

## **Abstract**

*Chemom. Intell. Lab. Syst.*, 117, 31-44 (2012)

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

### **Redundancy analysis for characterizing the correlation between groups of variables - Applied to molecular descriptors.**

Redundancy analysis (RA) estimates the extent of linear relationships between blocks of variables that are given for a set of objects (samples). RA has only rarely been used in chemometrics. Basic principles and limits of RA are discussed, and RA is briefly compared with canonical correlation analysis (CCA) and partial least-squares (PLS2) regression.

The significance of a redundancy index is estimated by permutation tests. For PLS2, an index determining the similarity of variable blocks can be derived that is equivalent to the canonical measure of correlation, CMC.

RA is applied to a set of 3708 molecular descriptors (created by software Dragon) for 6458 chemical structures (AMES database). The 27 descriptor groups are characterized by their redundancy indices, which allow a comparison of their multivariate information content. The results guide the selection of the most different descriptor groups, which perform better in a discrimination task (classification of mutagenicity) than the entire groups.