Spectral unmixing in Chemometrics

Workshop

DENNIS SEMYALO

Dept. Smart Agricultural Systems, Chungnam Natl. University

December 5, 2024

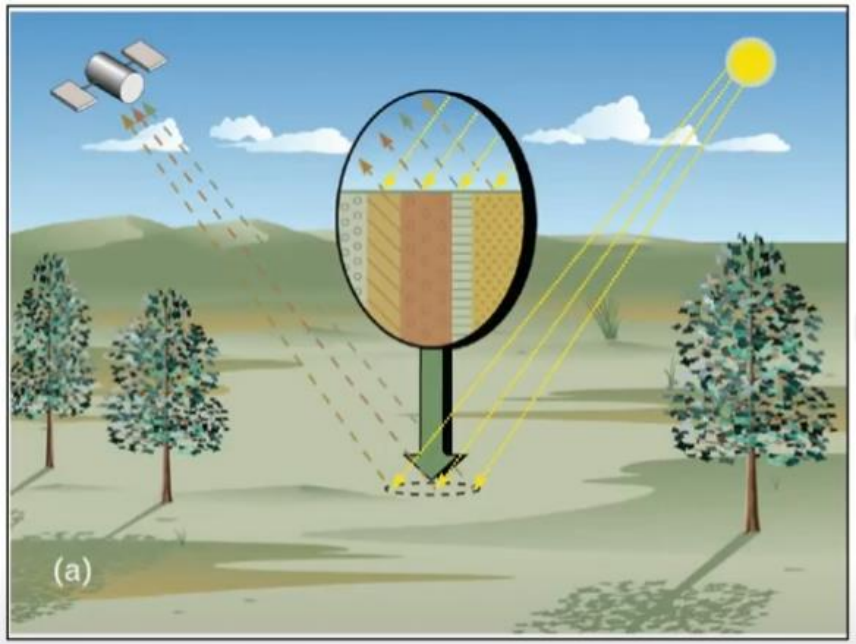
Overview

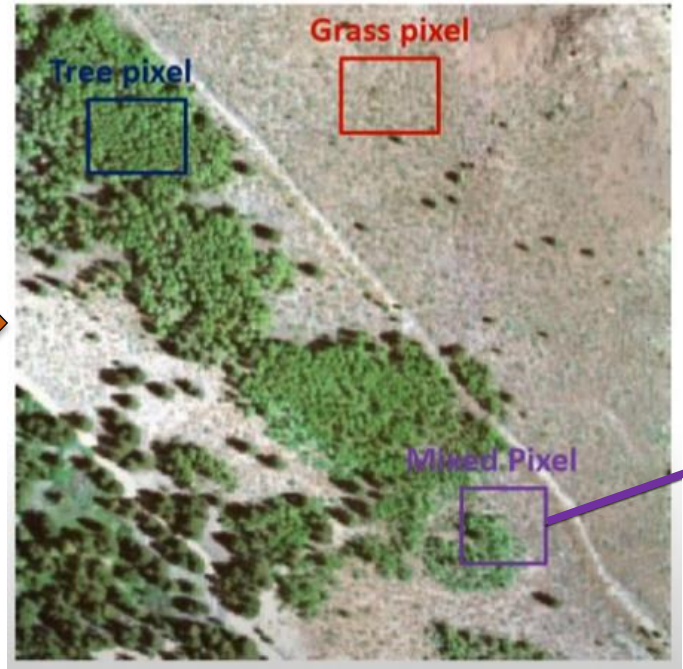
- Introduction
- Spectral unmixing technique
- Applications of spectral unmixing in chemometrics
- Spectral unmixing hands-on tutorial in Python
- Conclusions
- References

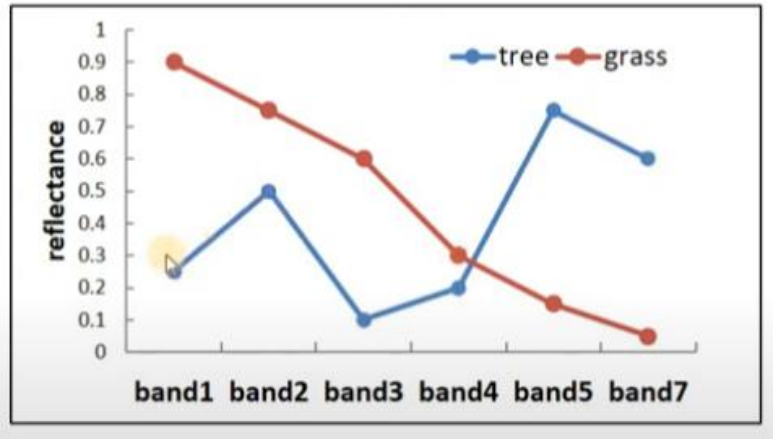
Introduction

- Spectral unmixing separates pure spectral signatures of constituents and their distribution in mixture measurements/samples.
- Useful in remote sensing, chemometrics, & environmental data applications

