JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

# Updating a Synchronous Fluorescence Spectroscopic Virgin Olive Oil Adulteration Calibration to a New Geographical Region

Matthew Ross Kunz,[†] Joshua Ottaway,[†] John H. Kalivas,[*,†] Constantinos A. Georgiou,[‡] and George A. Mousdis[§]

[†]Department of Chemistry, Idaho State University, Pocatello, Idaho 83209, United States

[‡]Chemistry Laboratory, Agricultural University of Athens, 75 Iera Odos, 118 55 Athens, Greece

[§]Theoretical and Physical Chemistry Institute, National Hellenic Research Foundation, 48 Vassileos Constantinou Avenue, 116 35 Athens, Greece

**ABSTRACT:** Detecting and quantifying extra virgin olive adulteration is of great importance to the olive oil industry. Many spectroscopic methods in conjunction with multivariate analysis have been used to solve these issues. However, successes to date are limited as calibration models are built to a specific set of geographical regions, growing seasons, cultivars, and oil extraction methods (the composite primary condition). Samples from new geographical regions, growing seasons, etc. (secondary conditions) are not always correctly predicted by the primary model due to different olive oil and/or adulterant compositions stemming from secondary conditions not matching the primary conditions. Three Tikhonov regularization (TR) variants are used in this paper to allow adulterant (sunflower oil) concentration predictions in samples from geographical regions not part of the original primary calibration domain. Of the three TR variants, ridge regression with an additional 2-norm penalty provides the smallest validation sample prediction errors. Although the paper reports on using TR for model updating to predict adulterant oil concentration, the methods should also be applicable to updating models distinguishing adulterated samples from pure extra virgin olive oil. Additionally, the approaches are general and can be used with other spectroscopic methods and adulterants as well as with other agriculture products.

**KEYWORDS:** Virgin olive oil, adulteration, synchronous fluorescence spectroscopy, Tikhonov regularization, model updating, geographical effect, food quality control, wavelength selection

## INTRODUCTION

With increasing international demand for extra virgin olive oil, the price has increased well above those of the most common edible oils. The International Olive Oil Council defines extra virgin olive oil as being produced by physical means with no chemical treatment, having no more than 0.8% acidity, and having been judged to have a superior taste.[1] In attempts to increase profits, extra virgin olive oil is adulterated with lower grade olive oil and/or less costly edible oils such as sunflower or hazelnut oils. It has been well documented that multivariate analysis methods in conjunction with spectroscopic measurements are able to predict adulterant concentrations in extra virgin olive oil samples.[2−17] Frequently used spectroscopic approaches are fluorescence,[2,3] Raman,[4−7] infrared,[8−11] and near-infrared.[12−14]

In these spectroscopic studies, a series of extra virgin olive oil samples from one or more geographical region, growing season, and cultivar have adulterant oil added, such as sunflower, corn, or olive pomace, over a range of concentrations. The samples are spectroscopically measured, and a multivariate calibration model is formed using a method such as partial least-squares (PLS). This model is then used to predict adulterant oil concentrations in new samples. Such an approach is successful as long as new samples are from the same population used to form the calibration model.

Olive oils have distinct chemical compositional differences based on which region the olive oil comes from. For example, the fatty acids present in the triacylglycerol and respective locations on the glycerol backbone vary due to chemical or physical factors, such as soil composition, climate, elevation, and time at which the olive oil was harvested.[15,16] Because of this variance, many spectroscopic studies have been able to demonstrate successful geographical classifications of extra virgin olive oils [16−26] as well as distinguish between virgin olive oils, pure olive oils, and olive pomace oils[27] and cultivar identification.[28] Common spectroscopic approaches are fluorescence,[17,27] infrared,[18,19,28] near-infrared,[19−21,28] and nuclear magnetic resonance (NMR).[22−26] In two geographical classification studies, harvest year was found to have a greater influence on geographically grouping virgin olive oils compared to the cultivar and actual origin.[16,26] Other studies have shown that geographical classification with multiple harvest years is less accurate than single-year classification models.[22,24] A geographical classification study based on certain harvest years was not able to accurately classify samples from the same regions but different harvest years.[21] A seasonal effect has also recently been observed in another classification study.[20]

Other studies have shown that fluorescence,[29] Raman,[4,6,30,31] infrared,[10,11,31] near-infrared,[12,14] and NMR[32] can be used to determine

**Received:** September 30, 2010
**Accepted:** December 22, 2010
**Revised:** December 20, 2010
**Published:** January 20, 2011





dx.doi.org/10.1021/jf1038053 | *J. Agric. Food Chem.* 2011, 59, 1051–1057

whether a sample of extra virgin olive is pure or adulterated. Again, the success of these investigations relies on fact that the validation sets used are from the same population as the calibration set used to form the discriminating model.

Predicting adulterant concentration in extra virgin olive oil is the focus of this paper. The calibration model used for predictions is likely to be specific to the calibration sample set of geographical regions, growing seasons, cultivars, and extraction methods (the primary conditions). Thus, such a model will probably not be able to accurately predict new samples originating from new geographical regions, growing seasons, etc. (secondary conditions). In a recent study, it was found that sunflower adulteration concentration determination was more accurate with signal regional models compared to a model based on all Greek regions.[6] In another study, the best prediction models were those based on one harvest year to predict sunflower adulteration concentration from samples in that same year compared to a multiple harvest year model.[8] Because it is indeed expected that samples will derive from new secondary conditions, then methods need to be developed allowing a primary calibration model to accurately predict adulterant concentrations for new samples from the secondary condition population.

