Improvement of Prediction Errors Based on Standardized Infrared Spectra for a Calibration-free Approach

Shojiro SHIBAYAMA[1], Kimito FUNATSU[1*]

[1]School of Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

Abstract Continuous manufacturing (CM) in the pharmaceutical industry has been paid attention to, because it is expected to reduce the costs of manufacturing. One of technical hurdles in CM is establishment and maintenance of predictive models for process monitoring. Conventionally, calibration models with optic spectra such as infrared or Raman spectra have been used as the predictive models for process monitoring. The calibrated models predict product qualities such as active pharmaceutical ingredient's content, moisture content, particle size, and so on. However, any changes in rots, ratio of ingredients, or operation conditions may affect the relationship between sensor information and the product qualities, which results in deterioration of predictive models. Operators must update calibration models to assure predictive accuracy; however, calibration always requires data acquisition. Thus, the use of calibration models intrinsically increases economical costs. To tackle this problem, the authors have been attempting to propose a calibration-free approach with infrared spectra, which employs an equation in physics. To apply the calibration-free approach to real processes, it is important that a model provides accurate and reliable prediction. In this study, we propose a method to improve predictive accuracy of a calibration-free approach after assessing predictive errors using a rational indicator. We verified that the post-processing method succeeded in non-ideal binary mixtures.

1 Introduction

Continuous manufacturing (CM) has been paid attention to nowadays because of its efficiency in terms of human labour. However, there are several technical hurdles to realize CM. One of them is to construct and maintain predictive models for process monitoring. Conventionally, calibration models are used as process monitoring. Calibration models constructed by statistical regression models such as partial least squares predict product quality such as active pharmaceutical ingredient (API) content, moisture content, particle size, and so on from optic spectra. Constructing and maintaining the calibration models requires economic costs for data acquisition. To avoid this problem, use of calibration-free methods is effective.

One of useful calibration-free methods involving optic spectra is iterative optimization technology (IOT) [Muteki *et al.* 2013]. IOT predicts ingredients' content in mole fractions from infrared spectra of a mixture and pure components. IOT makes use of a theoretical equation called Beer's law. In other words, the constraint of Beer's law and pure components' spectra enables to predict ingredients' composition without calibration.

Other quantitative analysis methods for infrared spectra are multi-variate curve resolution-alternating least squares (MCR-ALS) [Juan *et al.* 2014] and indirect hard modelling (IHM) [Kriesten *et al.* 2008]. MCR-ALS and IHM are not the best methods as calibration-free methods, because they have many parameters to be fitted. On the other hand, IOT is expected to be effective as a calibration-free method, because IOT has as many constraints and equation as parameters.

Nonetheless, IOT has several limitations. One of them is application to non-ideal mixtures. IOT could be applicable for more various mixtures, if it is able to be applied to non-ideal mixtures. However, IOT assumes that a mixture spectrum follows Beer's law at all wavelengths, though it is not possible for real mixtures because of molecular interactions' effects. Thus, wavelength selection is required to eliminate wavelengths where Beer's law does not hold [Kaneko *et al.* 2015, Shibayama 2016]. However, once wavelength selection is conducted with a couple of spectra, the selected wavelengths are only fit to spectra similar to ones in non-ideal mixtures that were used for wavelength selection. In other words, wavelength selection spoils an advantage of calibration-free approaches, independence from calibration samples, though wavelength selection is necessary. As a result, IOT is not effective with the selected wavelengths, if composition changes significantly. If some indicators are effective to assess prediction errors for IOT, the selected wavelengths may be updated as prediction accuracy is improved.

In this study, we propose a way to update selected wavelengths. There is a rational scaling method for excess

---

absorbances that has the same form with the objective function of IOT. We found that the scaled excess absorbances can be used as a suitable indicator to assess prediction errors for IOT. We verified the update of selected wavelengths works with one set of binary mixtures consisting of acetic acid and water, which behave in a non-ideal way against Beer's law.

