

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

II. Local PLS and nonlinear models

N. T. Anderson, K. B. Walsh, J. R. Flynn and J. P. Walsh

Postharvest Biology and Technology, Volume 171, January 2021, 111358

Abstract

A range of modelling techniques were used in the estimation of dry matter content of intact mango fruit from short wave near infrared spectra, collected using an interreflectance geometry, with models developed on a data set collected across three seasons ($n = 10,243$) and tested on that of a fourth season ($n = 1,448$). Model types included Artificial Neural Network (ANN), Gaussian Process Regression (GPR), Local Optimized by Variance Regression (LOVR), Local Partial Least Squares Regression (LPLS), Local PLS Scores (LPLS-S) and Memory Based Learner (MBL), with manual tuning of parameters undertaken. Additionally, two commercially available cloud-based chemometric packages for automated model development were trialled. All of these models gave a better result than use of a global PLS model. The best result (lowest RMSEP) was achieved with an ensemble of ANN, GPR and LPLS-S, with the best individual model result achieved by LOVR, with RMSEP of 0.839 % and 0.881 %, respectively, compared to the global PLS result of 1.014 %. The best precision was achieved with the LPLS model, with a SEP of 0.846 %, compared to the global PLS result of 1.012 %. LOVR was twice as fast as a generalized latent variable selection method LPLS-S-cv in prediction of independent validation set (at 58.7×10^{-3} s compared to 163×10^{-3} s). The ANN model was satisfactory in all categories (prediction speed, model build speed, and prediction statistics) and insensitive to tuning, e.g., 33 of the 70 parameter combinations were within 0.05 units of RMSEP of the minimum combination. However, the ANN learning rate was low. For applications that require ‘real-time’ prediction, such as fruit packlines, use of ANN and GPR models is recommended. For non-cloud based handheld NIR devices lacking the computational power to perform local modelling, ANN is recommended, and LOVR or a

model ensemble recommended in cloud based implementation. The automated cloud-based systems performed well (RMSEP of 0.850 % and 0.963 % for Hone Create and DataRobot, ensemble models, respectively), without human intervention for the choosing and tuning of models.