Several multivariate calibration maintenance and transfer methods have been proposed that may be applicable to the extra virgin olive oil problem just described and are well reviewed.[33] The original broad basic goal of calibration maintenance and transfer in analytical chemistry is to allow the primary calibration model to remain useful for the analyte in secondary conditions such as a new temperature, pressure, particle size, pH, or humidity, new spectrally responding species, a different instrument, or the same instrument at a later point in time from the primary calibration development. Essentially, the primary calibration model needs to be maintained to accurately predict under new chemical, physical, environmental, and/or instrumental effects not spanned in the primary calibration domain. This paper expands fundamental calibration maintenance and transfer to include new geographical regions. Three common approaches exist and are described next, mostly in the context of geographical regions, but these methods can also be described in terms of climate, tree variety, and other sources of extra virgin olive oil differences.

The first approach is to build a global model robust to geographic regional effects from all extra virgin olive oil production regions. Building such a model requires many samples from each region to properly span anticipated regional variances. This approach is unrealistic due to the number of samples required to fully span all geographic regions. Additionally, for growing season effects, it is impossible to form a model that spans future climatic effects. When too many regions and other spectroscopic variances are included in the calibration set, the predictive quality in terms of accuracy has been shown to degrade compared to having only a few regions in the calibration set.[6] This result suggests forming single regional models, but the number of regions that produce extra virgin olive oil is too great to build a model for each individual region. Also, as with one model for all regions, this would entail a large number of samples for each region and does not address climatic or tree variety effects.

Global modeling can also be accomplished by using spectral preprocessing methods such as multiplicative scatter correction, finite impulse response filters, derivatives, and orthogonal correction.[33] Work has also shown that by selecting specific wavelengths from the respective spectral range, a model can be made robust to the primary and secondary conditions. This is important as two of the TR methods studied include wavelength selection as part of the calibration maintenance and transfer process.

The second approach is to manipulate sample spectra measured from a new region to fit the regional spectra used to form the primary model. However, to accomplish this conversion, a transformation matrix is needed. The transformation matrix allows mapping from one spectral domain to another and is based on a small set of samples measured in both the primary and secondary conditions. Clearly this is not realistic when one is correcting for a regional effect as it is impossible to measure the same extra virgin olive oil sample from two different geographical regions.

The third approach is model updating by augmenting sample spectra from a new region to the primary regional samples, allowing an updated model to be formed for predicting samples from the new region. Similar to the robust global model approach, many samples would normally be required to span the new region to fully account for the new spectral regional information. This amounts to a full recalibration. However, using the newly developed Tikhonov regularization (TR) approaches,[34−36] only a few samples from the new region should be needed to update the model. In essence, the approach is to expend extra effort collecting many samples from the primary region(s) and form a well-defined accurate model (a "golden" model) for the primary region. Then, only a few samples from the new secondary region (or growing season, tree variety, etc.) are used with TR to update the model.

Using a near-infrared spectral data set predicting ethanol in three-component mixtures, two TR-based methods successfully updated a primary model formed at one temperature to predict at new temperatures.[34−36] Similarly, these TR variants successfully updated a primary model formed on one instrument to predict sample analyte concentrations using spectra measured on another instrument. These studies were based on corn samples measured on three near-infrared spectrometers. In both data sets, a TR method including wavelength selection performed best. Unlike most other calibration maintenance and transfer methods, model updating with TR can correct for situations when new variances arise such as new spectrally responding species or the adulterant concentration is outside the primary concentration range.

Although geographical difference can be as great as between countries, for example, extra virgin olive oil samples from Spain, Italy, Portugal, and Greece, this paper concentrates on regional differences of extra virgin olive oil within Greece. Specific regions studied consist of using Zakynthos as the primary calibration region to predict samples from Attica and from a commercial source of unknown region(s) (supermarket samples). The adulterant studied in this paper is sunflower oil. Also studied in this paper is a new third TR variant that includes a new version of wavelength selection. Thus, two of the three studied TR variants include wavelength selection as part of the model updating.

## TIKHONOV REGULARIZATION VARIANTS

A standard relationship for the primary multivariate calibration model is stated as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \quad (1)$$

where $\mathbf{y}$ represents an $m \times 1$ vector of quantitative analyte values, such as the concentration of adulterant, for $m$ calibration

samples, $\mathbf{X}$ denotes the $m \times n$ matrix of spectra measured at $n$ wavelengths or frequencies, and $\mathbf{b}$ symbolizes the $n \times 1$ vector of primary model coefficients. The $m \times 1$ vector $\mathbf{e}$ indicates normally distributed errors with mean zero and covariance matrix $\sigma^2\mathbf{I}$. When $n \gg m$, the common spectral situation, then methods such as ridge regression (RR, a form of TR), PLS, or principal component regression (PCR) can be used to estimate $\mathbf{b}$ by $\hat{\mathbf{b}} = \mathbf{X}^{+}\mathbf{y}$, where $\mathbf{X}^{+}$ is the respective pseudo inverse of $\mathbf{X}$. The estimated model vector $\hat{\mathbf{b}}$ needs to accurately predict analyte values in any new sample spectrum $\mathbf{x}_{new}$ by $\hat{y}_{new} = \mathbf{x}^{t}_{new}\hat{\mathbf{b}}$. Wavelengths can be selected and have been shown to improve prediction accuracy. If a small enough wavelength set is selected such that $n \leq m$ and $\mathbf{X}$ is well conditioned, then multiple linear regression (MLR) can be used to estimate the model vector.

