

Orthogonal PLS (OPLS)

Separating predictive from non-predictive variation in QSAR, predictive toxicology and “Omics”-data analysis

Orthogonal PLS, OPLS, is a recent modification of the PLS method [1]. In comparison with PLS, OPLS produces models which are more transparent and therefore easier to interpret.

Enhanced discrimination

The objective of OPLS is to divide the systematic variation in the X-block into two model parts, one part which models the co-variation between X and Y, and another part which expresses the X-variation that is not related (orthogonal) to Y. This may lead to better class-resolution in a discriminant problem (Figure 1).

Lennart Eriksson¹, Johan Trygg² and Svante Wold^{1,2}

1) Umetrics AB, Umeå, Sweden, 2) Umeå University, Umeå, Sweden

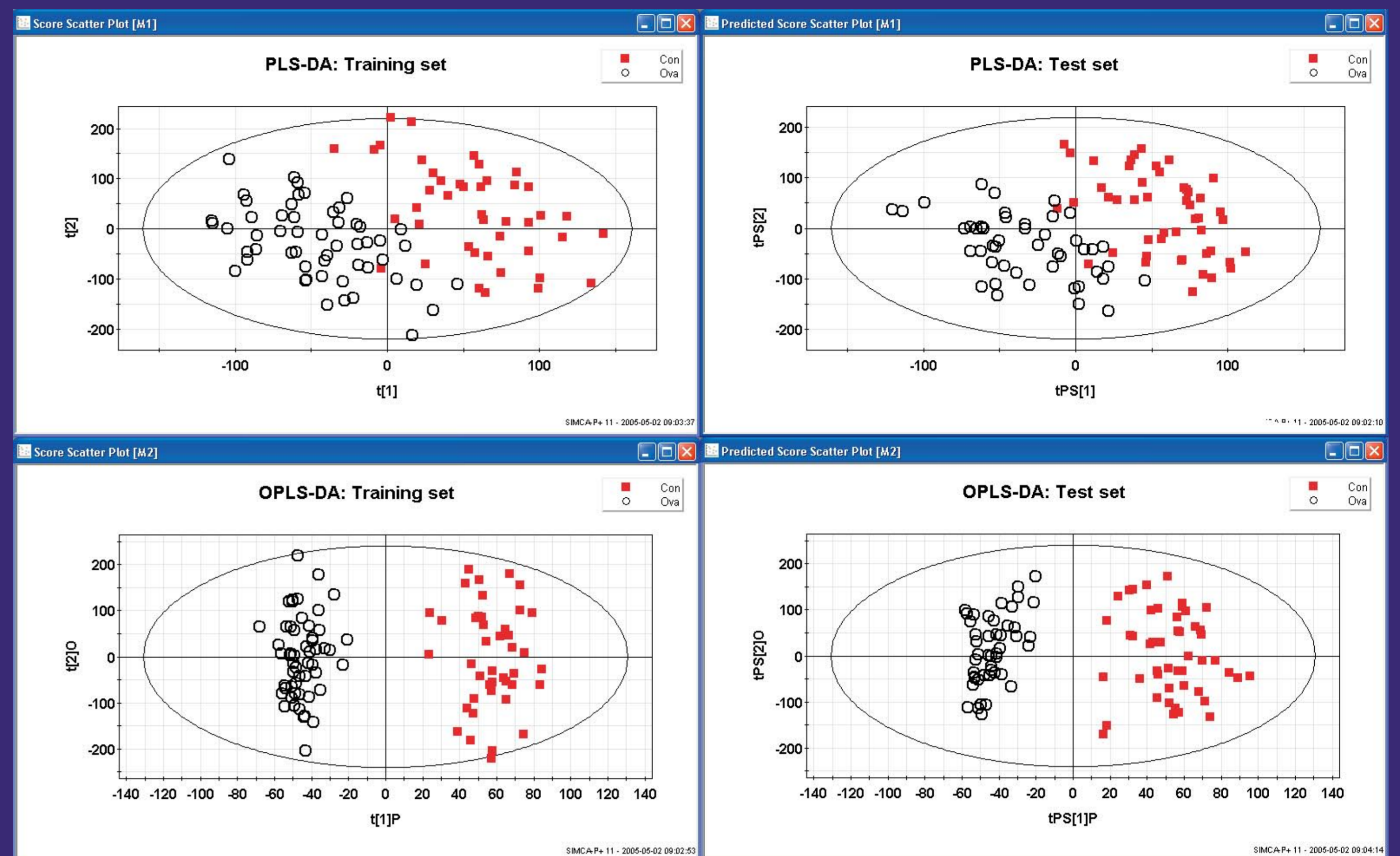


Figure 1: Ovarian cancer data set [3,4]. Proteomic profiles of 100 cancer patients and 92 unaffected women (controls) were analyzed using PLS-DA and OPLS-DA. The training set was identified through multivariate design. OPLS-DA enhances class separation.

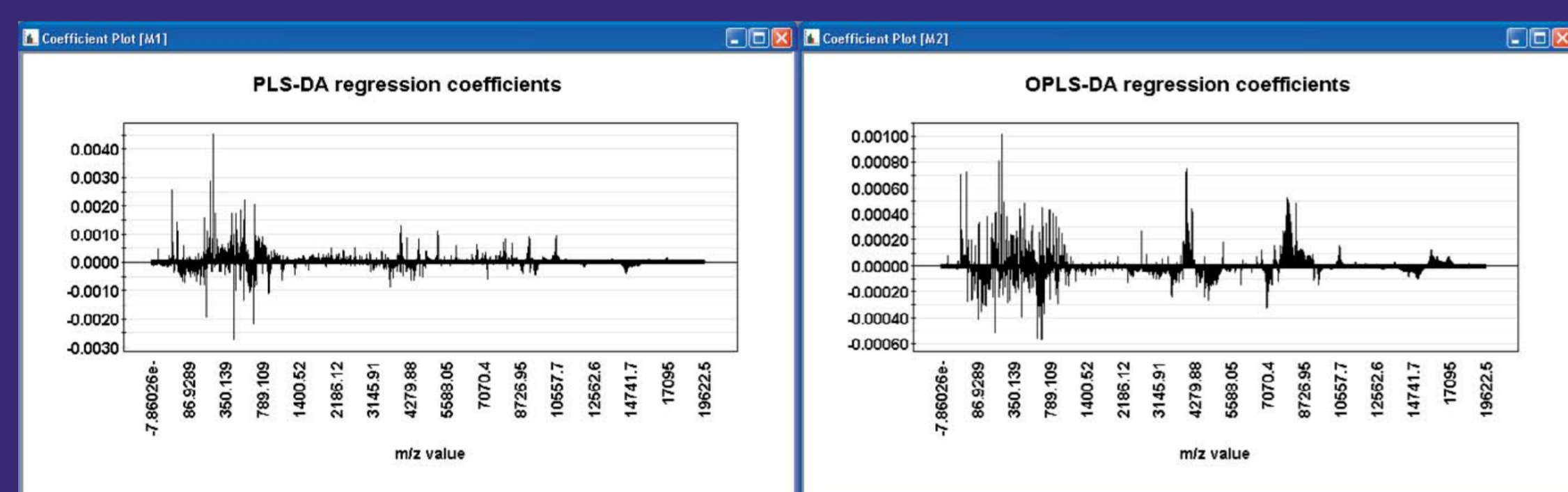


Figure 2: PLS-DA and OPLS-DA regression coefficients. Each proteomic profile consists of 15,154 variables (m/z values obtained from SELDI-TOF mass spectrometry).

