



HAL
open science

Locally-Weighted Partial Least Squares Regression for infrared spectra analysis

Aurélie Thébault, Damien Juery

► **To cite this version:**

Aurélie Thébault, Damien Juery. Locally-Weighted Partial Least Squares Regression for infrared spectra analysis. 1ères Rencontres R, Jul 2012, Bordeaux, France. hal-00717517

HAL Id: hal-00717517

<https://hal.science/hal-00717517>

Submitted on 13 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Locally-Weighted Partial Least Squares Regression for infrared spectra analysis

A. Thébault and D. Juery

UMR MISTEA
INRA Montpellier
2 place Pierre Viala, 34060 Montpellier Cedex 2
Contact author : aurelie.thebault@supagro.inra.fr

Keywords : Partial least square regression, Neighbour selection, Local calibration

Recent developments in infrared spectroscopy offer advantages in terms of speed and simplicity for routine analysis of chemical or biological data. Many chemometric tools have been developed to predict chemical and biological properties of samples from their infrared spectrum. Among them, the Partial Least Squares regression (PLS) is widely used to get information from large spectral databases. However, while PLS is efficient for predicting biological or chemical parameters from homogeneous spectral databases, it performs poorly on heterogeneous databases, leading to the use of local PLS. In local PLS, the chemical or biological parameters of an unknown sample are predicted from a subset of selected, spectrally similar, samples from a largest heterogeneous spectral database.

Several R-packages such as **pls** (Mevik et al., 2012), **plspm** (Sanchez and Trinchera, 2012), **plsRglm** (Bertrand et al., 2011), **ChemoSpec** (Hanson, 2012) and **soil.spec** (Terhoeven-Urselmans, 2012) already deal with spectral data analyses but none of them propose local calibration. Based on the LOCAL calibration algorithm (Shenk et al., 1997) and the Locally Weighted Regression (LWR ; Naes et al., 1990), we developed a powerful routine associating local calibration and weighting of the calibration samples for prediction of chemical or biological parameters from large and heterogeneous spectral databases.

The Locally-Weighted PLS routine allows optimization of spectral pre-processing within a large choice of available pre-treatments. The selection of the local calibration samples can be done either through the use of a metrics (Mahalanobis distance or correlation coefficient) or through Dirichlet Process (Neal, 2000). An optimization function of the number of local calibration samples is also available. The locally-weighted PLS routine achieves a trade-off between Locally Weighted Regression (Naes et al., 1990) and LOCAL algorithm (Shenk et al., 1997). Weights can be assigned to the selected calibration samples following the Locally Weighted Regression theory. Since the best number k of PLS components is unknown *a priori*, we adopt a model averaging strategy adapted from the LOCAL algorithm: the final estimator is based on the aggregation of many PLS estimators with an increasing number of components.

The steps of the Locally-Weighted PLS will be briefly presented. In a second time, the routine will be applied on a real dataset to present main functions and options.

References

- [1] Mevik, B.H., Wehrens, R., Liland, K.H. (2012). pls – Partial least squares and principal component regression. R package version 2.3-0.
<http://www.cran.r-object.org/package=pls/>.
- [2] Sanchez, G., Trinchera, L. (2012). pls pm– Partial least squares data analysis methods. R package version 0.2-2.
<http://www.cran.r-object.org/package=plspm/>.

- [3] Bertrand, F., Meyer, N., Maumy-Bertrand, M. (2011). plsRglm – Partial least squares regression for generalized linear models. R package version 0.7-6.
<http://www.cran.r-object.org/package=plsRglm/>.
- [4] Hanson, B.A. (2012). ChemoSpec – Exploratory chemometrics for spectroscopy. R package version 1.50-2.
<http://www.cran.r-object.org/package=ChemoSpec/>.
- [5] Terhoeven-Urselmans, T. (2012). soil.spec – Spectral data exploration and regression functions. R package version 1.4.
<http://www.cran.r-object.org/package=soil.spec/>.
- [6] Shenk, S., Berzaghi, P., Westerhaus, M.O. (1997). Investigation of a LOCAL calibration procedure for near infrared instruments. *J. Near Infrared Spectrosc.* 5, 223-232.
- [7] Naes, T., Isaksson, T., Kowalski, B. (1990). Locally weighted regression and scatter correction for near-infrared reflectance data. *Analytical Chemistry* 62(7), 664-673.
- [8] Neal, R.M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *J. Computational and Graphical Statistics* 9(2), 249-265.