To update a primary model to a secondary condition (geographical region, growing season, etc.), only a few samples from the new region need to be spectrally measured if these new samples are properly weighted by $\lambda$. Letting $\mathbf{M}$ and $\mathbf{y}_{\mathbf{M}}$ denote the new spectra and adulterant concentrations, respectively, eq 1 is modified to include the weighted augmentation as (ignoring the $\mathbf{e}$ term)

$$\begin{pmatrix} \mathbf{y} \\ \lambda\mathbf{y}_{\mathrm{M}} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \lambda\mathbf{M} \end{pmatrix}\mathbf{b} \qquad (2)$$

The weight meta-parameter (tuning parameter) prevents the new calibration model from being overly influenced by the original primary region. In recent calibration maintenance and transfer work,[34] it was found that eq 2 needed a second tuning parameter ($\eta$) to stabilize an inverse computation in the TR algorithm, forming

$$\begin{pmatrix} \mathbf{y} \\ 0 \\ \lambda\mathbf{y}_{\mathrm{M}} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \eta\mathbf{I} \\ \lambda\mathbf{M} \end{pmatrix}\mathbf{b} \qquad (3)$$

This TR approach and another TR variant used in this study are well described[34−36] and briefly summarized next, allowing a brief introduction to a new third TR variant.

**TR 2.** A solution to eq 3 satisfies

$$\min(||\mathbf{Xb} - \mathbf{y}||_2^2 + \eta^2||\mathbf{b}||_2^2 + \lambda^2||\mathbf{Mb} - \mathbf{y}_{\mathrm{M}}||_2^2) \qquad (4)$$

where $||\bullet||$ denotes the vector $p$-norm and the subscript $p = 2$ symbolizes the vector 2-norm (Euclidean norm). If the third term is removed, expression 4 reduces to RR. Thus, expression 4 can be thought of as RR with an additional 2-norm penalty on predicting a few samples from the new region. As with using PLS for eq 2, two meta-parameters need to be determined to form the final model. Determination of these parameters is based on an L-curve approach and is briefly discussed under Results and Discussion.

**TR 2-1.** An alteration of TR 2 is to use a 1-norm penalty on the model vector instead of a 2-norm penalty. In this case, expression 4 becomes

$$\min(||\mathbf{Xb} - \mathbf{y}||_2^2 + \tau||\mathbf{b}||_1 + \lambda^2||\mathbf{Mb} - \mathbf{y}_{M}||_2^2) \qquad (5)$$

where $\tau$ symbolizes the meta-parameter weighting the 1-norm penalty. By switching to the 1-norm, $\mathbf{b}$ is now restricted to being a sparse solution forcing wavelength selection. The TR 2-1 method simultaneously forms the model and selects wavelengths. Both $\tau$ and $\lambda$ are determined by the L-curve approach to meta-parameter selection. If the third term is removed, the TR expression is commonly known as the least absolute shrinkage and selection operator (LASSO). In this case, a primary calibration model can be formed with wavelengths selected[37] and will be referred to as TR LASSO in this paper.

**TR 2b.** The final and third variant of TR studied in this paper is new and has not been investigated for modeling updating. This variation is similar to TR 2 except an $n \times n$ diagonal matrix $\mathbf{L}$ is used in the model 2-norm penalty. Expression 4 is modified to

$$\min(||\mathbf{Xb} - \mathbf{y}||_2^2 + \eta^2||\mathbf{Lb}||_2^2 + \lambda^2||\mathbf{Mb} - \mathbf{y}_{\mathrm{M}}||_2^2) \qquad (6)$$

where the diagonal of $\mathbf{L}$ contains a TR 2 model vector as $1/|\hat{b}_i|$. After first forming a model vector solving expression 4, the goal is to form the next model vector using expression 6 such that wavelengths with large model vector coefficients are weighted more than small coefficients. In a basic multivariate calibration study with no concern for calibration maintenance and transfer, such an approach was found to be useful. Bands of important wavelengths were selected compared to the 1-norm TR LASSO approach that tends to select single wavelengths.[37] Both $\eta$ and $\lambda$ in expression 6 are determined by the L-curve approach.

## MATERIALS AND METHODS

**Software.** MatLab 7.8 (The MathWorks, Natick, MA) was used for the three TR algorithms including the least angle regression (LAR) used in TR 2-1 to incorporate the 1-norm. These algorithms are have been previously described.[35,37]

**Data Centering.** Data are mean centered using the local mean centering approach. Local mean centering is the process whereby $\mathbf{X}$, $\mathbf{y}$, $\mathbf{M}$, and $\mathbf{y}_{\mathbf{M}}$ are each mean centered to respective means. Validation spectra measured in the same region as $\mathbf{M}$ are then mean centered to the mean of $\mathbf{M}$.

**Data Set.** Extra virgin olive oil samples and synchronous fluorescence spectra are the same as those previously described and used to form a PLS model predicting sunflower oil concentration.[2] Greek regional extra virgin olive oil samples and local shop samples of Greek origin were varied in sunflower oil concentration ranging from 0.5 to 95% (w/w). The spectral range used is from 250 to 502 nm at $\Delta\lambda = 80$ nm.

