Models fused with successive CARS-PLS for measurement of the soluble solids content of Chinese bayberry by vis-NIRS technology

Lei-ming Yuan, Fei Mao, Guangzao Huang, Xiaojing Chen, Di Wu, Shaojia Li, Xinqi Zhou, Qiaojun Jiang, Dingpeng Lin and Ruyi He

Postharvest Biology and Technology, Volume 169, November 2020, 111308

Abstract

Variables selection methods have been proven successfully in the field of visible-near infrared spectroscopy (vis-NIRS) to optimize the predictive performance of regression models. However, because only selected spectral variables have been used, and discarding of residual spectral variables result in loss of spectral information. In this work, soluble solids content (SSC) of Chinese bayberry was non-destructively measured by a portable vis-NIRS equipment in the interactance spectral acquiring mode, and combined with a consensus modeling approach. The first member model was developed with the full spectra by applying a competitive adaptive reweighting algorithm (CARS), and the remainder developed successively with the residual spectral variables until the performance of the CARS- partial least square (PLS) model was not improved over that of the residual-spectral-based PLS model. A series of consensus models were developed with different number of top member-models in a fusing strategy of distributing the weightings. Results showed the residual spectral wavelengths after variables selection still reserved some useful information. In total, five CARS-PLS member models were developed. All consensus models performed better than any univocal member model, and the second consensus model F_2 that fused the top two- member models performed best. Compared to the full-spectral-based PLS model, the F_2 model promoted its performance with RMSECV of 0.80 by 11.3 % in the calibration set, and an RMSEP of 0.85 by 9.1 % in prediction set. The fusing strategy combined with member models that were successively developed with the discarded spectral variables utilized more useful information and improve the predictive performance.