

Robust generalized multiplicative scatter correction algorithm on pretreatment of near infrared spectral data

ABSTRACT

Recently, the use of Near Infrared (NIR) spectral sensor in agricultural process is getting much attention, particularly for fruit quality evaluation. The sensor requires a spectrometer to produce some sufficient information called spectrum as interaction between physical matters of the sample with the electromagnetic spectrum. In fact, the presence of experimental error or/and measurement error due to the heterogeneous particle size, moisture content variability, sample density, the instrument noise and pretreatment experience are often cannot be avoided. These would damage the spectra collected which results to decrease the performance in model selection, and increases the prediction error as the harmful influence of possible outlier and leverage points in dataset. To encounter these, a robust pretreatment of NIR spectral data is needed to correct the spectra before it is used for post-processing using any statistical method. In this paper, several different classical pretreatment methods were evaluated and a new robust Generalized Multiplicative Scatter Correction (GMSC) algorithm was proposed to correct the additive and/or multiplicative baseline effects in the spectral data. A dataset of NIR spectral on oil palm (*Elaeis guineensis* Jacq.) fruit bunch was used in the simulation. In the simulation, a number of repetitions using the single and double cross validation with robust partial least square are also applied. The Desirability Indices as statistical measures are presented for evaluating the methods.

Keyword: Near infrared; Spectral data; Robust statistics; Generalized multiplicative scatter correction; Standard normal variate; Multiplicative scatter correction; Detrend; Derivative; Partial least squares