Fusion of Synchronous Fluorescence Spectra with Application to Argan Oil for Adulteration Analysis

Tyler D. Stokes[a], Mellou Foteini[b], Brett Brownfield[a] John H. Kalivas[a], George Mousdis[c] , Aziz Amine[d], and Constantinos Georgiou[b]

[a]Department of Chemistry, Idaho State University, 921 S. 8th Avenue, Pocatello, ID 83209, USA
[b]Chemistry Laboratory, Agricultural University of Athens, 75 Iera Odos, 11855 Athens, Greece
[c]Theoretical & Physical Chemistry Institute, National Hellenic Research Foundation, 48 Vassileous Constantinous Avenue, 11635 Athens, Greece
[d]Laboratoire Génie des Procédés et Environnement, Université Hassanll-Mohammedia, Morocco
[a]stoktyle@isu.edu, [a]kalijohn@isu.edu

Abstract

When synchronous fluorescence (SyF) spectroscopy is used for quantitative and qualitative analysis, selection of a useful wavelength interval between the excitation and emission wavelengths (Δλ) is needed. Presented is a fusion approach to combine Δλ intervals thereby negating the selection process. This study uses the fusion of SyF spectra for the quantitative analysis of the corn oil adulterant content. The SyF spectra were acquired by varying the excitation wavelength in the region 300-800 nm using Δλ wavelength intervals from 10 to 100 nm in steps of 10 nm producing 10 sets of SyF spectra. For quantitative analysis, two calibration approaches are evaluated with these 10 SyF spectral datasets. Multivariate calibration by partial least squares (PLS) and a univariate calibration process where the SyF spectra are summed over respective SyF spectral ranges, the area under the curve (AUC) method. For adulteration detection and quantitation of the corn oil, prediction errors decrease with fusion compared to individually using the 10 Δλ interval SyF spectral data sets. For this data set, the AUC method generally provides smaller prediction errors than PLS at individual Δλ intervals as well as with fusion of all 10 Δλ intervals.

Objectives

- Develop and showcase a fusion approach to bypass the Δλ selection process required when using synchronous fluorescence spectroscopy (SyF) for multivariate analysis
- Showcase an effective method for quantitating corn oil adulterants in Argan Oil

Synchronous Fluorescent Spectroscopy (SyF)

- When a SyF spectrum is measured a Δλ interval is maintained
  - Δλ interval: a linearly increasing wavelength difference between the emission and excitation wavelengths
- Currently Δλ intervals are individually evaluated and selected for multivariate analysis
  - Slows down the classification and/or calibration and prediction processes

Modeling Approach

Partial Least Squares (PLS)

$$\mathbf{y} = \mathbf{X}\mathbf{b} \rightarrow \hat{\mathbf{b}} = \mathbf{X}^{+}\mathbf{y}$$
$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$$

- $\mathbf{X}^{+}$ Is calculated by decomposing **X** and **y**
- The number of PLS models is dependent on the rank of **X**, i.e. the minimum number of rows or columns in
- Requires latent variable selection

Area Under the Curve (AUC)

$$\mathbf{y} = \mathbf{x}b \rightarrow \hat{b} = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{y}$$
$$\hat{\mathbf{y}} = \mathbf{x}\hat{b}$$

- Spectral responses are summed across the measured spectral ranges
- **X :** calibration samples
- **y :** actual analyte concentration
- $\hat{\mathbf{y}}$ **:** analyte concentration prediction
- $\hat{\mathbf{b}}$ **:** estimated model regression vector

Model Measures

- **Bias :**
  - R^2
  - Root Mean Square Error (RMSE) RMSEC, RMSEV

$$\mathrm{RMSE}_i = \sqrt{\frac{\sum_{k=1}^{n}\left(y_k - \hat{y}_k\right)^2}{n}}$$

Model Measures Cont'd

- **Variance :**
  - Euclidean 2-norm ($\|\hat{\mathbf{b}}\|_2$)
- **U-Curves :**