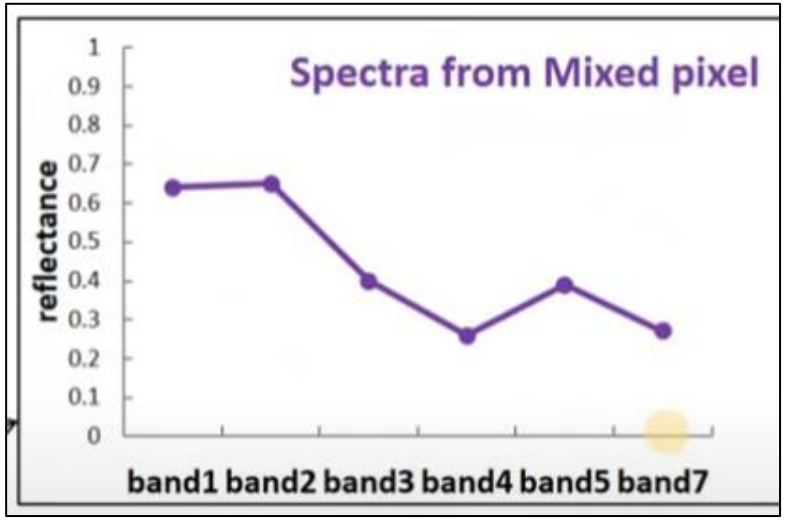
Remote sensing







Spectral unmixing



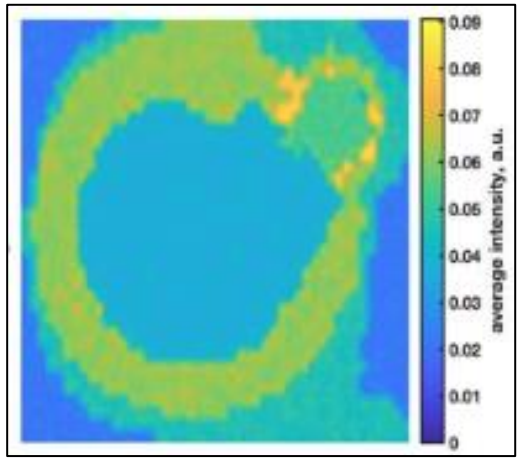
Common terms - synonyms

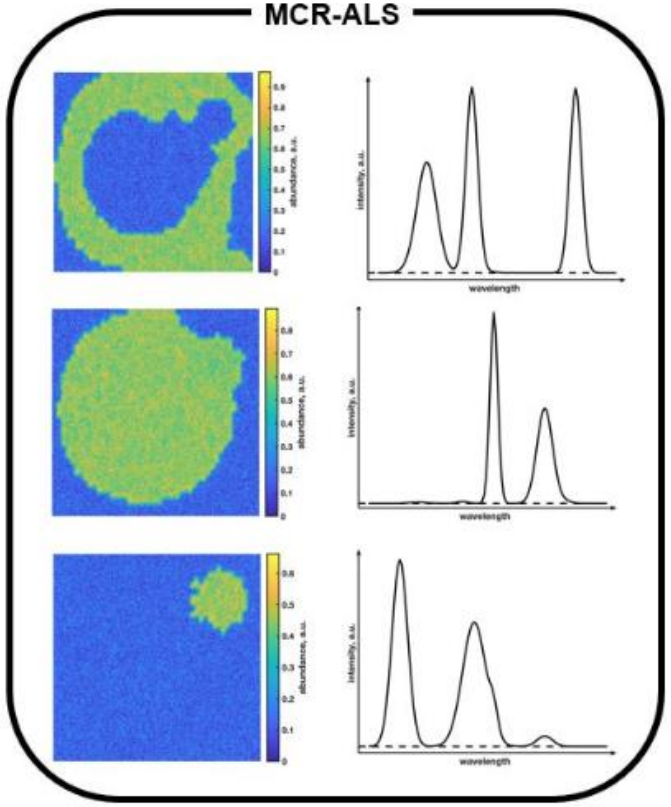
- Pure components/ pure analytes/pure constituents
- Pure spectral signatures/endmembers/ constituent spectra/ spectra profile
- Concentration/fractions of endmembers/abundances/ distribution profile