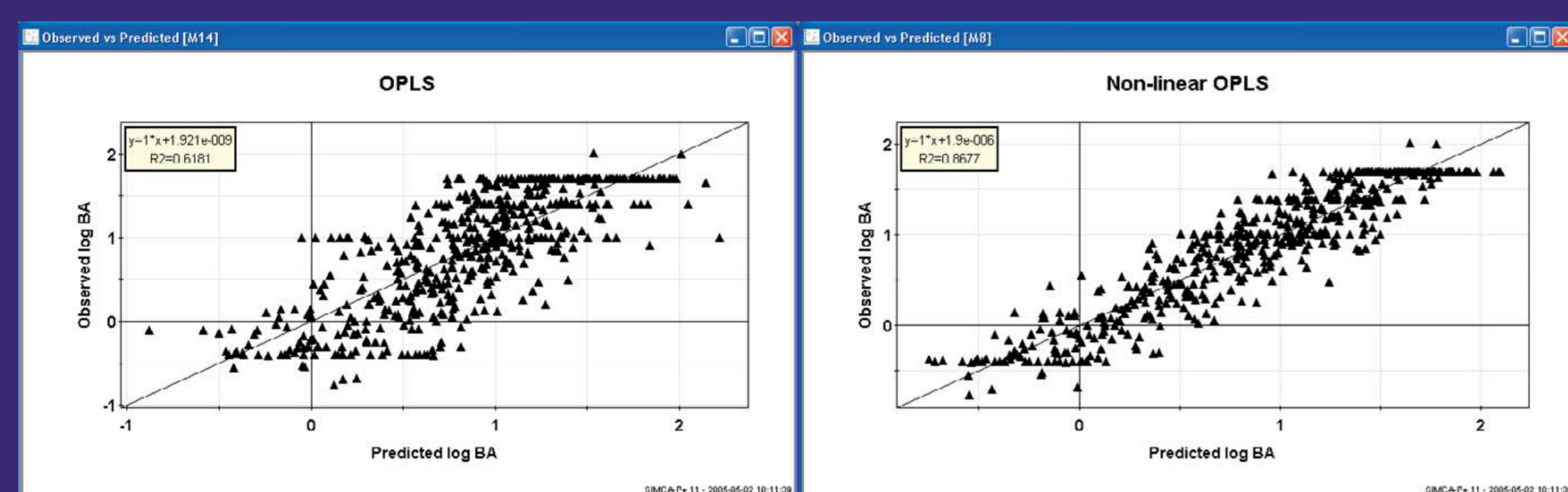


Figure 3: Relationship between observed and predicted log biological activity in an ADME investigation. In this case, the non-linear OPLS method leads to a stabilized QSAR model.

Improved interpretation

OPLS leads to a more straight-forward and realistic model interpretation as it is possible to focus on the variation in X that really correlates with Y (Figure 2).

Analysis of the non-correlating variation

OPLS gives an opportunity to analyze the orthogonal variation in the X-data that does not correlate with Y, and understand what the sources are for that.

Extension of OPLS

Non-linear and hierarchical extensions of OPLS may lead to stabilized QSAR models in complex problems in pharmaceutical industry (Figure 3).

Discussion

The OPLS method does not change predictive power, but its great asset lies in making model interpretation much easier [2]. Additionally, the number of predictive PLS components becomes smaller. In the case of a single-Y problem, OPLS assembles the predictive X-variation in the first component, and, when dealing with multi-Y problems, in the first few predictive components.

Software from Umetrics



SIMCA-P – The standard in Multivariate Data Analysis



MODDE – Software for Design of Experiments and Optimisation

References

- [1] Trygg and Wold, J. Chemom, 2002, 16:119-128
- [2] Eriksson, et al., Multi- and megavariate data analysis – Principles and Applications, www.umetrics.com
- [3] <http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp>
- [4] Whelehan O., et al., Detection of Ovarian Cancer using Chemometric Analysis of Proteomic Profiles, Manuscript in preparation



www.umetrics.com