SPORT pre-processing can improve near-infrared quality prediction models for fresh fruits and agro-materials

Puneet Mishra, Jean Michel Roger, Douglas N. Rutledge and Ernst Woltering

Postharvest Biology and Technology, Volume 168, October 2020, 111271

Abstract

Near-infrared spectroscopy (NIRS) is a key non-destructive technique for rapid assessment of the chemical properties of food materials. However, a major challenge with NIRS is the mixed physicochemical phenomena captured by the interaction of the light with the matter. The interaction often results in both absorption and scattering of the light. The overall NIRS signal therefore contains information related to the two phenomena mixed. To predict chemical properties such as dry matter, Brix and lipids, light refelction/absorption is used. Therefore, when the aim of the data analysis is to predict chemical components, it is necessary to remove as much as possible the scattering effects from the spectra. Several pre-processing techniques are available to do this, but it is often difficult to decide which one to choose. In this article we present the use of a recently developed pre-processing approach, sequential pre-processing through orthogonalization (SPORT), to improve the predictive power of multivariate models based on NIR spectra of food materials. The SPORT approach utilizes sequential orthogonalized partial least square regression (SOPLS) for the fusion of data blocks corresponding to several spectral preprocessing techniques. The results were compared with commonly used pre-processing techniques in the analysis of food materials by NIRS. The comparison was made by analyzing 5 different datasets comprised of apples, apricots, olive oils and grapes associated with chemical properties such as dry matter (DM), Brix, lipids and citric acid. The datasets were from both reflection and transmission measurements. The results showed that the fusion-based pre-processing methodology is an ideal choice for pre-processing of NIRS data. For four out of five datasets, the prediction accuracies (high R^2_{pred} and low RMSEP) were improved. The improvement led to as much as a 20 % increase in R^2_{pred} and a 25 % decrease in RMSEP compared to the standard 2nd derivative pre-processing. The pre-processing fusion was more effective for the

reflection mode compared to the transmission mode. Multiple pre-processing techniques provided complementary information, and therefore, their fusion using the SPORT approach improved the model performance. The methodology is not only applicable to food materials but can in fact be used as a general pre-processing approach for all types of modeling of spectral data.