Extra virgin olive oil samples come from three sources: the two geographical regions Zakynthos and Attica and purchased from a local shop in Athens. Of the 28 samples used from Zakynthos, 10 were randomly selected to form the primary regional calibration model and the remaining 18 samples were used to validate the primary model. The calibration and validation samples respectively ranged from 4.741 to 89.949% and from 9.714 to 94.540% sunflower oil. A total of 10 Attica samples are available. Of these, 4 were randomly selected to serve as the standardization set ($\mathbf{M}$) and the remaining 6 formed the validation set. The standardization set and validation samples, respectively, ranged from 6.412 to 72.556% and from 2.6596 to 89.7207% sunflower oil. Thirteen local shop samples exist; 4 were randomly selected for the standardization set, and the remaining 9 samples served as the validation set. The standardization set and validation samples, respectively, ranged from 1.095 to 77.956% and from 0.720 to 91.873% sunflower oil. Plotted in Figure 1 are mean regional spectra.

**Meta-parameter $\lambda$, $\eta$, and $\tau$ Values.** For TR 2 and TR 2b, $\eta$ and $\lambda$ values ranged exponentially from 0 to $3.75 \times 10^6$ in 80 increments for $\eta$ and 100 increments for $\lambda$. Meta-parameter $\lambda$ for TR 2-1 also ranged exponentially from 0 to $3.75 \times 10^6$ in 100 increments. Because the LAR algorithm is used for the 1-norm, each $\tau$ is determined by the LAR algorithm. Essentially, successive LAR models approach the least-squares solution and correlate to increasing $\tau$ values.

## RESULTS AND DISCUSSION

**Principal Component Analysis Plots.** Plotted in Figure 2 is the score plot of the primary region Zakynthos with the first

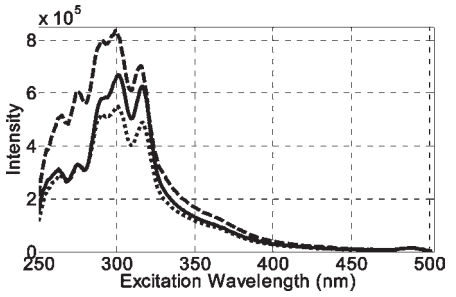


**Figure 1.** Regional mean synchronous fluorescence spectra: Zakynthos (solid line); Attica (dashed line); local shop (dotted line).

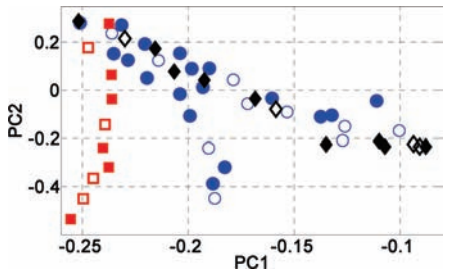


**Figure 2.** Score plot of PC2 against PC1 based on the uncentered Zakynthos samples adulterated with sunflower oil (blue circles). Attica (red squares) and local shop (black diamonds) samples are projected into the Zakynthos score space. Solid symbols are respective validation samples. For Attica and local shop samples, open symbols are the respective random samples selected for **M**.

principal component (PC) plotted against the second. The first and second PCs account for 70.71 and 17.38% of the variance in **X**, respectively. The Attica and local shop samples are projected into the primary score space. The score plot reveals that the local shop score values overlap with the Zakynthos sample score values. Hence, it is expected that the local shop samples should be fairly well predicted by a calibration based on samples from Zakynthos. Thus, model updating may not be necessary for the local shop samples. Conversely, samples from Attica are not expected to be accurately predicted as most of the sample score values are outside the Zakynthos samples. It is anticipated that model updating will allow new regional samples to be predicted with reduced errors.

**TR 2 and 2b Meta-parameter Selection.** The L-curve approach to meta-parameter selection is well documented (see refs 34−37 and references therein). Briefly, a prediction variance indicator such as the model 2-norm is plotted against a prediction error (bias) merit such as the root-mean-square error of calibration (RMSEC) or the RMSE for the samples in **M** (RMSEM). In such a plot, an L-shaped curve is obtained with the better models being in the corner region of the L-curve. These models represent a good variance/bias trade-off. The specific approach to TR 2 and TR 2b is the same and described in refs 34 and 36. It is based on the agreement of the RMSEC and the RMSEM plots. Examination of mean RMSEC and RMSEM values plotted against the range of $\lambda$ values and the mean model 2-norms finds an agreement in the bias/variance trade-off to determine a proper value for $\eta$. Once a proper $\eta$ value is determined, the $\lambda$ in the corner region of the corresponding $\eta$ L-curve with $\lambda$ as points on the curve is selected.

**TR 2-1 Meta-parameter Selection.** The L-curve protocol for meta-parameter selection is described in ref 35. The L-curve plots

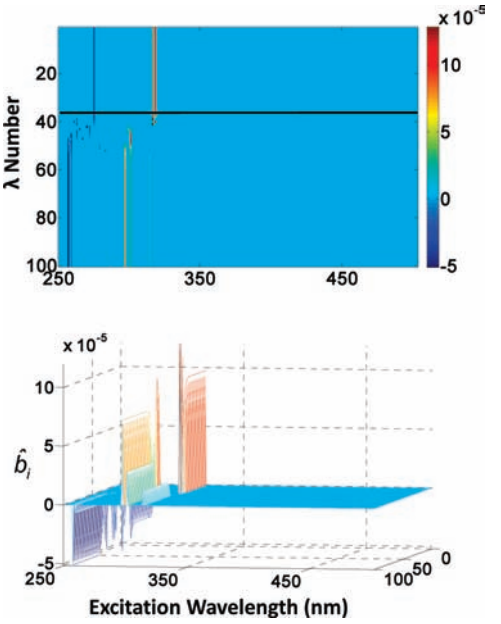


**Figure 3.** Attica TR 2-1 updated model vectors at a fixed $\tau$ value. As the $\lambda$ number decreases from bottom to top, the $\lambda$ values increase. The line indicates the $\lambda$ value selected. See text for selection process.

of the model 1-norm against RMSEC and RMSEM are formed where each L-curve represents a different $\lambda$ value and each point on a curve is a LAR model or $\tau$ value. As $\lambda$ increases in value, putting more weight on the secondary domain represented by **M**, L-curves move farther to the left and the RMSEM reduces. The opposite results for RMSEC L-curves. The first meta-parameter to select is $\tau$ (LAR model). The LAR model is selected from the corner region of the farthest left RMSEM L-curve. After the LAR model is selected, the $\lambda$ value is selected with the fixed $\tau$.