2 Methods

2.1 Iterative optimization technology

Iterative optimization technology (IOT) is a calibration-free strategy based on Beer's law. IOT minimizes an objective function. Beer's law is expressed as follows, using a mixture spectrum x_{mix} and pure component i's spectrum $x_{pure,i}$:

$$x_{mix} = \sum_{i=1...c} r_i x_{pure,i} + \varepsilon , \qquad (1)$$

where r_i and ε denote mole fraction of pure component i and error. IOT solves the following minimization problem with an optimization solver.

$$\min_{r1,r2,...,rc} |x_{mix} - \sum_{i=1...c} r_i x_{pure,i}|^2,$$

such that $\qquad\qquad\qquad\qquad\qquad\qquad (2)$

$$\sum_{i=1...c} r_i = 1, 0 \le r_i \le 1.$$

2.2 Wavelength selection based on excess absorption

One of obstacles for applications of IOT is unnecessary wavelengths. IOT is not effective when Beer's law does not hold at some wavelengths. The authors have been proposed two types of wavelength selection [Kaneko *et al.* 2015, Shibayama *et al.* 2016]. One indicator to determine if Beer's law holds is excess absorbance, which is difference between the observed mixture spectrum and mixture spectrum calculated using eq. (1) except ε, pure components' spectra, and observed mole fractions. IOT provides accurate prediction results after the wavelength selection.

2.3 Weighted IOT

We have found that excess absorbances, which have the same form with the objective function of IOT, can be scaled by the possible variance of Beer's law. By replacing the original objective function of IOT with the scaled excess absorbances, IOT provides more accurate prediction result that the original IOT. We call this new method weighted IOT (wIOT). The scaling enables us to evaluate excess absorbances across wavelengths equally.

2.4 Inference of errors for wIOT

A mixture spectrum x_{mix} can be estimated as $x_{mix,pred}$, by using estimated mole fraction $r_{pred} = (r_{pred,1}, r_{pred,2}, ..., r_{pred,c})$ with eq. (1). We have found that root mean square of x_{mix}-$x_{mix,pred}$ ($RMSE\ x_{mix}$) is an indicator to estimate prediction error in r [Shibayama *et al.* 2016, Shibayama and Funatsu 2018]. Once virtual mixture spectra are generated with eq. (1) and virtual mole fraction r, possible prediction errors in mole fraction can be estimated. The virtual mixture spectra is expressed as follows:

$$x_{mix,sim} = \sum_{i=1...c} r_{sim,i} x_{pure,i} + \varepsilon ,$$
$$\varepsilon_n \sim N(0, s^2), \qquad (3)$$

where $x_{mix,sim}$, $r_{sim,i}$, ε_n and s denote simulated mixture spectrum, simulated mole fraction of pure component i, measurement noise at wavelength n, and simulated standard deviation of the measurement noise. s is set various values to simulate disturbances of spectra.

This time, we replaced the indicator with the weighted $RMSE\ x_{mix}$. The scale is the same one with the objective function of wIOT. $RMSE\ x_{mix}$ after the scaling will show strong linear correlation with prediction errors in mole fraction.

2.5 Update of selected wavelengths

Prediction results for IOT depend on selected wavelengths, as described above. In addition to this, the weighted $RMSE\ x_{mix}$ has strong correlation with prediction errors in mole fraction. Thus, prediction results in mole fraction will decrease if the selected wavelengths are updated as $RMSE\ x_{mix}$ is minimized.

Figure 1. indicates how the update of selected wavelengths affects prediction results for wIOT. The red markers denote prediction results for wIOT. The orange area denotes prediction results for simulated spectra. The orange area is usually depicted as a group of dots. The points for simulated spectra usually forms a trapezoid, because the prediction errors in mole fraction and weighted $RMSE\ x_{mix}$ has strong correlation. Thus, prediction results can be improved by updating selected wavelengths as minimizing weighted $RMSE\ x_{mix}$. The update corresponds to the move represented by the arrow in the figure.

