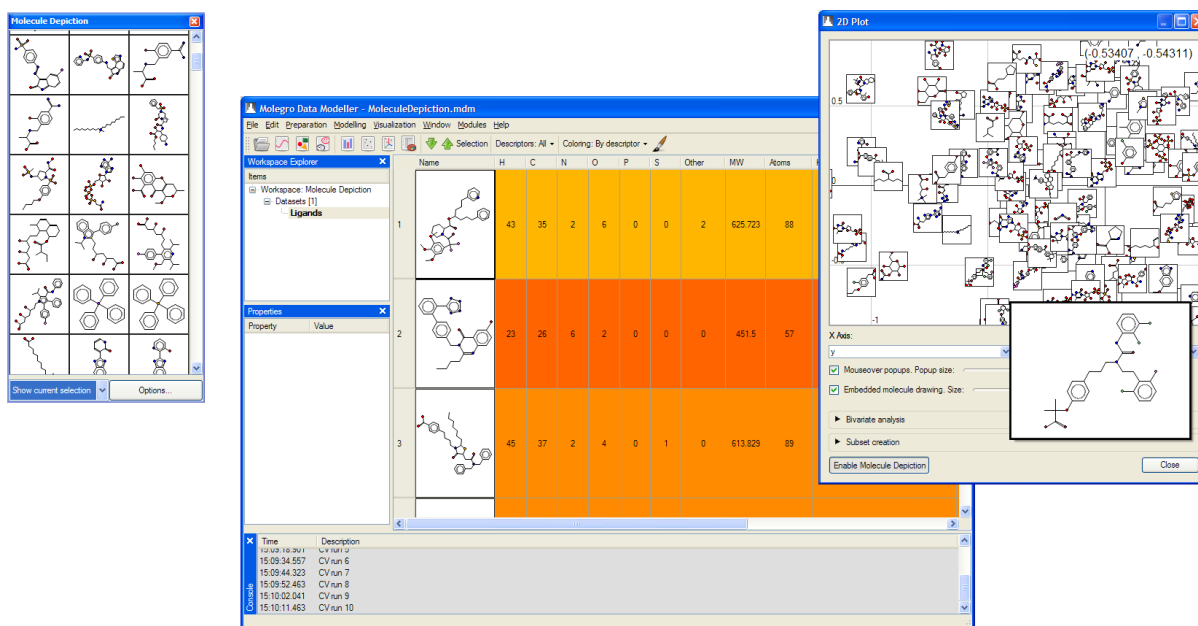
- Data preparation (repair missing values, outlier detection, subset creation, PCA).
- Data modelling: regression, classification, and clustering.
- Chemistry: molecule depictions and SMILES/SDF support.
- Data transformations.
- Data visualization (histograms, 2D/3D plots, project multi-dimensional data onto 2D/3D).
- Cross-platform: Windows, Mac, and Linux.



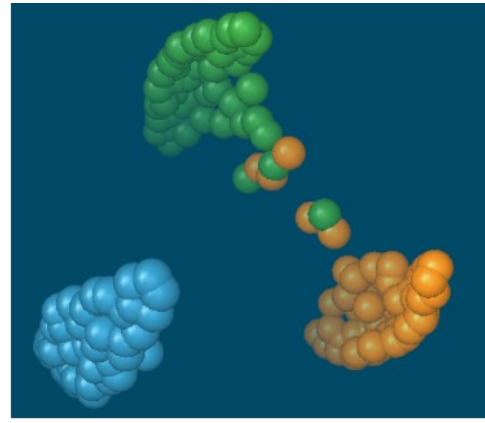
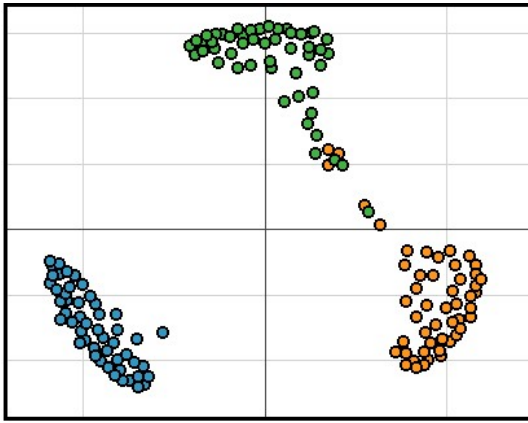
Dimensionality reduction using principal component analysis.

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.