Shown in Figure 3 are the model vectors for each $\lambda$ at the selected $\tau$. Figure 3 results from updating the Zakynthos primary regional calibration model to the extra virgin olive oil samples from Attica. In Figure 3, the $\lambda$ number decreases from bottom to top, but the actual value of $\lambda$ increases. When $\lambda$ is close to zero, the model is good at predicting the primary set (small RMSEC) and not good at predicting the secondary (larger RMSEM and root-mean-square error of validation (RMSEV)). As $\lambda$ increases in value, the updated model essentially retains its shape with only a change in the model vector 1-norm (from the 100th to the 55th $\lambda$ (0 to 0.0350) in Figure 3). The model vector changes shape as it starts transitioning to better predict the secondary condition (from the 54th to the 38th $\lambda$ (0.0492 to 11.795)). After this, the model vector has converged to a shape that best predicts the secondary condition. The value for $\lambda$ is selected as the smallest value at which the model vector converges to the updated shape. This is shown in Figure 3 at the $\lambda$ with the line (the 37th $\lambda$ (16.611)). With increasing $\lambda$ values (36th to 1st (23.395 to 3.751e6)), model vectors primarily increase in 1-norm and respective RMSEM values slightly decrease. Selecting the point of convergence guards against overfitting the model to **M**.

**Validation of Zakynthos Primary Calibration.** Shown in Figure 4 are three model vectors from the primary calibration of the Zakynthos regional samples. Respective L-curves were used to select meta-parameters. One model vector is a RR model. The second one is based on reusing the plotted RR model as the diagonal of **L** in expression 6. Because there is no calibration

maintenance and transfer occurring in forming this model, the third term is removed. This model will be referred to as RRb. The third model is a TR LASSO model using expression 5 with the third term removed, that is, only wavelength selection and, again, no model updating. These three model vectors are for comparison to models updated to the secondary regions and ascertaining model changes from the updating process. In Table 1 are the corresponding validation prediction errors. If the only interest is predicting samples from Zakynthos, then RR provides the best results. Apparently, using the wavelength selection approaches does not provide lower prediction errors. The $R^2$, slope, and intercept values listed in Table 1 result from plotting model-predicted sunflower oil concentrations against nominal values. The decreasing slope values observed for the wavelength selection approaches RRb and TR LASSO add a bias (offset) to predicting the validation samples. Wavelengths spanning interferent information may have been down-weighted too much, and additional calibration samples could be included.

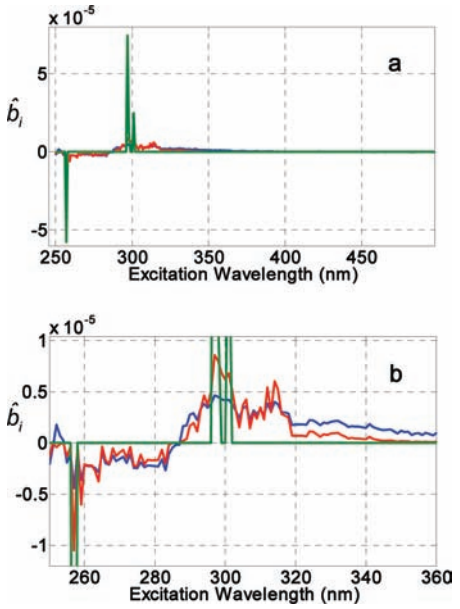


**Figure 4.** Model vectors for the Zakynthos region (a) and zoom (b): RR (blue); RRb (red); TR LASSO (green).

**Table 1. Zakynthos Primary Calibration Predicting Zakynthos Validation Samples**

| method | RMSEV | $R^2$ | slope | intercept | $\eta$ or $\tau$ | $\|\|\hat{b}\|\|_2$ |
|---|---|---|---|---|---|---|
| RR | 4.324 | 0.977 | 0.992 | 2.000 | 3.41e5 | 2.47e−5 |
| RRb | 4.426 | 0.975 | 0.963 | 3.250 | 1.51 | 2.86e−5 |
| TR LASSO | 5.754 | 0.961 | 0.939 | 5.289 | 5.83e5 | 9.79e−5 |

As noted in refs 2 and 17, the primary nonzero regions of mean regional spectra in Figure 1 and model vectors in Figure 4 are due to tocopherol and phenolic compounds present in extra virgin olive oil and sunflower oil. The TR LASSO model does emphasize the RRb bands of greatest intensity indicating that, perhaps, specific tocopherol and phenolic variances are emphasized.

**TR Updating Analysis.** Respective meta-parameter values are selected from L-curves, and shown in Figure 5 are the resultant updated local shop and Attica model vectors. Listed in Table 2 are model evaluation values for the local shop and Attica validation samples. Included are sample predictions without any model updating for comparison with updating. Compared to no updating, each updating method decreases the RMSEV. As expected from the score plot in Figure 2, the greatest decrease is for the samples from Attica.