In the real situations, prediction errors in mole fraction are never known, because IOT is used as online process monitoring. Thus, operators of the process must refer to indicator, weighted $RMSE\ x_{mix}$ here, to check indirectly if process monitoring with IOT is reliable enough. Besides, operators may infer the prediction errors by checking the prediction results for simulated spectra. This operation stands on a hypothesis that weighted $RMSE\ x_{mix}$ correctly keeps strong correlation with the prediction errors for IOT.
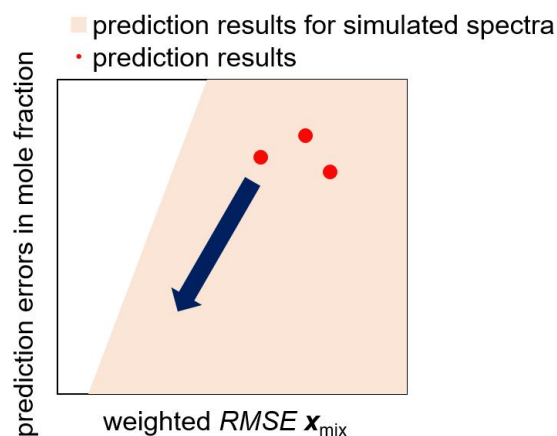
Figure 1. Concept of updating selected wavelengths.

The entire workflow in this study is shown in Figure 2. First of all, wavelength selection based on excess absorbances (WLSEA) is conducted. Second, wIOT is conducted to predict mole fraction r_{pred}. At last, the update of selected wavelengths is carried out to check if prediction accuracy is improved.
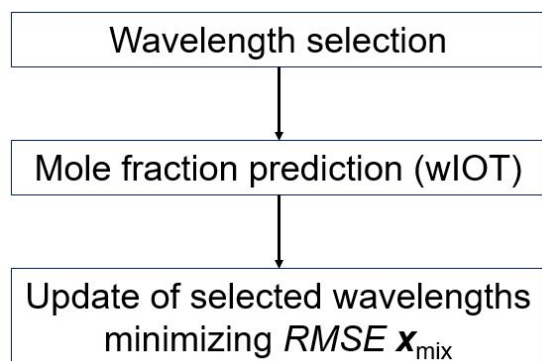

Figure 2. Flowchart of the entire analysis.

3 Results and discussion

3.1 Material

In this study, we used binary mixtures consisting of acetic acid (AA) and water for validation. The spectra were collected in different nineteen compositions [Shibayama *et al.* 2018]. Infrared spectra of AA and water will represent for industrial real ones, though AA and water are not the actual industry samples. This is because infrared spectra of AA and water follow Beer's law to some extent and because the spectra violate Beer's law due to molecular interactions. The mathematical features that appear in the spectra cover those found in actual industry samples.

We randomly took two samples out of the samples for wavelength selection. Others were treated as test samples.

3.2 Update of selected wavelengths

We first conducted wavelength selection using the two samples that were randomly sampled. Thus, the wavelengths selected first may be overfitted to the two samples. Peak positions and intensities of mixture spectra can change because of changes in compositions. Changes in composition affect magnitude of molecular interactions in absorbances. Thus, a set of the optimal wavelengths for IOT prediction is different across samples. Figure 3. indicates the relationship between the prediction results and weighted RMSE x_{mix}. The orange and blue markers denote prediction results for observed and simulated samples.
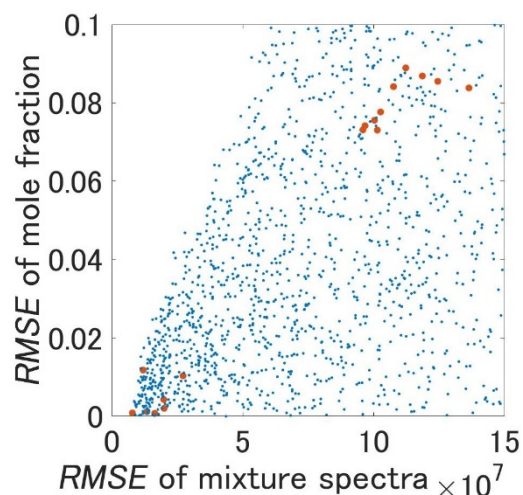

Figure 3. Relationship between weighted RMSE x_{mix} and prediction errors in mole fraction. Blue and orange markers denote prediction results for simulated virtual spectra and observed spectra.