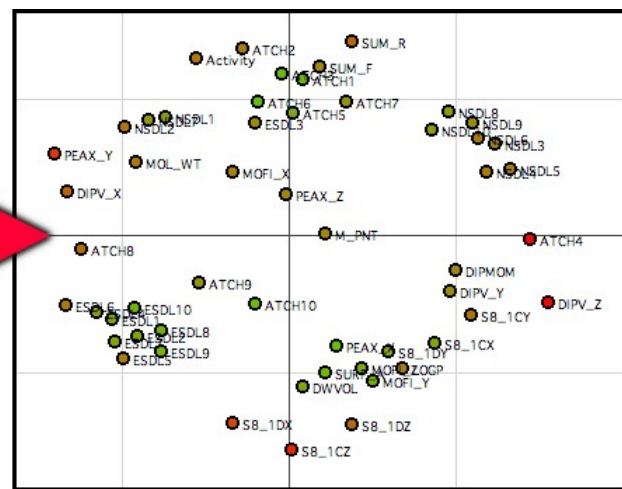
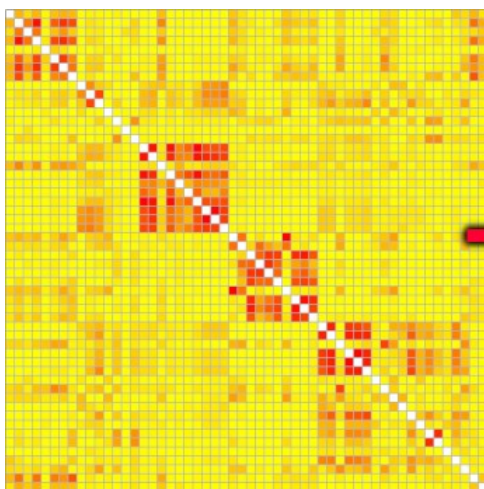
Molecule depictions in Molegro Data Modeller.



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

Data Modelling and Analysis

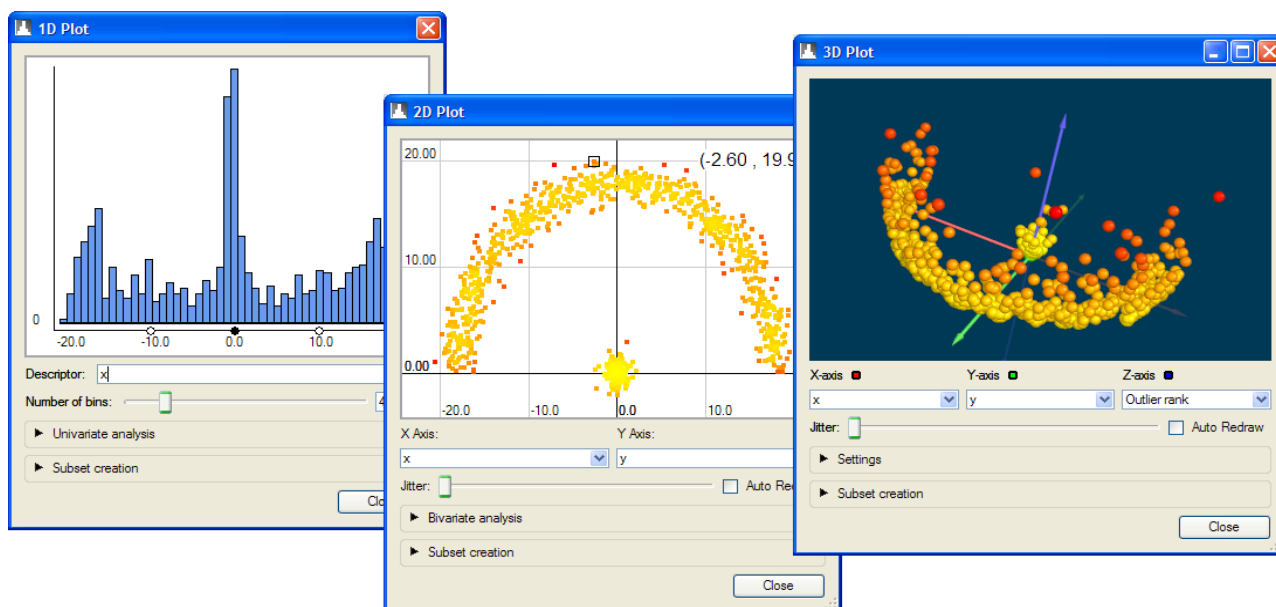
- Regression: partial least squares, neural networks, multiple linear regression, support vector machines.
- Classification: K-nearest neighbors and support vector machine classification.
- Clustering: K-means, density-, and threshold-based clustering.
- Cross-validation and leave-one-out validation is easy to apply. It is possible to create the folds manually or by random selection.
- 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 columns 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 10.4.0 or later (Intel and PowerPC).
- Linux: Most major distributions including Fedora Core, Red Hat, and Ubuntu. Both 32 and 64 bit versions are available.

Evaluation Copies

- To request a 30 day trial version visit: www.molegro.com/products.php

Additional Information

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

About Molegro

Molegro develops bioinformatics 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.

Molegro

C. F. Moellers Alle 8, Building 1110
DK-8000 Aarhus C
Denmark

Phone: (+45) 89 42 31 65
Fax: (+45) 89 42 30 77