Introduction

Chemometrics

- Application in chemometrics includes the use of spectroscopy, & hyperspectral imaging on chemical mixture samples
- Applicable on variety of spectroscopies such as infrared, ultraviolet-visible (UV-Vis), near-infrared (NIR), mass spectrometry, etc.
- Spectral unmixing decomposes hyperspectral images into distribution maps and spectral profiles associated with all the constituents in the samples (Badaró et al., 2021; De Juan et al., 2014).





Pure component spectra and distribution maps by spectral unmixing of simulated hyperspectral image

Spectral unmixing technique

Algorithm: Multivariate Curve Resolution Alternating Regression(MCR-AR)

- MCR also known as self-modeling mixture analysis (SMMA) enables spectral unmixing in chemometrics
- Mathematically, MCR methods are based on a bilinear model
- Eq. (1) solved using an iterative alternating regression (AR) scheme; fix C, calculate S, and vice versa
- Regressors include OLS (MCR-ALS), NNLS (MCR-NNLS), Lasso (MCR-Lasso), Ridge (MCR-Ridge), etc.

$$\mathbf{D} = \boldsymbol{C}\boldsymbol{S}^{\boldsymbol{T}} + \boldsymbol{\varepsilon} \qquad (1)$$

$$| \mathbf{D} \in \cup^{m,n} \; | \; \mathbf{C} \in \cup^{m,p} \; | \; \mathbf{S} \in \cup^{n,p} \; | \; \boldsymbol{\varepsilon} \in \cup^{m,n} \; |$$

- *D : 2D matrix built by unfolding the original hypercube,*
- *C: concentration matrix,*
- S^T *: matrix containing pure spectra information, or signatures*
- ε*: matrix expressing noise, error, or variance unexplained by the bilinear model (Olmos et al., 2017).*
- *m: number of independent measurements (rows/observations)*
- *n: number of elements in single measurements (columns/wavelengths)*
- *p: pure components/analytes*
- ∪*: universal set of numbers (real, imaginary, or complex)*

$$\boldsymbol{C}^{[k+1]} = \underset{\boldsymbol{C}^{[k]}}{\boldsymbol{argmin}}\, \boldsymbol{Q_c}\left(\boldsymbol{C}^{[k]}, \boldsymbol{S}^{[k]}\right) \qquad (2)$$

$$\boldsymbol{S}^{[k+1]} = \underset{\boldsymbol{S}^{[k]}}{\boldsymbol{argmin}}\, \boldsymbol{Q_s}\left(\boldsymbol{C}^{[k+1]}, \boldsymbol{S}^{[k]}\right) \qquad (3)$$

$$\boldsymbol{MSE} = \sum \left| \boldsymbol{D} - \boldsymbol{C}^{[k]} \left(\boldsymbol{S}^{[k]}\right)^{\boldsymbol{T}} \right|^{\boldsymbol{2}} \qquad (4)$$

- *k : iteration number*
- Q_c *: objective function for C*
- Q_s *: objective function for S*

Spectral unmixing technique

Algorithm: Multivariate Curve Resolution Alternating Regression(MCR-AR)

