Abstract
Sampling strategies for hyperspectral image models

J. Burger

BurgerMetrics SIA, Jelgava LV-3002, Latvia.
E-mail: james.burger@burgermetrics.com

Introduction

Applications of partial least squares regression (PLS) and PLS discriminant analysis (PLS-DA) to NIR spectra for quantification and classification purposes are well established and proven techniques. However creation of calibration models requires a known Y value for every spectrum included in the model training and test sets—either a quantity or class member value. This makes extension of PLS to hyperspectral imaging difficult, since accurate Y values for each pixel based spectrum are typically not known. Bulk sample values can be used to represent an average value, but in this case which hyperspectral image spectra should be used? This paper presents various strategies used for pixel spectra selection to optimise model building.

Materials and methods

Five varieties of Canadian wheat were imaged with a NIR based HyperLab hyperspectral imaging system. (BurgerMetrics SIA). Images contained 214 wavelength channels (960–1660 nm) and 320 × 300 pixels, with pixel spatial resolution of 100 microns. Individual mages were obtained for each of the five varieties, as well as augmented images with well-defined regions for each variety. Bulk hardness reference values were obtained from independent laboratory measurements. PLS and PLS-DA models were built using the interactive HyperLab hyperspectral visual calibration model software.

Results and discussion

Figure 1 shows an augmented image representing the 5 wheat classes. A 320 × 50 pixel image was selected from each class. The five coloured rectangles represent a general bulk analysis approach to spectral selection. This mimics what is available from a spot probe NIR instrument. A sixth class was selected to represent the background area separating the five wheat sample images. Individual PLS-DA models were made for each of the six classes—using a range of 8 to 10 factors. Classification threshold limits were selected using an interactive visualisation tool, which maps the predicted class members onto the training images.

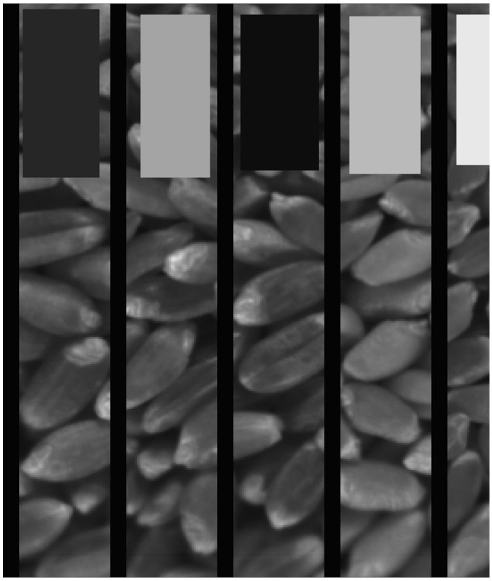

Figure 1. General "bulk" region of interest areas for five classes of wheat.

Figure 2 shows the preliminary prediction model obtained for the 5 wheat classes. Classes 2, 4 and 5 are clearly discriminated. Classes 1 and 3 highly overlapped each other, but were distinct from the other classes. Significant improvements to classification can be achieved using alternative spectral (pixel) training set selections. These strategies will be explained in detail and compared for both PLS-DA and PLS regression models.

Figure 2. PLS-DA prediction map for the five wheat classes.