

# Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content

N. T. Anderson, K. B. Walsh, P. P. Subedi and C. H. Hayes

Postharvest Biology and Technology, Volume 168, October 2020, 111202

---

## Abstract

Short wave near infrared spectroscopy has found use in non-invasive assessment of dry matter content (DMC, % fresh weight) of mango fruit, both as a guide to harvest maturity and ensure eating quality of ripened fruit. In this study, this application is optimised in terms of pre-processing of spectra, the source of variations important to model performance documented, and the performance of cultivar or physiological stage specific partial least squares regression (PLSR) models, global PLSR and an artificial neural network (ANN) model are compared. The data set consisted of 4675 samples acquired across four seasons, ten cultivars and two growing regions, with harvest populations used as cross validation groups. The data of the fourth season was reserved as an independent test set. Spectra pre-treatment of mean centred Savitzky-Golay second derivative (second order polynomial using a 17 point interval) and use of the wavelength range 684–990 nm gave the lowest RMSECV for PLSR models, although other ranges had similar statistics. The fruit physiological stage had the greatest impact on PLSR model performance, compared to cultivar, year or growing region, as estimated using a ‘variable importance metric’ devised and implemented using a random forest regression. The use of specific (cultivar or physiological stage) PLSR models improved prediction results of the independent validation set (RMSEP on DMC decreased from 1.01 to 0.88 %), and was similar to the result of a global ANN model (0.89 %). The use of an ANN model is recommended in terms of ease of use of a single model across all cultivars.