  Bias-variance trade-off
  - C1
  - C2

$$\mathrm{C1}_i = \left(\frac{\|\mathbf{b}_i\| - (\|\mathbf{b}\|)_{\min}}{(\|\mathbf{b}\|)_{\max} - (\|\mathbf{b}\|)_{\min}}\right) + \left(\frac{\mathbf{RMSEC}_i - \mathbf{RMSEC}_{\min}}{\mathbf{RMSEC}_{\max} - \mathbf{RMSEC}_{\min}}\right)$$

$$\mathrm{C2}_i = \left(\frac{(\|\mathbf{b}_i\|) - (\|\mathbf{b}\|)_{\min}}{(\|\mathbf{b}\|)_{\max} - (\|\mathbf{b}\|)_{\min}}\right) + \left(\frac{(1-\mathbf{R}_i^2) - (1-\mathbf{R}_{\min}^2)}{(1-\mathbf{R}_{\max}^2) - (1-\mathbf{R}_{\min}^2)}\right)$$

MV Model Selection

- The PLS approach creates numerous models with latent variables (LV).
- A set of LV must be selected to form the model from the calibration set and then predict the validation set.
- To automate the LV selection a U-curve approach was used

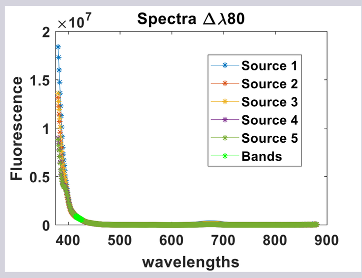
$$\text{Selected LV} = \min[\text{mean}(\mathrm{C1}+\mathrm{C2})]$$

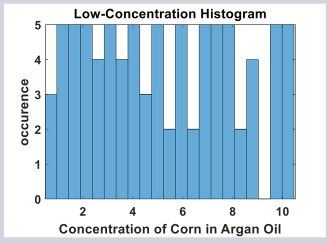
- C1 balances prediction error for the calibration set, RMSEC, in conjunction with the model regression vector **b**.
- C2 balances the calibration model fit with the variance indicator
- C1 and C2 guard against selecting an over fitted or under-fitted model