Inputs: $\mathbf{C}^{[0]}$ or $\mathbf{S}^{[0]}$; $\mathbf{D}$; Q; L_C; L_S; L_B

1: **for** $k \leftarrow 0$ to k_{max} **do**
2: **if** $k > 0$ **or** $\mathbf{S}^{[0]}$ inputted **then**
3: $\mathbf{C}^{[k+1]} \leftarrow \underset{\mathbf{C}^{[k]}}{\operatorname{argmin}}\, Q(\mathbf{C}^{[k]}, \mathbf{S}^{[k]})$
4: $\mathbf{C}^{[k+1]} \leftarrow L_C\{\mathbf{C}^{[k+1]}\}$
5: **if** $L_B\left(\mathbf{C}^{[\cdot]}, \mathbf{S}^{[\cdot]}\right)$ is TRUE **then**
6: **break**
7: **end if**
8: **else**
9: $\mathbf{C}^{[k+1]} \leftarrow \mathbf{C}^{[k]}$
10: **end if**
11: $\mathbf{S}^{[k+1]} \leftarrow \underset{\mathbf{S}^{[k]}}{\operatorname{argmin}}\, Q(\mathbf{C}^{[k+1]}, \mathbf{S}^{[k]})$
12: $\mathbf{S}^{[k+1]} \leftarrow L_S\{\mathbf{S}^{[k+1]}\}$
13: **if** $L_B\left(\mathbf{C}^{[\cdot]}, \mathbf{S}^{[\cdot]}\right)$ is TRUE **then**
14: **break**
15: **end if**
16: **end for**

Provide initial estimate for S

Solve for C with S fixed

Apply constraints on C

Stop if the convergence criterion is met

Solve for S with C fixed

Apply constraints on S

Stop if the convergence criterion is met

Repeat loop if convergence not met

D: input data matrix

S: pure spectral profile

C: pure concentration profile

$S^{[0]}$: initial estimate for S

$C^{[0]}$: initial estimate for C

k: iteration number

Q: objective function for S & C

LS: constraint applied to S

L_C: constraint applied to C

L_B: convergence criterion

convergence criterion

- *Number of iterations meets preset maximum*
- *Error function (e.g. MSE) exceeds a defined value*
- *Relative change of MSE after each iteration*
- *Etc.*

Applications of spectral unmixing in chemometrics

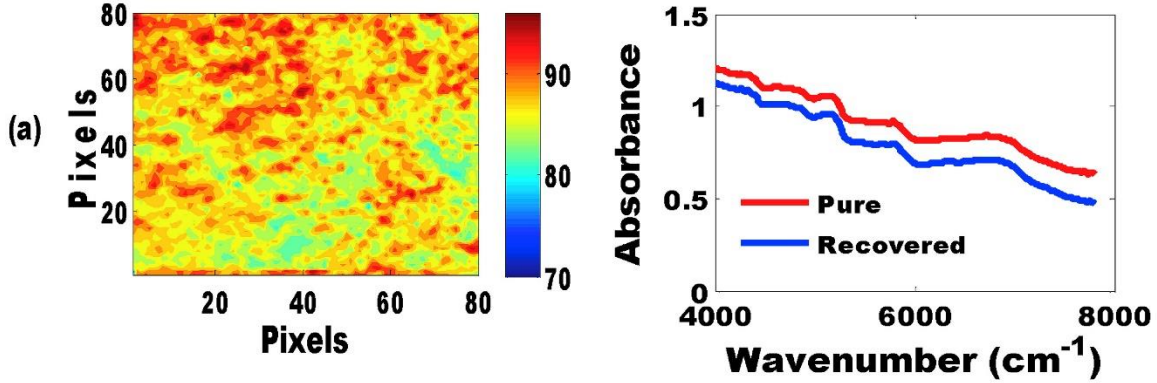
Purpose	source
Detection of adulterants (starch, urea, and melamine) in milk powder	(Forchetti & Poppi, 2017)
Fiber distribution in pasta	(Badaró et al., 2021)
Study of vegetal tissues (Lignins, Vesicles, and chlorophylls) in rice leaves	(Gómez-Sánchez et al., 2021)
Analysis of the constituents (sucrose, lactose, butter, and whey) of commercial chocolate samples	(Zhang et al., 2015)
Evaluate glucosinolate level to detect postharvest senescence in broccoli	(Guo et al., 2022)
Detection of blood in fish muscle	(Skjelvareid et al., 2017)
Distribution of tissue cryosections of the zebrafish head, including eyes, brain, and mouth	(Olmos et al., 2017)

- Some studies used spectral unmixing for feature extraction purposes (Guo et al., 2022).

