Food Chemistry 148 (2014) 289–293

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Food adulteration analysis without laboratory prepared or determined reference food adulterant values

John H. Kalivas [a,*], Constantinos A. Georgiou [b], Marianna Moira [b], Ilias Tsafaras [b], Eleftherios A. Petrakis [b], George A. Mousdis [c]

[a] Department of Chemistry, Idaho State University, Pocatello, ID 83209, USA
[b] Chemistry Laboratory, Agricultural University of Athens, 75 Iera Odos, 118 55 Athens, Greece
[c] Theoretical and Physical Chemistry Institute, National Hellenic Research Foundation, 48 Vassileos Constantinou Ave., 116 35 Athens, Greece

ARTICLE INFO

Article history:
Received 4 July 2013
Received in revised form 11 October 2013
Accepted 14 October 2013
Available online 23 October 2013

Keywords:
Food adulteration
Food quality control
Green chemical analysis
Extra virgin olive oil
Synchronous fluorescence spectroscopy
Tikhonov regularization

ABSTRACT

Quantitative analysis of food adulterants is an important health and economic issue that needs to be fast and simple. Spectroscopy has significantly reduced analysis time. However, still needed are preparations of analyte calibration samples matrix matched to prediction samples which can be laborious and costly. Reported in this paper is the application of a newly developed pure component Tikhonov regularization (PCTR) process that does not require laboratory prepared or reference analysis methods, and hence, is a greener calibration method. The PCTR method requires an analyte pure component spectrum and non-analyte spectra. As a food analysis example, synchronous fluorescence spectra of extra virgin olive oil samples adulterated with sunflower oil is used. Results are shown to be better than those obtained using ridge regression with reference calibration samples. The flexibility of PCTR allows including reference samples and is generic for use with other instrumental methods and food products.

1. Introduction

The adulteration of food products is of primary concern for consumers, food processors, regulatory agencies, and industries. Adulteration typically involves replacing or diluting high-cost ingredients with less expensive ones. Thus, once a food product is identified as being adulterated and with what substance, rapid quantitative analysis of the adulterant is needed.

As an example, it is well documented that to increase profits, extra virgin olive oil (EVOO) is adulterated with lower grade olive oil and/or less costly edible oils such as sunflower or corn oils. Because of the importance of the EVOO adulteration problem, numerous studies have shown that multivariate calibration methods in conjunction with spectroscopic measurements, can quantitate for adulterant concentrations in EVOO (Baeten, Meurens, Morales, & Aparicio, 1996; Davis, McIntyre, & Morgan, 2000; Gurdeniz & Ozen, 2009; Heise, Damm, Lampen, Davis, & McIntyre, 2005; Lai, Kemsley, & Wilson, 1995; Lerma-García, Ramis-Ramos, Herrero Martínez, & Simó-Alfonso, 2010; López-Díez, Bianchi, & Goodacre, 2003; Poulli, Mousdis, & Georgiou, 2006, 2007; Küpper, Heise, Lampen, Davis & McIntyre, 2000; Yang and Irudayaraj, 2001). In these studies, calibration samples are formed by adding the identified adulterant over a range of concentrations to EVOO samples from one or more geographical region, growing season, and/or cultivar. These samples are then spectrally measured. A multivariate calibration model is formed using a method such as partial least squares (PLS) or ridge regression (RR) (Hastie, Tibshirani, & Friedman, 2009; Kalivas, 2009; Næs, Isaksson, Fern, & Davies, 2002). This model is then used to predict adulterant oil concentrations in new samples. This approach is successful as long as new samples are from the same population used to form the calibration model, i.e., the new samples have the same matrix effects. Additionally, numerous samples need to be prepared causing additional time and costs before the quantitative analysis can be performed.