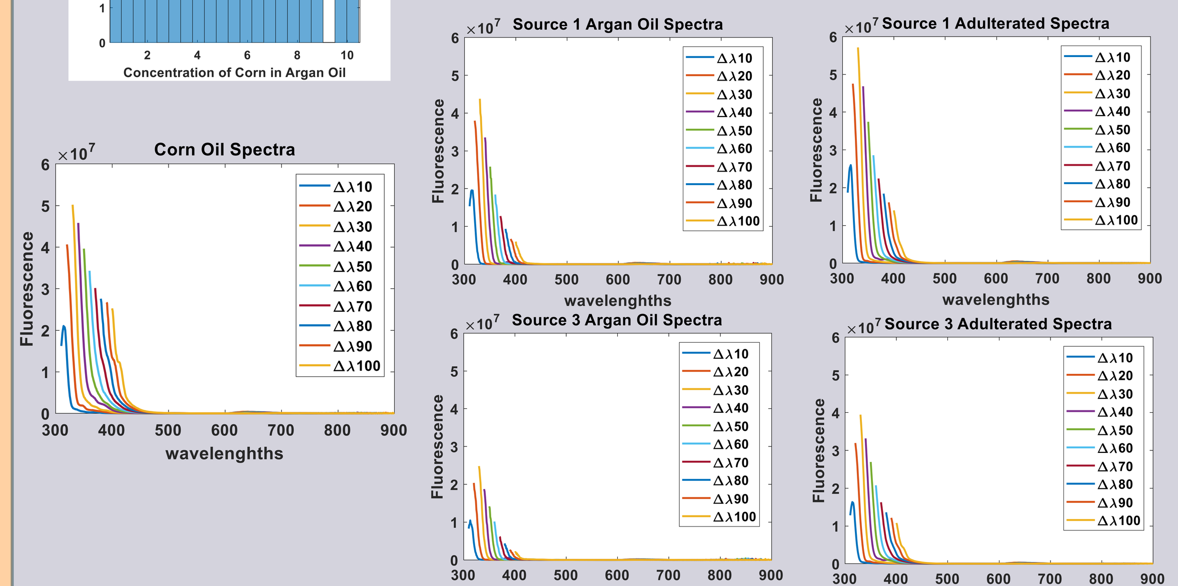
Argan Oil Data Set

- 5 sources from separate Moroccan farms
- Each source was adulterated with corn oil purchased from a local store
- Source 1: 27 samples
- Source 2: 30 samples
- Source 3: 29 samples
- Source 4: 33 samples
- Source 5: 34 samples
- SyF spectra obtained by simultaneously measuring excitation and emission wavelengths in the range 300 to 800 nm varying the Δλ interval from 10 to 100 nm in increments of 10 nm.

Wavelengths (nm)		Δλ Interval
Full Range	Selected Band	
310-710	365-375	10
320-720	345-355	20
330-730	360-370	30
340-740	360-370	40
350-750	380-390	50
360-760	385-395	60
370-770	410-420	70
380-780	415-425	80
390-790	415-425	90
400-800	435-445	100

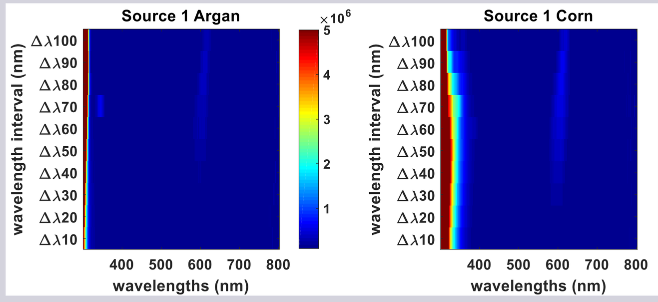






- Each source is from a unique farm in Morocco
- Argan spectra differentiates from Corn spectra

Total Synchronous Fluorescence Spectra Potential



- Argan Excitation Region: 300 – 330 nm
- Corn Excitation Region: 300 – 410 nm
- Potential for differentiation within 300 – 410 nm range

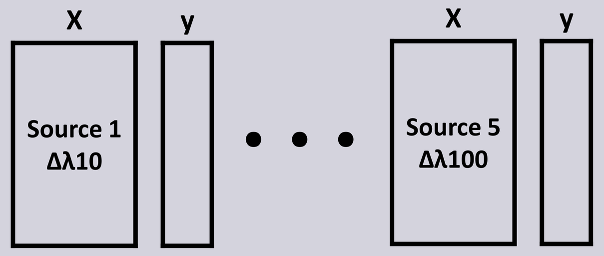
Organizational Techniques

1. **Δλ-Wise, Source-Wise**: Δλ and sources separate
2. **All Δλ, Source-Wise**: Δλ combined and sources separate
3. **Δλ-Wise, All Sources**: Δλ separate and sources combined
4. **All Δλ, All Sources**: Δλ combined and sources combined