For both secondary conditions, it is observed that the two respective RR (no model updating) $R^2$ values are as good as or nearly the same as the $R^2$ values with updating. However, the larger RMSEV values with no updating stem from the intercept term being greater than those with updating. Thus, there is a larger bias in the RR predictions with no updating. In summary, the TR methods are able to update a model built for one geographical region to predict samples from another geographical region.

Similar to Figure 4 and as expected, the TR 2-1 local shop updated model vector keys in on only a few wavelengths and the TR 2b model is less broad-banded compared to the TR 2 model. The TR 2 local shop updated model vector is similar in shape to the Zakynthos RR model except in the 250 nm region. Compared to the TR 2 local shop model, the TR 2b local shop updated model has increased emphasis in the 300 nm relative to the 315 nm region. Contrarily, the TR 2-1 local shop updated model vector shows an emphasis on the 315 nm region compared to the 300 nm region. The TR 2-1 model also selects the 275 and 315 nm regions. The shifts in wavelength emphases for the respective local shop updated models compared to the Zakynthos primary calibration models shown in Figure 4 may explain the reduction in prediction error. That is, wavelengths more specific to regional tocopherol and phenolic compounds are emphasized. Additionally, it does not appear that selecting wavelengths provides any distinct advantages in model updating for this data set; that is, the TR 2 model predicts as well as the TR 2b and TR 2-1 models.

For model updating to the Attic region, similar trends are observed, albeit, some different spectral regions are emphasized

**Table 2. Updated Zakynthos Models Predicting Local Shop and Attica Validation Samples**

| method | region | RMSEV | $R^2$ | slope | intercept | $\eta$ or $\tau$ | $\lambda$ | $\|\|\hat{b}\|\|_2$ |
|---|---|---|---|---|---|---|---|---|
| TR 2 | local shop | 1.545 | 0.998 | 1.023 | −1.418 | 9.54e5 | 1.51 | 2.09e−5 |
| TR 2b | local shop | 1.564 | 0.998 | 1.013 | −1.062 | 2.30 | 1.07 | 2.29e−5 |
| TR 2-1 | local shop | 1.955 | 0.999 | 0.957 | 0.640 | 2.84e7 | 257.12 | 8.79e−5 |
| RR[a] | local shop | 6.040 | 0.999 | 1.174 | −9.047 | | 3.41e5 | 2.47e−5 |
| TR 2 | Attica | 1.746 | 0.998 | 0.977 | −0.308 | 6.77e5 | 23.40 | 3.86e−5 |
| TR 2b | Attica | 1.971 | 0.998 | 0.991 | −1.121 | 3.00 | 32.95 | 4.52e−5 |
| TR 2-1 | Attica | 2.240 | 0.998 | 1.005 | −1.920 | 5.43e6 | 16.61 | 1.19e−4 |
| RR[a] | Attica | 14.871 | 0.974 | 0.885 | 18.121 | | 4.41e5 | 2.47e−5 |

[a] Zakynthos as calibration samples and no updating.


dx.doi.org/10.1021/jf1038053 |J. Agric. Food Chem. 2011, 59, 1051–1057

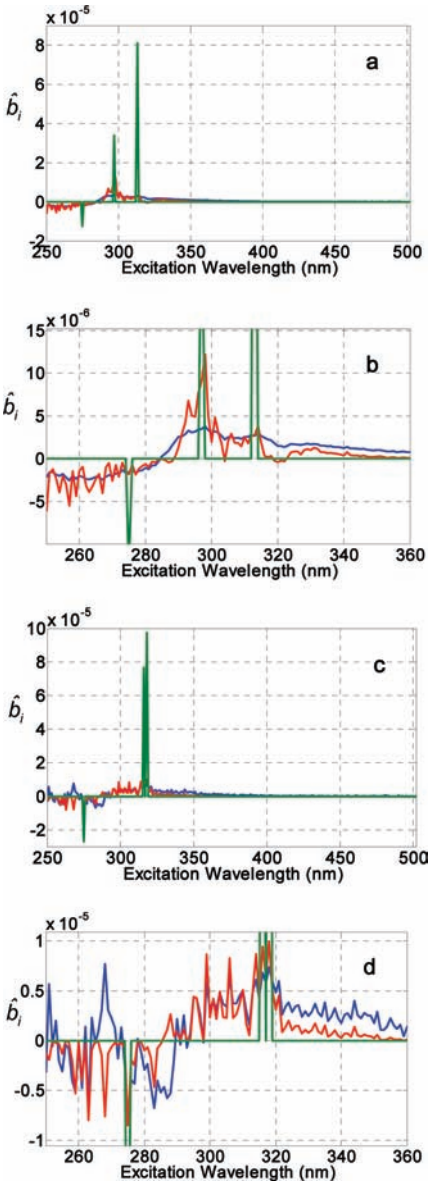


**Figure 5.** Updated TR model vectors for the local shop (a) with zoom (b) and Attica region (c) with zoom (d): TR 2 (blue); TR 2b (red); TR 2-1 (green).

and deemphasized. Most observable are the more jagged peaks in the TR 2 and TR 2b updated model vectors and, hence, much less broad-banded. Again, the differences may be indicative of regional differences in tocopherol and phenolic compounds, the primary sources of spectral responses in this wavelength range. The Attica region has the greatest changes from the Zakynthos RR and RRb models as the Attica samples are more uniquely different from the Zakynthos samples as noted by the previous score plot. As with the Zakynthos and local shop samples, wavelength selection does not provide any unique advantages for the Attica data set.