In this EVOO example, and in other food analysis cases, spectroscopic analysis has significantly improved the green factor of analyses compared to other methods requiring stabilization in solvents, extractions, and treatment with chemicals (Zandomeneghi, Carbonaro, & Caffarata, 1996). However, even with spectroscopic methods, numerous calibration samples requiring tedious laboratory work and often chemicals still need to be prepared. Recent studies have shown that it is possible to form a calibration model without reference samples (Boulet & Roger, 2010; Marbach, 2002, 2005; Ottaway, Farrell, & Kalivas, 2013). Thus, significant steps have been made towards providing greener chemical analysis including the potential for analysis of food adulterants. In the non-food analysis published studies, the calibration model is formed using an analyte

* Corresponding author. Tel.: +1 208 282 2726; fax: +1 208 282 4373.
E-mail address: kalijohn@isu.edu (J.H. Kalivas).

0308-8146/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodchem.2013.10.065

pure component spectrum and non-analyte spectra, i.e., samples where the analyte amount is known to be zero. In the most recent work of Ottaway et al. (2013), a dynamic weighting scheme is developed to obtain a balance between model prediction accuracy, model shrinkage (model vector size and hence complexity), and orthogonality to the non-analyte matrix effects (non-analyte spectra are predicted by the model to have zero analyte). This process of forming calibration models with the three way balance is named pure component Tikhonov regularization (PCTR) and is presented in this paper as a viable calibration method for food adulterants using EVOO as the example. Additionally, PCTR is easily adaptable to conditions that alter the sample matrix such as a new growing season or geographical region.

2. Mathematics

2.1. Multivariate calibration

In quantitative food adulteration, spectroscopic multivariate calibration relates the concentration of the adulterant $\mathbf{y}$ to the measured spectra $\mathbf{X}$ by:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ denotes an $m \times 1$ vector of quantitative analyte values for m reference samples, $\mathbf{X}$ represents the $m \times n$ matrix of spectra measured at the n wavelengths, $\mathbf{b}$ signifies the $n \times 1$ vector of model vector of coefficients to be estimated, and $\mathbf{e}$ symbolizes the $m \times 1$ vector of normally distributed errors with mean zero and covariance matrix $\sigma^2\mathbf{I}$ with $\mathbf{I}$ representing the $m \times m$ identity matrix. The relationship described in Eq. (1) assumes $\mathbf{y}$ and $\mathbf{X}$ are column-wise mean centered or constrained to the origin and hence, no $\mathbf{y}$-intercept term. The estimated regression vector symbolized as $\hat{\mathbf{b}}$, is commonly computed by PLS or RR. Once $\hat{\mathbf{b}}$ has been estimated, it can then be used to predict future sample spectra provided the new food sample matrix is part of the calibration domain. As previously noted, the samples making up $\mathbf{X}$ and $\mathbf{y}$ need to characterize the sample matrix and measurement conditions (chemical, physical, instrumental, and environmental) of any sample to be predicted. If the conditions changes, either new samples are needed or some sort of calibration maintenance is required (Brown, 2009; Kalivas, 2012).

In order to accomplish a multivariate calibration using Eq. (1), a large number of samples are commonly required to effectively characterize the analyte and matrix effects, i.e., food product variety, growing season, and/or geographical region, etc. This calibration process can be time consuming, require chemical treatments, and costly. Therefore, an objective of multivariate calibration is to maintain low prediction errors and simultaneously, limit the number of reference samples.

As with any vector, there are two features of the calibration model vector $\hat{\mathbf{b}}$ that are key. One is the model complexity characterized by the Euclidean length represented by $\|\hat{\mathbf{b}}\|$ for the L_2 norm. The other is the model vector direction which in conjunction with the magnitude, affect the accuracy of sample predictions made with the model vector. Generally, for a given $\mathbf{X}$ and $\mathbf{y}$ data set, the greater the model size, the more complex (and potentially overfitted) the calibration model. Conversely, a model with too much shrinkage is too small in size and the model is underfitted with poor predictive accuracy. The tradeoff in model direction and shrinkage is characterized by the bias/variance tradeoff for the model prediction. As noted previously, calibration models are commonly estimated using PLS or RR. These methods are termed biased as each requires selection of a meta-parameter (tuning parameter) to balance the bias/variance tradeoff, i.e., the direction and magnitude are balanced by the respective tuning parameter value (Kalivas & Palmer, 2013). In conjunction with magnitude and direction of the model vector, another model vector feature is the orthogonality to the current non-analyte sample matrix and/or measurement conditions (Brown, 2004; Skibsted et al., 2005; Zeaiter, Roger, & Bellon-Maurel, 2005). Presented next is PCTR that provides a flexible method to balance model size and direction with the degree of model orthogonality to the non-analyte space while simultaneously sustaining accurate predictions. To accomplish the balance, PCTR uses a pure component analyte spectrum and non-analyte spectra. The flexibility of PCTR also allows including reference samples if such samples are available.