The prediction results projected to the up right region in the figure are inaccurate results. The goal of update is to decrease the prediction errors of inaccurate results. We re-selected wavelengths as minimizing weighted x_{mix}. The wavelength selection was conducted in the same manner with WLSEA. First, excess absorbances for the two samples were sorted ascendingly. Then, wavelengths with excess absorbances from small to large were taken. Eventually, wavelengths were selected as weighted *RMSE* x_{mix} converged to minimal.

Figure 4. indicates the relationship between prediction results before and after update of selected wavelengths. The results with large prediction errors were located in the right area in the figure. The prediction errors for the results were successfully decreased from around 0.09 to 0.03. on the other hand, prediction results with small prediction errors were not improved. This implies that small weighted *RMSE* x_{mix} only defines the maximum possible prediction errors, not the exact prediction errors. Thus, prediction errors may even increase, when minimizing weighted *RMSE* x_{mix} for the samples with small weighted *RMSE* x_{mix}.
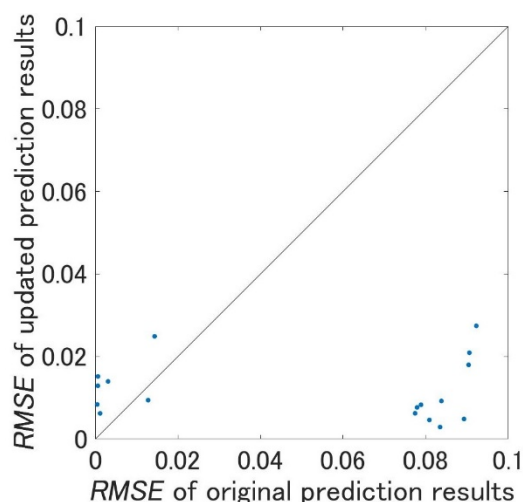
Figure 4. Prediction errors in mole fractions before and after the update of selected wavelengths. The markers correspond to mixtures.

4 Conclusion

We proposed a method to update selected wavelengths for IOT. We verified that the updating method improved predictive accuracy by minimizing an indicator, weighted $RMSE\ \boldsymbol{x}_{mix}$, through a case study with mixtures of AA and water. The prediction errors in mole fraction were decreased around 0.09 to 0.03. The dependency on calibration samples after wavelength selection was an issue; however, it was confirmed that the updating method can solve the issue. The scaling method for IOT and $RMSE\ \boldsymbol{x}_{mix}$ is effective for only binary mixtures. Thus, the authors will extend the method to multi-components systems.

Acknowledgment

Reference

K. Muteki, D.O. Blackwood, B. Maranzano, Y. Zhou, Y.A. Liu, K.R. Leeman, and G.L. Reid; "Mixture component prediction using iterative optimization technology (Calibration-Free/Minimum Approach)," *Ind. Eng. Chem. Res.* **52** 12258–12268 (2013).

A. de Juan, J. Jaumot, R. Tauler, "Multivariate Curve Resolution (MCR). Solving the mixture analysis problem," *Anal. Methods.* **6** 4964–4976 (2014).

E. Kriesten, F. Alsmeyer, A. Bardow, W. Marquardt, "Fully automated indirect hard modeling of mixture spectra," *Chemom. Intell. Lab. Syst.* **91** 181–193 (2008).

H. Kaneko, K. Muteki, and K. Funatsu; "Improvement of iterative optimization technology (for process analytical technology calibration-free/minimum approach) with dimensionality reduction and wavelength selection of spectra," *Chemom. Intell. Lab. Syst.* **147** 176–184 (2015).

S. Shibayama, H. Kaneko, and K. Funatsu; "Iterative optimization technology combined with wavelength selection based on excess absorption for a process analytical technology calibration-minimum approach," *Chemom. Intell. Lab. Syst.* **156** 137–147 (2016).

S. Shibayama, H. Kaneko, K. Funatsu, "Formulation of the excess absorption in infrared spectra by numerical decomposition for effective process monitoring," *Comput. Chem. Eng.* **113** 86–97 (2018).

S. Shibayama, and K. Funatsu; "Applicability Domains of a Minimal-Calibration Model for Effective Online Monitoring of Pure Components' Concentrations in the Pharmaceutical Continuous Manufacturing Processes", *13th Int. Symp. Process Syst.*, 919–924 (2018).