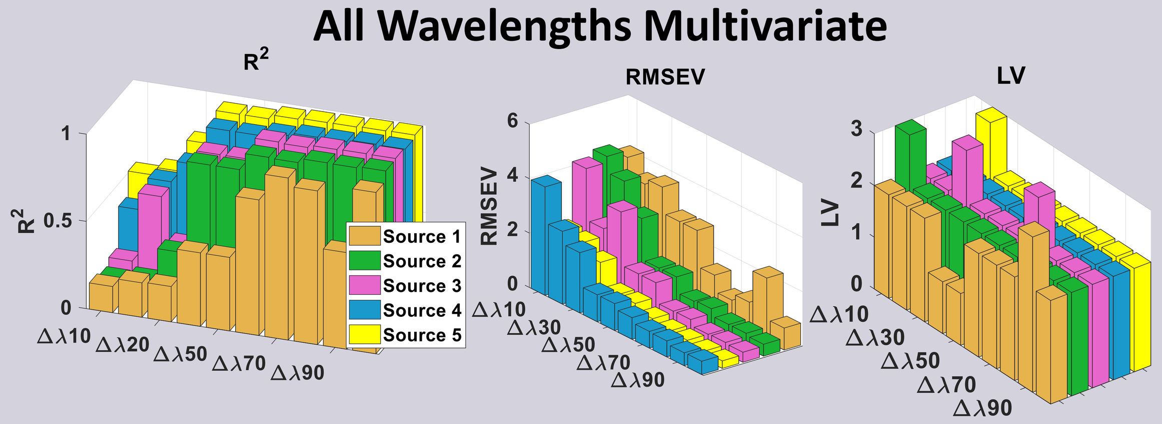
Experimental Results

(1) Δλ-Wise, Source-Wise

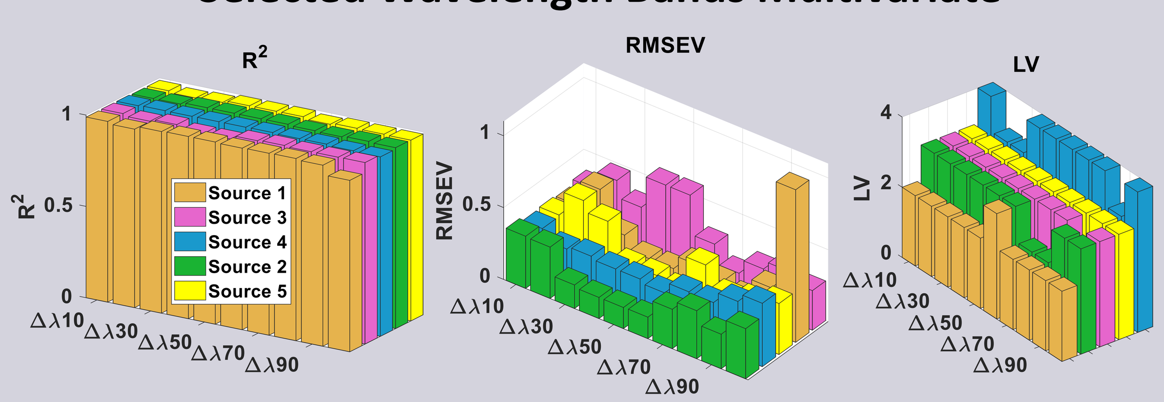
Δλ separate and Sources separate



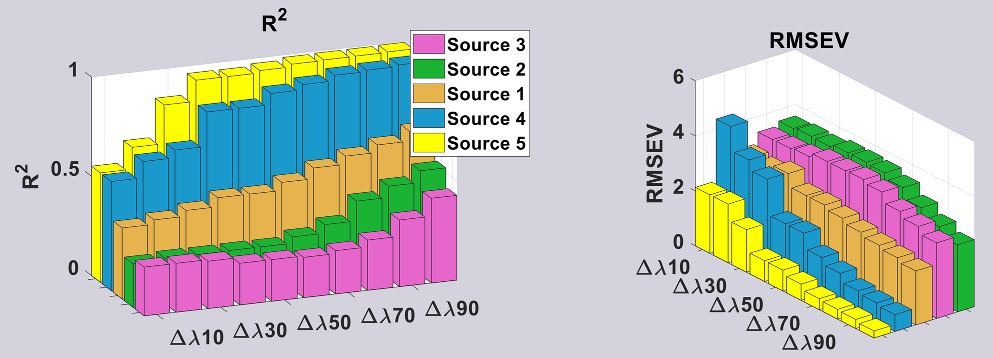
All Wavelengths Multivariate



