

# Partial least squares density modelling (PLS-DM) – an efficient approach to one-class classification

*Paolo Oliveri, Cristina Malegori, Monica Casale*

*DIFAR Department of Pharmacy, University of Genova, Genova, Italy*

In the present work, a novel class-modelling method, called partial least squares density modelling (PLS-DM), is presented. Class-modelling (also referred to as one-class classifiers) is one of the two families in which it is possible to divide the qualitative data modelling and it is the choice to be preferred when the focus is on a single class – a typical case for the verification of authenticity claims.

PLS-DM is based on a partial least squares (PLS) regression, in which a distance-based sample density measurement is used as the response variable. A kernel approach involving a potential-function estimation of probability density is subsequently applied on PLS scores, and is considered, jointly with residual  $Q$  statistics, to develop efficient class models effectively capable to deal with data characterised by non-normal and non-uniform distributions.

The work critically discusses the influence of adjustable model parameters (e.g., the pre-processing, the number of latent variables, and the smoothing coefficient of potential functions) on the resulting performances by means of evaluation of sensitivity and specificity within a cross-validation cycle and by application of the Pareto optimality criterion; moreover, performances of the optimal model are evaluated on an external test set.

The potential of PLS-DM is illustrated presenting ad-hoc case studies related to verification of food authenticity claims, including a critical comparison with well-established class-modelling methods, such as soft independent modelling of class analogy (SIMCA) and unequal dispersed classes (UNEQ).

## Acknowledgments

Financial support by the Italian Ministry of Education, Universities and Research (MIUR) is acknowledged – Research Project SIR 2014 “Advanced strategies in near infrared spectroscopy and multivariate data analysis for food safety and authentication”, RBSI14CJHJ (CUP: D32I15000150008).

