

## **Abstract**

*J. Chemometrics*, **24**, 111-120 (2010)

Liebmann B., Filzmoser P., Varmuza K.:

### **Robust and classical PLS regression compared**

Classical PLS regression is a well-established technique in multivariate data analysis. Since classical PLS is known to be severely affected by the presence of outliers in the data or deviations from normality, several PLS regression methods with robust behavior towards data contamination have been proposed.

We compare the performance of the classical SIMPLS approach with the partial robust M regression (PRM).

Both methods are applied to three different data sets including outliers intentionally created. A simulated data set with known true model parameters allows insight in the modeling performance with increasing data contamination. QSPR data are modified with a cluster of outlying observations. A third data set from near infrared (NIR) spectroscopy is likely to include noise and experimental errors already in the original variables, and is further contaminated with outliers.

To provide a sound comparison of the considered methods we apply repeated double cross validation. This validation procedure judiciously optimizes the model complexity (number of PLS components) and estimates the models' prediction performance based on test-set predicted errors.

All studied robust regression models outperform the classical PLS models when outlying observations are present in the data. For uncontaminated data, the prediction performances of both the classical and the robust models are in the same range.