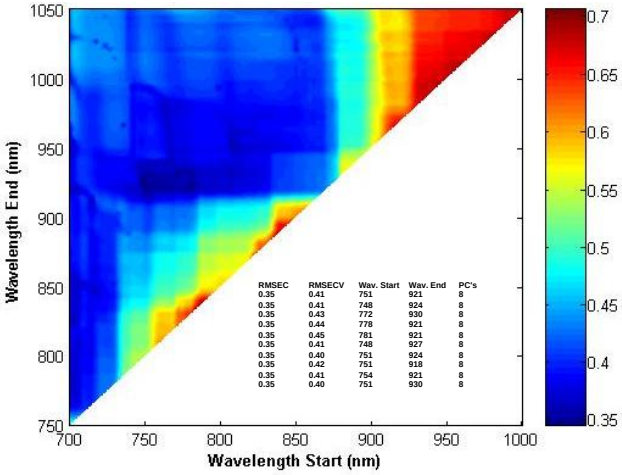
The sensitivity of updated NIR PLS models for fruit sorting to initial parameter settings

J. A. Guthrie[1,2], C.J. Liebenberg[2] and K.B. Walsh[2]

[1] **1 Delivery, Queensland Department of Primary Industries and Fisheries, PO Box 6014 CQ Mail Centre, Rockhampton 4702.**
2 Plant Sciences Group, Central Queensland University, Rockhampton, 4702, Australia

Partial least squares regression models on NIR spectra are often optimised (for wavelength range, mathematical pre-treatment and outlier elimination) in terms of calibration performance (R^2, RMSEC, RMSECV), and least frequently in terms of validation performance with reference to totally independent populations. For example, a searching window method applied to a calibration set suggest the use of a 750 to 951 nm window for assessment of total soluble solids concentration (ºBrix) of intact mandarin fruit (Fig. 1).

Fig. 1. Spectral window for ºBrix in intact mandarin

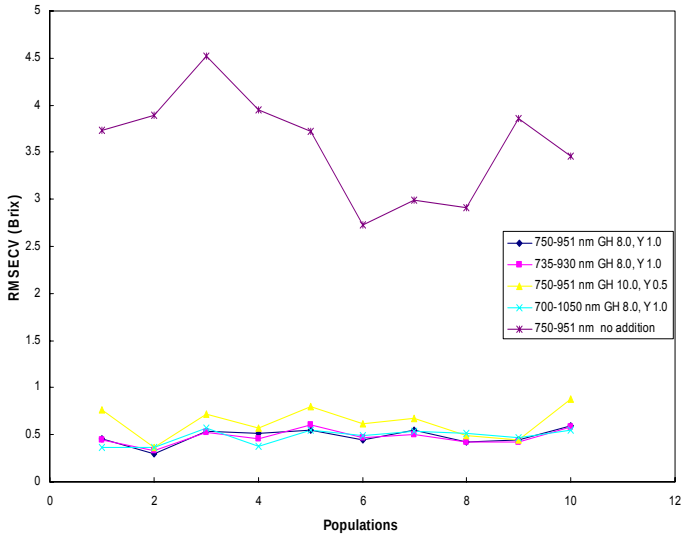


Calibration robustness varies with commodity with intact mandarin fruit being worse than thin skinned fruit (e.g. peach and plum, Golic *et al*. 2006). Thus calibration statistics can be encouraging but performance in validation must be documented. Model parameters should thus be optimized in terms of model validation, not model calibration statistics. For example, a population varying in maturity may yield better calibration statistics for a wavelength (λ) window that includes spectral features relevant to chlorophyll. Such a model may perform poorly across populations varying in maturity levels.

Model updating requires selection of λ range, data pre-treatment (derivatising, gap size etc.) and rules for outlier removal and number of PLS factor selection.

In this study with intact mandarin fruit, 3 λ windows and an outlier removal treatment were investigated with regard to model performance, using a Matlab script that enables automated calibration updating.

Fig. 2. RMSECV values for prediction of 10 populations with varying spectral windows and outlier removal



Without model updating, prediction error was high across the 10 validation populations (primarily bias, data not shown). The wavelength window was relatively unimportant in prediction across new populations (Fig. 2), relative to calibration (Fig. 1), although the use of the optimised window (750-951 nm) resulted in a lower number of samples identified as outliers. The outlier removal strategy requires optimisation with respect to the setting of the Y residual and Mahalanobis distance criteria.

References.

Golic, M. and Walsh, K.B. (2006). Robustness of calibration models based on near infrared spectroscopy for the in-line grading of stonefruit for total soluble solids content. *Analytica Chimica Acta* **555**,286-291.

Guthrie, J.A., Walsh, K.B., Reid, D.J. and Liebenberg, C.J. (2005). Assessment of internal quality attributes of mandarin fruit. 1. NIR calibration model development. *Australian Journal of Agricultural Research* **56**, 405-416.

PLANT SCIENCES GROUP
Central Queensland University
Rockhampton, Australia

Department of
Primary Industries and Fisheries
Queensland Government