Although the TR approaches are applied to synchronous fluorescence spectra for determining sunflower oil adulteration concentration, the methods are general and can be used with other spectroscopic methods and adulterants as well as other agriculture products. Additionally, the methods should be applicable to updating models to new geographical regions, harvest years, etc., where the primary model distinguishes adulterated samples from pure extra virgin olive oil. Such a study is currently under investigation. The TR methods should also be applicable to models providing geographical classification where the models need to be made to work in new secondary conditions such as a new growing season and/or instrument.

## AUTHOR INFORMATION

### Corresponding Author

*Phone: (208) 282-2726. Fax: (208) 282-4373. E-mail: kalijohn@isu.edu.

### Funding Sources

This material is based upon work supported by the National Science Foundation under Grant CHE-0715149 (cofunded by the MPS Chemistry and DMS Statistics Divisions, by the MSPA Program, and the Office of International Science and Engineering), which we gratefully acknowledge. The research was also partially supported by Grant F09-12 from the Undergraduate Research Committee, Idaho State University. The synchronous fluorescence spectra were acquired under financial support of the Greek General Secretariat of Research and Technology and Minerva SA (Athens, Greece) through a PENED 2003 grant.

## REFERENCES

(1) *Handbook of Olive Oil Analysis and Properties*; Harwood, J., Aparicio, R., Eds.; Aspen Publishers: Gaithersburg, MD, 2000.

(2) Poulli, K. I.; Mousdis, G. A.; Georgiou, C. A. Synchronous fluorescence spectroscopy for quantitative determination of virgin olive oil adulteration with sunflower. *Anal. Bioanal. Chem.* **2006**, *386*, 1571–1575.

(3) Poulli, K. I.; Mousdis, G. A.; Georgiou, C. A. Rapid synchronous fluorescence method for virgin olive oil adulteration assessment. *Food Chem.* **2007**, *105*, 369–375.

(4) López-Díez, E. C.; Bianchi, G.; Goodacre, R. Rapid quantitative assessment of the adulteration of virgin olive oils with hazelnut oils using Raman spectroscopy and chemometrics. *J. Agric. Food Chem.* **2003**, *51*, 6145–6150.

(5) Baeten, V.; Meurens, M.; Morales, M. T.; Aparicio, R. Detection of virgin olive oil adulteration by Fourier transform Raman spectroscopy. *J. Agric. Food Chem.* **1996**, *44*, 2225–2230.

(6) Davis, A. N.; McIntyre, P.; Morgan, E. Study of the use of molecular spectroscopy for the authentication of extra virgin olive oil. Part I: Fourier transform Raman spectroscopy. *Appl. Spectrosc.* **2000**, *54*, 1864–1867.

(7) Heise, H. M.; Damm, U.; Lampen, P.; Davis, A. N.; McIntyre, P. S. Spectral variable selection for partial least squares calibration applied to authentication and quantification of extra virgin olive oils using Fourier transform Raman spectroscopy. *Appl. Spectrosc.* **2005**, *59*, 1286–1294.

(8) Lai, Y. W.; Kemsley, E. K.; Wilson, R. H. Quantitative analysis of potential adulterants of extra virgin olive oil using infrared spectroscopy. *Food Chem.* **1995**, *53*, 95–98.

(9) Lerma-García, M. J.; Ramis-Ramos, G.; Herrero Martínez, J. M.; Simó-Alfonso, E. F. Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy. *Food Chem.* **2010**, *118*, 78–83.

(10) Gurdeniz, G.; Ozen, B. Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data. *Food Chem.* **2009**, *116*, 519–525.

(11) Ozen, B. F.; Mauer, L. J. Detection of hazelnut oil adulteration using FT-IR spectroscopy. *J. Agric. Food Chem.* **2002**, *50*, 3898–3901.

(12) Wesley, I. J.; Pacheco, F.; McGill, A. E. J. Identification of adulterants in olive oils. *J. Am. Oil Chem. Soc.* **1996**, *73*, 515–518.

(13) Wesley, I. J.; Barnes, R. J.; McGill, A. E. J. Measurement of adulteration of olive oils by near-infrared spectroscopy. *J. Am. Oil Chem. Soc.* **1995**, *72*, 289–292.

(14) Downey, G.; McIntyre, P.; Davies, A. N. Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the eastern Mediterranean by visible and near-infrared spectroscopy. *J. Agric. Food Chem.* **2002**, *50*, 5520–2225.

(15) Mavromoustakos, T.; Zervou, M.; Theodoropoulou, E.; Panagiotopoulos, D.; Bonas, G.; Day, M.; Helmis, A. $^{13}$C NMR analysis of the triacylglycerol composition of Greek virgin olive oils. *Magn. Reson. Chem.* **1997**, *35*, S3–S7.

(16) Tsimidou, M.; Karakostas, K. X. Geographical classification of Greek virgin olive oil by non-parametric multivariate evaluation of fatty acid composition. *J. Sci. Food Agric.* **1993**, *62*, 253–257.

(17) Dupuy, N.; Le Dréau, Y.; Ollivier, D.; Artaud, J.; Pinatel, C.; Kister, J. Origin of French virgin olive oil registered designation of origins predicted by chemometric analysis of synchronous excitation-emission fluorescence spectra. *J. Agric. Food Chem.* **2005**, *53*, 9361–9368.

(18) Tapp, H. S.; Defernez, M.; Kemsley, E. K. FTIR spectroscopy and multivariate analysis can distinguish the geographic origin of extra virgin olive oils. *J. Agric. Food Chem.* **2003**, *51*, 6110–6115.