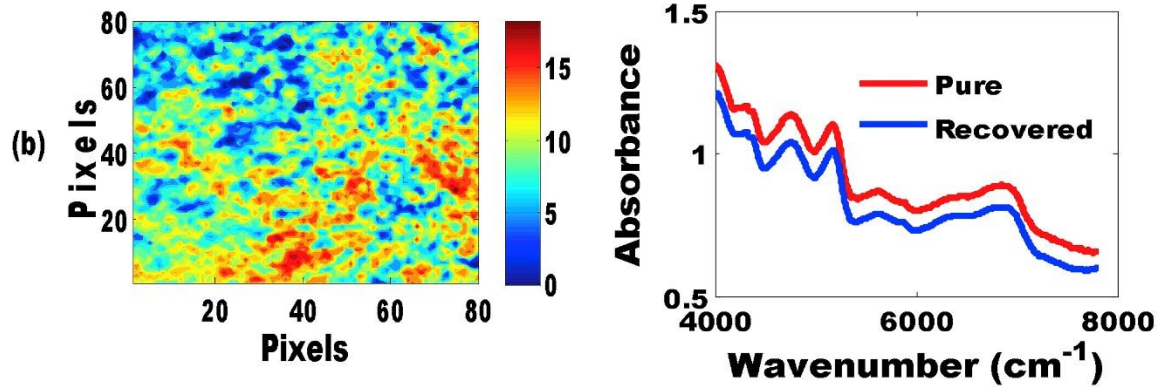
Applications of spectral unmixing in chemometrics

Detection of adulterants (starch, urea, and melamine) in milk powder (Forchetti & Poppi, 2017) : NIR-HSI

