A Physics-Based Data Simulation and Augmentation Method for Enhancing Calibration Model Development on a Visible Spectrometer

S Pinault, D Heran, D Brouckaert, F Chauchard-Rios, Ryad Bendoula

A PHYSICS-BASED DATA SIMULATION AND AUGMENTATION METHOD FOR ENHANCING CALIBRATION MODEL DEVELOPMENT ON A VISIBLE SPECTROMETER

S. Pinault[a,b], D. Heran[b], D. Brouckaert[a], F. Chauchard-Rios[a], R. Bendoula[b]
[a]INDATECH – Chauvin Arnoux, France
[b]INRAE, France
simon.pinault@chauvin-arnoux.com

Spectrometer calibration is a hot topic for PAT industrials and researchers because of the difficulties and challenges it brings [1]. For industrial applications, a low number of calibration samples and/or limited variability often compromise the chemometric calibration model development and hence its performances. Data augmentation is gaining popularity for overcoming this issue on visible or near-infrared spectrometers, for instance by using Deep Learning [2], pure components estimation [3] or interpolation [4].

This work presents a new data augmentation method based on a mechanistic physics-based approach. The method uses an optics model to simulate realistic synthetic spectra. Those simulated spectra are then used to augment measured calibration sets. The study focuses on a data set made out of a small number of samples with a limited variability. Moreover, we study the extreme case where the independent test data is at the edge of the variability range.

The data set consists of 55 phantoms samples displaying varying absorption and scattering properties, inspired by those made by Watté and al. [4]. They were inspected in reflectance and transmittance mode simultaneously, thanks to a double-integration sphere and visible spectrometer setup. Synthetic spectra were computed according to the Kubelka-Munk model [5] with Saunderson correction. Two data augmentation modes were investigated, namely completion and superposition. The augmented data sets are compared with a baseline set, in terms of correlation (R^2) and prediction (SEP) accuracy, by training partial least squares (PLS) regression models. This study investigates the benefits of augmenting the calibration data with synthetic data generated from various calibration sets.

The proposed method shows encouraging results in the extreme case where the test data are at the edge of the variability range (extrapolation). The strengths and weaknesses of the method will be addressed.

This work contributes to the recent trends in development of intelligent instruments for the PAT industry.

References

[1] W. Chew, P. Sharatt, Analytical Methods, 2010, 2, 1412-1438

[2] C. Jiang, J. Zhao, Sensors, 2023, 23(7), 3686

[3] R. Schiemer, M. Rüdt, Frontiers in Bioengineering and Biotechnology, 2024, 12, 1228846

[4] R. Watté, N. Do Trong, Optics Express, 2013, 21(26), 32630

[5] P. Kubelka, Journal of the Optical Society of America, 38(5), 448