Self-adaptive models for predicting soluble solid content of blueberries with biological variability by using near-infrared spectroscopy and chemometrics

Wei Zheng, Yuhao Bai, Hui Luo, Yuhua Li, Xi Yang and Baohua Zhang

Postharvest Biology and Technology; Volume 169, November 2020, 111286

Abstract

Biological variability is the natural characteristic of agricultural products. Non-destructive determination of fruit/vegetable soluble solid content (SSC) using spectral detection method is still a challenge due to the spectral variation caused by abundant biological variations, such as different cultivars, geographic origins and harvest seasons. In this paper, a self-adaptive model was established by combining five correcting methods for biological variability elimination, self-selection strategy and model search technology. Thus, the model can automatically adapt to the change of diverse biological variation compared to others. Furthermore, 100 cycles of selection accompanied with the random algorithm were set up to randomly select the calibration sets and prediction sets to ensure the reliability of the results. For the same batch of blueberry samples, five correcting models showed different prediction performances and all achieved satisfactory prediction accuracy compared to the individual-variation model and the hybrid-variation model. The consequence of the self-adaptive model showed consistency when considering multiple variation as well as variation with only cultivars or seasons. The best models in the three cases (multiple variation, only cultivars and only seasons) were all based on the preprocessing method, which was selected for 70, 57 and 47 times respectively. The results indicated that the biological variability had an impact on SSC prediction and that correcting models could improve the prediction accuracy. For the blueberry samples, the most suitable model selected according to the adaptive results was the preprocessing-based model. Within the study conditions, the self-adaptive model can select the most reliable model with the best prediction performance with respect to different variations.