(19) Dupuy, N.; Galtier, O.; Ollivier, D.; Vanloot, P.; Artaus, J. Comparison between NIR, MIR, concatenated NIR and MIR analysis and hierarchical PLS model. Application to virgin olive oil analysis. *Anal. Chim. Acta* **2010**, *666*, 23–31.

(20) Woodcock, T.; Downey, G.; O'Donnell, C. P. Confirmation of declared provenance of European extra virgin olive oil samples by NIR spectroscopy. *J. Agric. Food Chem.* **2008**, *56*, 11520–11525.

(21) Casale, M.; Casolino, C.; Ferrari, G.; Forina, M. Near infrared spectroscopy and class modeling techniques for the geographical authentication of Ligurian extra virgin olive oils. *J. Near Infrared Spectrosc.* **2008**, *16*, 39–47.

(22) Mannina, L.; Marini, F.; Gobbino, M.; Sobolev, A. P.; Capitani, D. NMR and chemometrics in tracing European olive oils: the case study of Ligurian samples. *Talanta* **2010**, *80*, 2141–2148.

(23) Sacchi, R.; Mannina, L.; Fiordiponti, P.; Barone, P.; Paolillo, L.; Patumi, M.; Segre, A. Characterization of Italian extra virgin olive oils using $^{1}$H-NMR spectroscopy. *J. Agric. Food Chem.* **1998**, *46*, 3947–3951.

(24) Mannina, L.; Paumi, M.; Proietti, N.; Bassi, D.; Segre, A. L. Geographical characterization of Italian extra virgin olive oils using high-field $^{1}$H-NMR spectroscopy. *J. Agric. Food Chem.* **2001**, *49*, 2687–2696.

(25) Vlahov, G.; Del Re, P.; Simone, N. Determination of geographical origin of olive oils using $^{13}$C nuclear magnetic resonance spectroscopy. I − Classification of olive oils of the Puglia region with denomination of protected origin. *J. Agric. Food Chem.* **2003**, *51*, 5612–5615.

(26) Alonso-Salces, R. M.; Moreno-Rojas, J. M.; Holland, M. V.; Reniero, F.; Guillou, C.; Héberger, K. Virgin olive oil authentication by multivariate analyses of $^{1}$H NMR fingerprints and $\delta^{13}$C and $\delta^{2}$H data. *J. Agric. Food Chem.* **2010**, *58*, 5586–5596.

(27) Guimet, F.; Boqué, R.; Ferré, J. Cluster analysis applied to the exploratory analysis of commercial Spanish olive oils by means of excitation−emission fluorescence spectroscopy. *J. Agric. Food Chem.* **2004**, *52*, 6673–6679.

(28) Casale, M.; Sinelli, N.; Oliveri, P.; Di Egidio, V.; Lanteri, S. Chemometrical strategies for feature selection and data compression applied to NIR and MIR spectra of extra virgin olive oils for cultivar identification. *Talanta* **2010**, *80*, 1832–1837.

(29) Sayago, A.; García-González, D. L.; Morales, M. T.; Aparicio, R. Detection of the presence of refined hazelnut oil in refined olive oil by fluorescence spectroscopy. *J. Agric. Food Chem.* **2007**, *55*, 2068–2071.

(30) Zou, M. Q.; Zhang, X. F.; Qi, X. H.; Ma, H. L.; Dong, Y.; Liu, C. W.; Guo, X.; Wang, H. Rapid authentication of olive oil adulteration by Raman spectrometry. *J. Agric. Food Chem.* **2009**, *57*, 6001–6006.

(31) Baeten, V.; Pierna, J. A. F.; Dardenne, P.; Meurens, M.; García-González, D. L.; Aparicio-Ruiz, R. Detection of the presence of hazelnut oil in olive oil by FT-Raman and FT-MIR spectroscopy. *J. Agric. Food Chem.* **2005**, *53*, 6201–6206.

(32) Fragaki, G.; Spyros, A.; Siragakis, G.; Salivaris, E.; Dais, P. Detection of extra virgin olive oil adulteration with lampante olive oil and refined olive oil using nuclear magnetic resonance spectroscopy and multivariate statistical analysis. *J. Agric. Food Chem.* **2005**, *53*, 2810–2816.

(33) Brown, S. D. Transfer of multivariate calibration models. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*; Brown, S. D., Tauler, R., Walczak, B., Eds.-in-Chief; Kalivas, J. H., Section Ed.; Elsevier: Amsterdam, The Netherlands, 2009; pp 345−377.

(34) Kalivas, J. H.; Siano, G. S.; Andries, E.; Goicoechea, H. C. Calibration maintenance and transfer using Tikhonov regularization approaches. *Appl. Spectrosc.* **2009**, *63*, 800–809.

(35) Kunz, M. R.; Kalivas, J. H.; Andries, E. Model updating for spectral calibration maintenance and transfer using 1-norm variants of Tikhonov regularization. *Anal. Chem.* **2010**, *89*, 3642–3649.

(36) Kunz, M. R.; Ottaway, J.; Kalivas, J. H.; Andries, E. Impact of standardization sample designs on Tikhonov regularization variants for spectroscopic calibration maintenance and transfer. *J. Chemom.* **2009**, *24*, 218–229.

(37) Ottaway, J.; Andries, E.; Kalivas, J. H. Spectral multivariate calibration with wavelength selection using variants of Tikhonov regularization. *Appl. Spectrosc.* **2010**, *64*, 1388–1395.