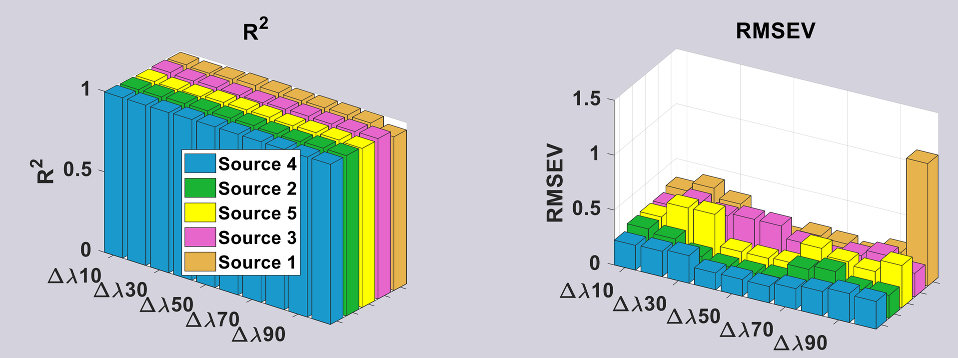
Selected Wavelength Bands Multivariate



All Wavelengths Area Under the Curve

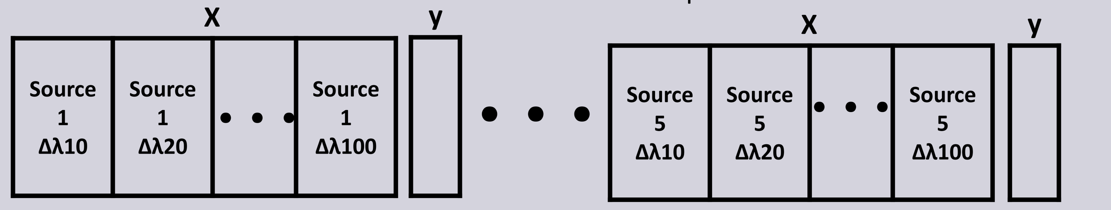


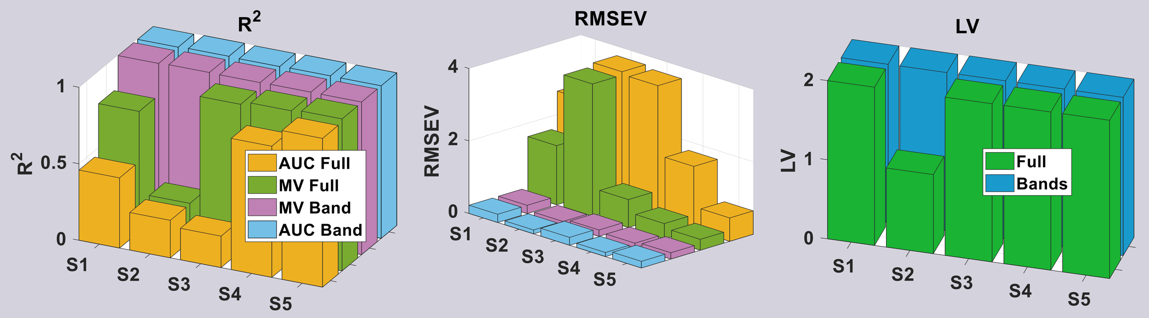
Selected Wavelength Bands Area Under the Curve



