L7. Uncertainty in multivariate calibration: application to embedded NIR data

Juan A. Fernandez Pierna[1], *V. Baeten*[2], *G. Sinnaeve*[2], *P. Dardenne*[2]
[1]*University of Agronomical Sciences, Statistics and Informatics Department, Gembloux, Belgium*
[2]*Walloon Agricultural Research Centre (CRA-W), Quality of Agricultural Products Department, Gembloux, Belgium*

The primary goal of using a regression model in multivariate calibration, for instance PLS, is to predict the value of a property of interest, the so-called predictand, and its uncertainty. The uncertainty of a calculated value is defined as a parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand. In most of the cases this uncertainty is calculated as a function of the different sources of uncertainty present in the model.

In the univariate context, prediction uncertainty is quantified by a sample-specific standard error of prediction. Unfortunately, multivariate models are inherently much more complex than their univariate analogues. Monte Carlo simulation techniques such as the bootstrap and the noise addition method can give an estimate of this uncertainty but also some approximate mathematical expressions have been proposed in the literature. One of these proposals is the correction made by De Vries and Ter Braak (1995) on the expression derived by Martens (1989) and used in the Unscrambler® software package. Another proposal is the simplification of Faber and Bro (2002) of an expression derived earlier under the errors-in-variables (EIV) model (1996).

The purpose of this study is to show how these two proposals work and to assess their results from the ones obtained using the bootstrapping and the noise addition methods. This study has been performed using embedded near-infrared (NIR) data sets. This kind of data is produced by a NIR instrument installed on a forage plot harvester. Thanks to this instrument the collection, compression and scanning of forage samples is performed during harvesting. In such a way, properties as the dry matter content in forages, for instance, can be measured without having to handle the samples or transport them to a laboratory.