



DETERMINATION OF THE GEOGRAPHIC ORIGIN OF TOBACCO LEAVES BY NIR

Eric Massicotte, Alcohol and Tobacco Section, Science and Engineering Directorate, Canada Border Services Agency

Abstract

The Canada Border Services Agency (CBSA) was allocated funding to develop chemical and biological methods to determine the geographic origin of seized tobacco and to compare tobacco products. This project was identified as the “Tobacco Origin Project”. One portion of the Tobacco Origin Project was designed to determine the geographic origin of flue-cured tobacco leaves using several analytical methods.

This poster presents the results obtained from Near Infrared (NIR) spectra in order to predict the geographic origin of tobacco leaves. Approximately 1000 samples of tobacco leaves from the top 10 flue-cured producing countries including Canada were analysed by NIR. A Partial Least Square Discriminant Analysis (PLSDA) model was developed. The results indicate that this model could be used as a good screening tool to determine if the tobacco originates in Canada, USA or another country (International).

Objective

- Evaluate the feasibility of using NIR as a screening tool to predict the geographical origin of tobacco leaves.

Experimental

- Each tobacco leaf was manually stripped (i.e. the stem portion was removed from the leaf).
- The stripped portion was ground and analysed using the integrated sphere interface of a Bruker MPA Near Infrared Analyser. The spectra was acquired for 1 min. at a resolution of 4 cm^{-1} from 10000cm^{-1} to 4000cm^{-1} .
- The spectra were saved under a .SPC format and imported directly into the PLS-ToolBox software using the DataSet Editor. A Class corresponding to the geographic origin of the leaves was assigned to each spectrum. The dataset was split into a calibration and a validation set.
- A group of pre-processing steps were applied to each spectrum (1st Derivative, GLS Weighting and Mean Center).
- A PLSDA model was created from the calibration set and evaluated using the validation set.