2.2. Pure component TR (PCTR)

For a measured spectrum $\mathbf{x}$ and assuming a linear Beer–Lambert law type relationship is valid, then $\mathbf{x}$ can be expressed as:

$$\mathbf{x}^t = y_a\mathbf{k}_a^t + \mathbf{y}_N^t\mathbf{K}_N + \mathbf{r}^t \tag{2}$$

where, in the spectroscopic situation, y_a and $\mathbf{k}_a$ respectively denote the analyte concentration and pure component analyte spectrum at unit concentration, $\mathbf{y}_N$ and $\mathbf{K}_N$ symbolize the interferent concentrations and pure component non-analyte spectra at unit concentration as rows in $\mathbf{K}_N$, and $\mathbf{r}$ represents the random spectral noise. The non-analyte spectra in $\mathbf{K}_N$ can be pure component interferent spectra as well as spectra representing instrumental and/or environmental sources affecting $\mathbf{x}$ such as scatter, baseline shifts, background, temperature, etc., (all the components of $\mathbf{x}$ not due to the analyte). While pure component interferent spectra are sometimes obtainable, other pure component non-analyte spectra making up $\mathbf{K}_N$ are usually not. Additionally, the amount of respective non-analyte components in samples are typically not known. If the spectra are scaled by the respective quantities in $\mathbf{y}_N$, then Eq. (2) becomes:

$$\mathbf{x}^t = y_a\mathbf{k}_a^t + \mathbf{1}^t\mathbf{N} + \mathbf{r}^t \tag{3}$$

where the 1 signifies a vector of ones with as many ones as there are spectra in $\mathbf{N}$. In food adulteration, $\mathbf{k}_a$ would be the pure component spectrum of the adulterant and spectra for $\mathbf{N}$ would simply be the spectra of the food item without the adulterant.

Prediction for y_a, expressed as $\hat{y}_a$, is computed by multiplying $\mathbf{x}$ in Eq. (2) by an estimated model vector $\hat{\mathbf{b}}$ written as:

$$\hat{y}_a = \mathbf{x}^t\hat{\mathbf{b}} = y_a\mathbf{k}_a^t\hat{\mathbf{b}} + \mathbf{1}^t\mathbf{N}\hat{\mathbf{b}} + \mathbf{r}^t\hat{\mathbf{b}} \tag{4}$$

Based on Eq. (4), three conditions need to be satisfied in order to obtain an accurate prediction with $\hat{y}_a = y_a$. These conditions are: (1) $\mathbf{k}_a^t\hat{\mathbf{b}} = 1$ (minimum bias), (2) $\mathbf{N}\hat{\mathbf{b}} = \mathbf{0}$ (orthogonality), and (3) $\mathbf{r}^t\hat{\mathbf{b}} = 0$ (orthogonality and/or low model complexity). Unfortunately, not all three conditions can usually be simultaneously satisfied and a compromise is needed in forming the model from Eq. (1). Additionally, depending on how much the matrix effects, represented by the non-analyte spectra in $\mathbf{N}$, for the new sample deviate from the current matrix effects present in $\mathbf{X}$ used to form the model with Eq. (1), will dictate the level of accuracy in the prediction from Eq. (4), i.e., the success of the prediction depends on the analysis (sample) specific situation matrix matching the calibration matrix span (Ottaway et al., 2013).

In order to balance the three conditions, a variant of Tikhonov regularization (TR) was recently developed named pure component TR (PCTR). The PCTR approach incorporates and balances these tradeoffs directly in the minimization expression:

$$\min(\|\mathbf{k}_a^t\mathbf{b} - 1\|_2^2 + \eta^2\|\mathbf{b}\|_2^2 + \lambda^2\|\mathbf{N}\mathbf{b} - \mathbf{0}\|_2^2) \tag{5}$$

with the solution