(2) All Δλ, Source-Wise

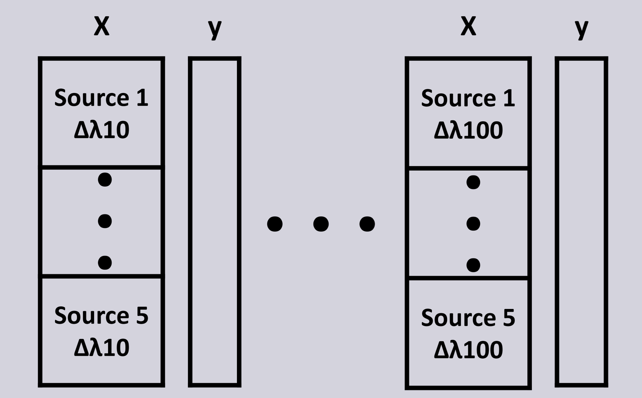
Δλ combined and Sources separate

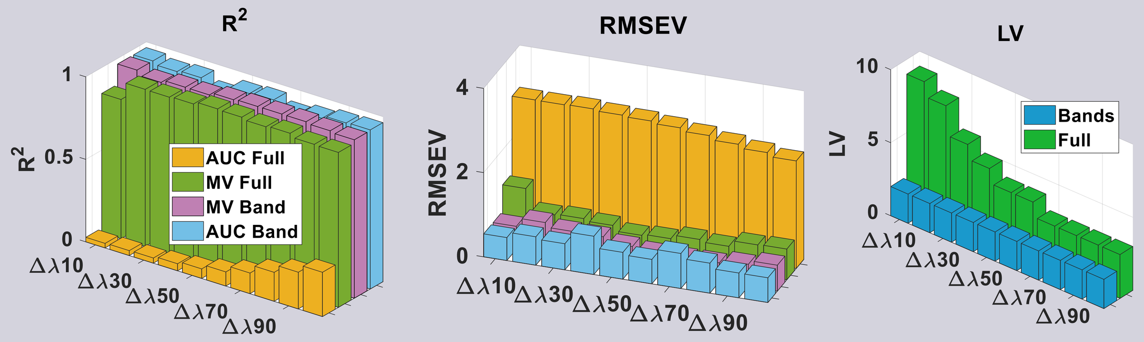




(3) Δλ-Wise, All Sources

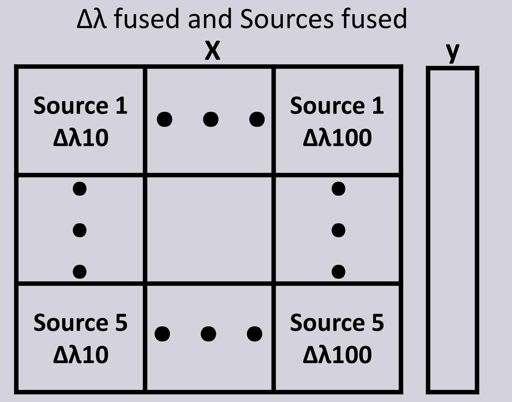
Δλ separate and Sources combined

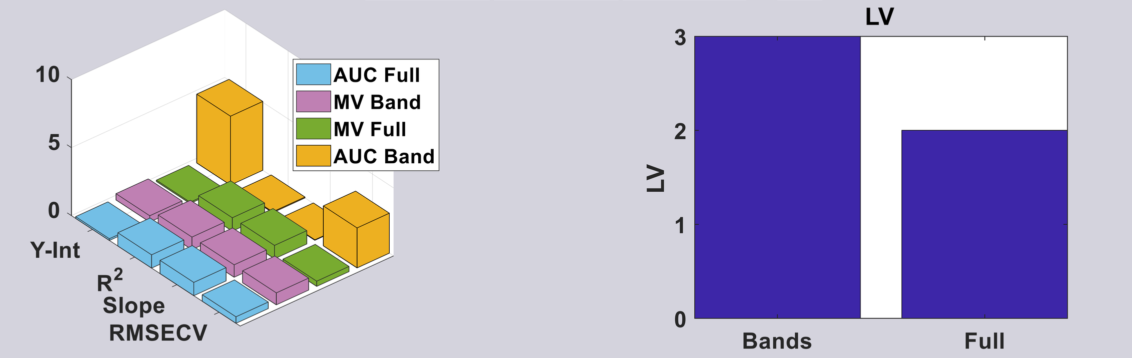




(4)All Δλ, All Sources

Δλ fused and Sources fused





Conclusions

- Low level fusion provided lower prediction errors
- The bands provided more consistent prediction errors for individual Δλ
  - Full wavelength results were Δλ independent when the sources are fused
- AUC performed better than PLS at the bands, in general
- Low level fusion was showcased to be an effective means to bypass the Δλ selection process

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. CHE-1506417 (co-funded by CDS&E) and is gratefully acknowledged by the authors.