Fig.1 Predicting if samples are from Canada

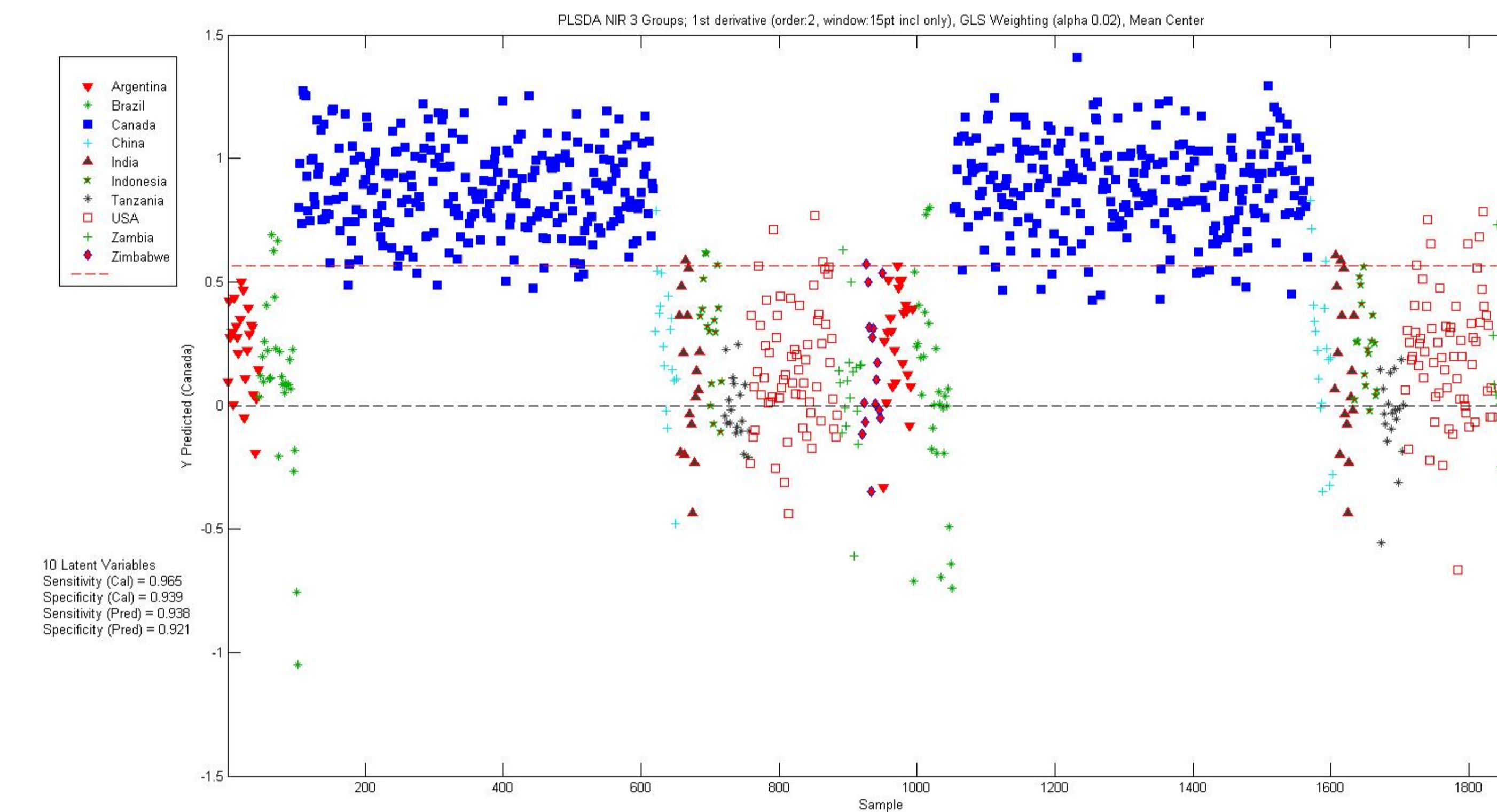


Fig.2 Predicting if samples are from USA

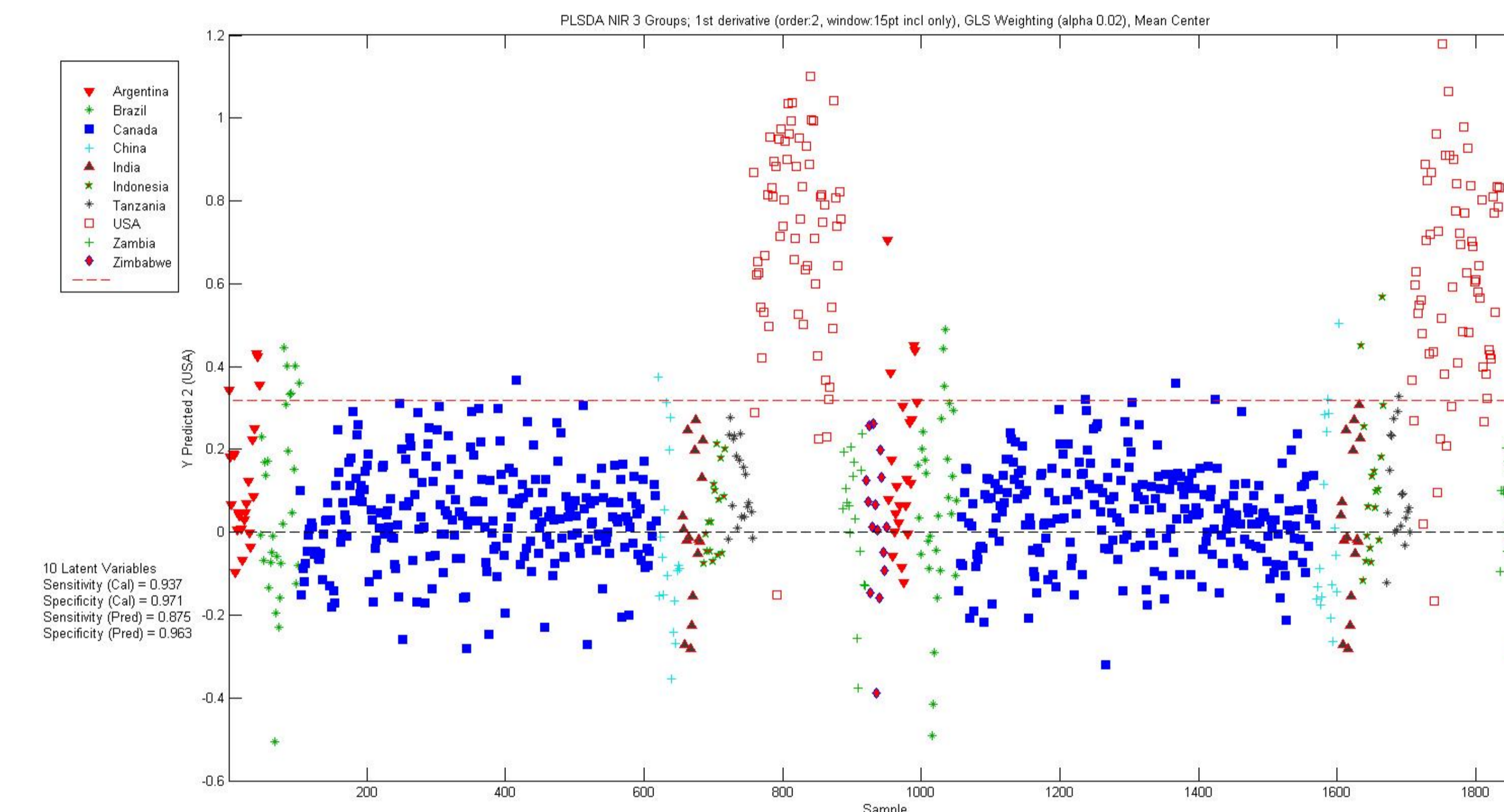
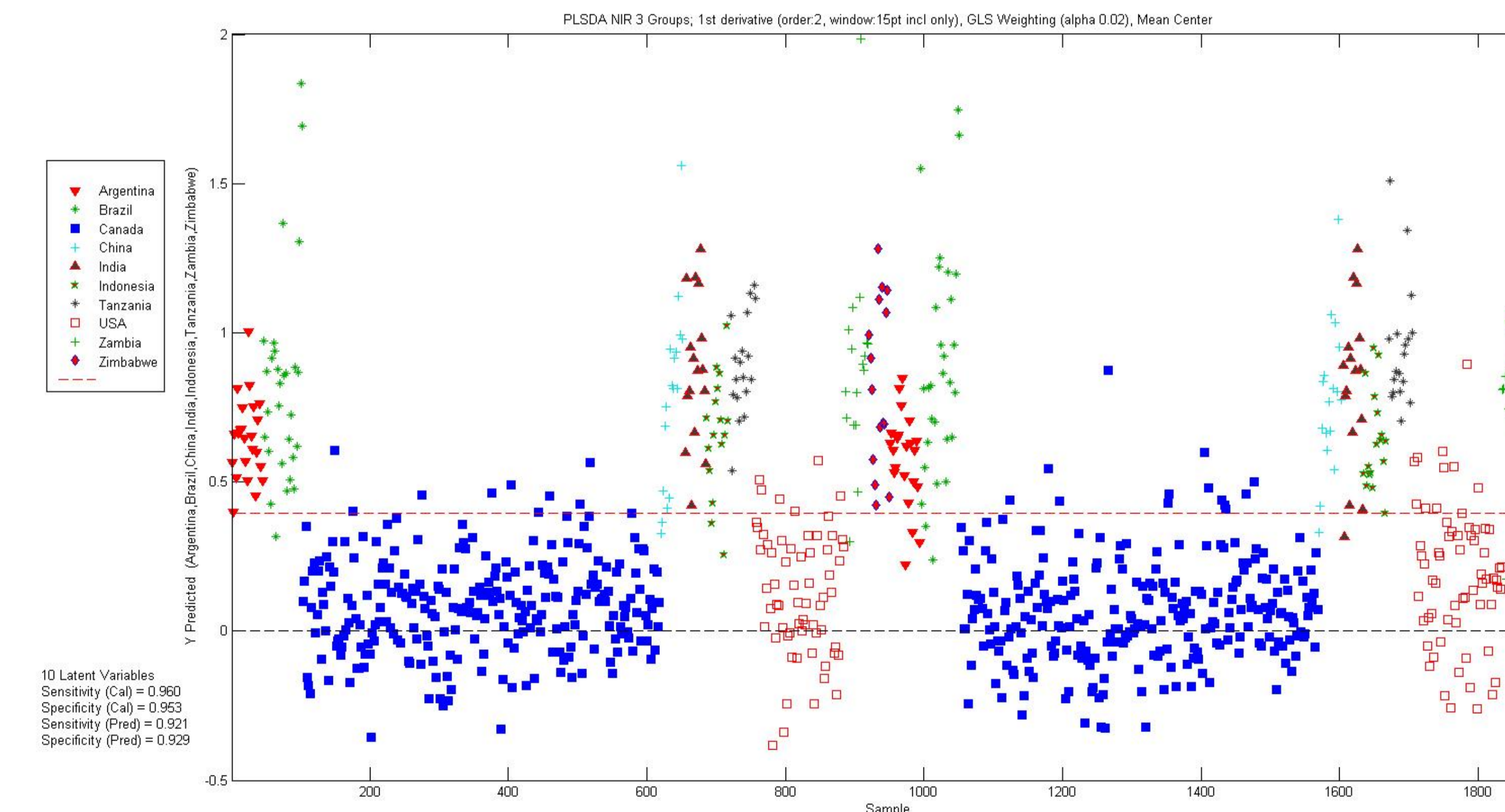


Fig. 3 Predicting if samples are from International



Results and discussion

- The model was built using 3 groups based on their geographical origin; Canada, USA and International. Although all International samples are grouped together, they are still displayed with a different icon.
- The first half of the samples (calibration set) was used to generate the model and the second half of the samples (validation set) was used to evaluate the model.
- A model using 10 Principal Components (PCs) was used. The number of PCs used was determined using two criteria: the shape of the “Variance Captured (%) / Latent Variable Number” plot and the performance of the model based on the calibration set. The validation set was kept completely independent and was not used to optimise the model.
- The results are presented in Figures 1 - 3
- The sensitivity and specificity was found to be better than 0.92 for all the groups (Canada, USA, International) except for the sensitivity of the USA group (0.875). This could be explained by a larger variability in the sample group coupled with a limited number of samples.

Conclusion

- This method may be used as a screening tool to determine if a tobacco leaf was grown in Canada, USA or another country.
- The performance of this model is very good considering that the sample preparation and analysis time for this method is less than 5 minutes per sample. Other models based on levels of trace elements, isotopic ratio and concentration of tobacco specific compounds show better performance but required significantly more resources to complete.