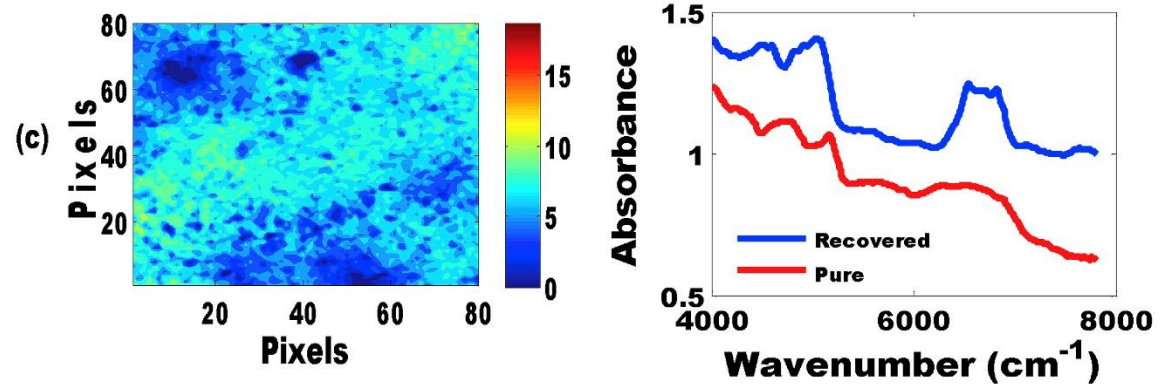
(a) milk powder



(b) starch



(c) urea

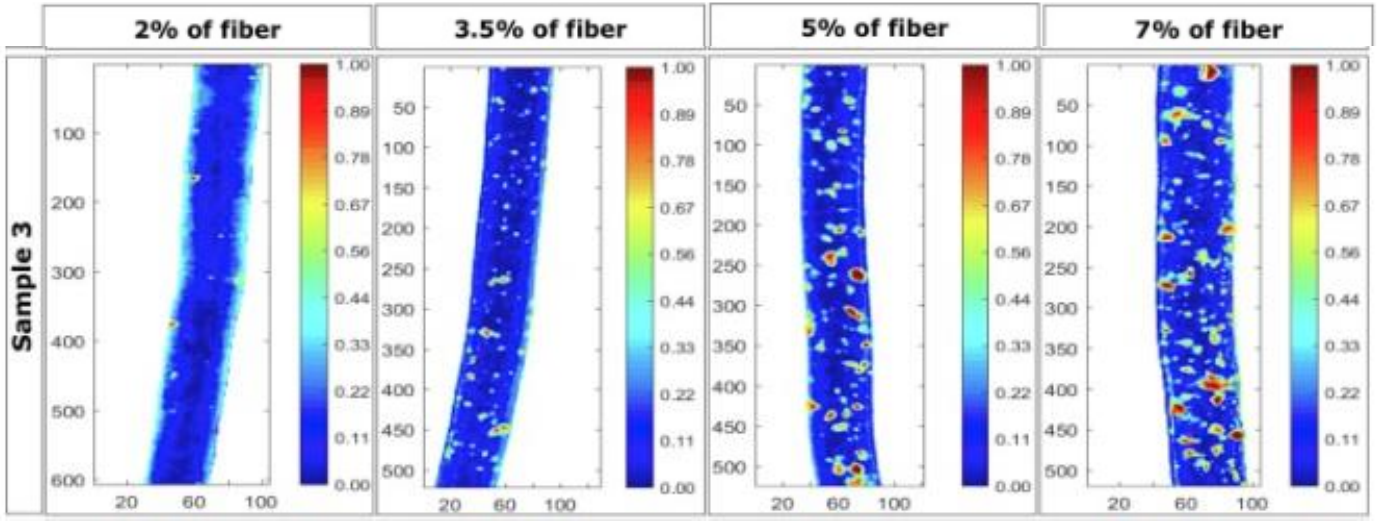


Distribution maps

Expected and recovered by MCR spectra

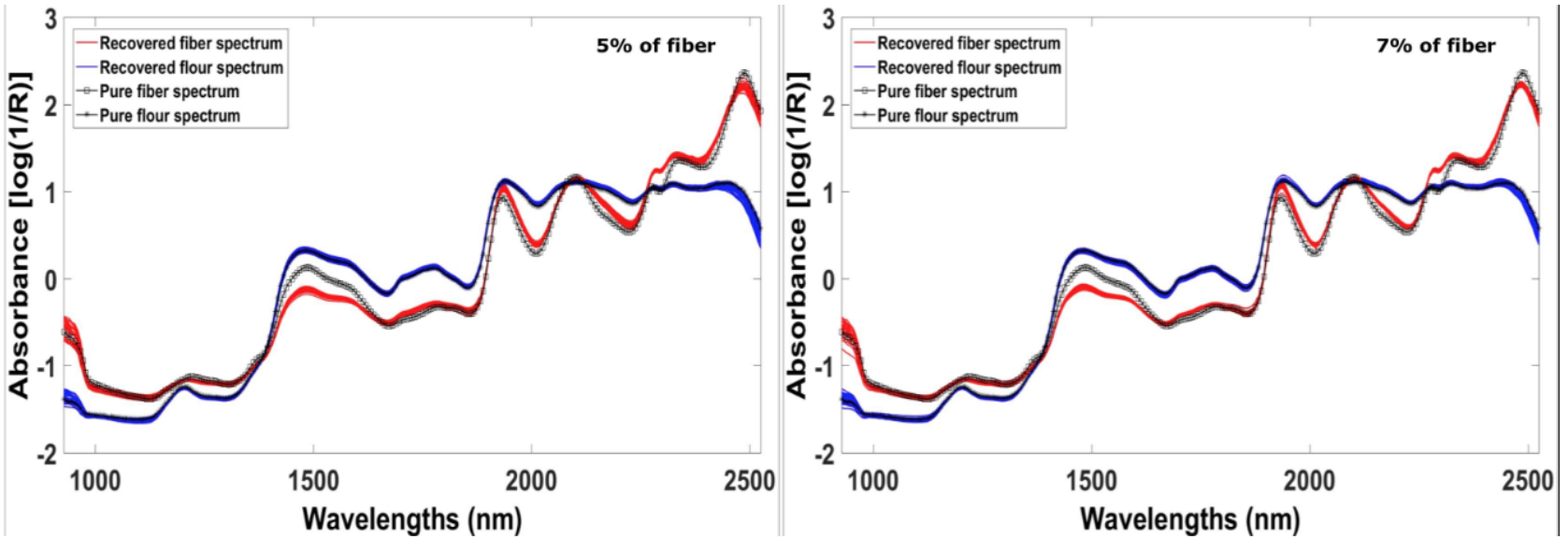
Applications of spectral unmixing in chemometrics

Fiber distribution in pasta (Badaró et al., 2021): NIR-HSI



Concentration map

Pure and recovered spectra from fiber and flour



Spectral unmixing hands-on tutorial in Python

- Developed software program for spectral unmixing based on pyMCR, an open-source Python library

Open-source software library

Setup and installations

python >= 3.4 up to 3.9

pip install pyMCR

Or pip3 install pyMCR

Install additional libraries (Dependencies)

pip install numpy

pip install scipy

pip install matplotlib

Verify Installation

```
import pymcr

print(pymcr.__version__)
```

Spectral unmixing hands-on tutorial in Python

Simulated HSI dataset

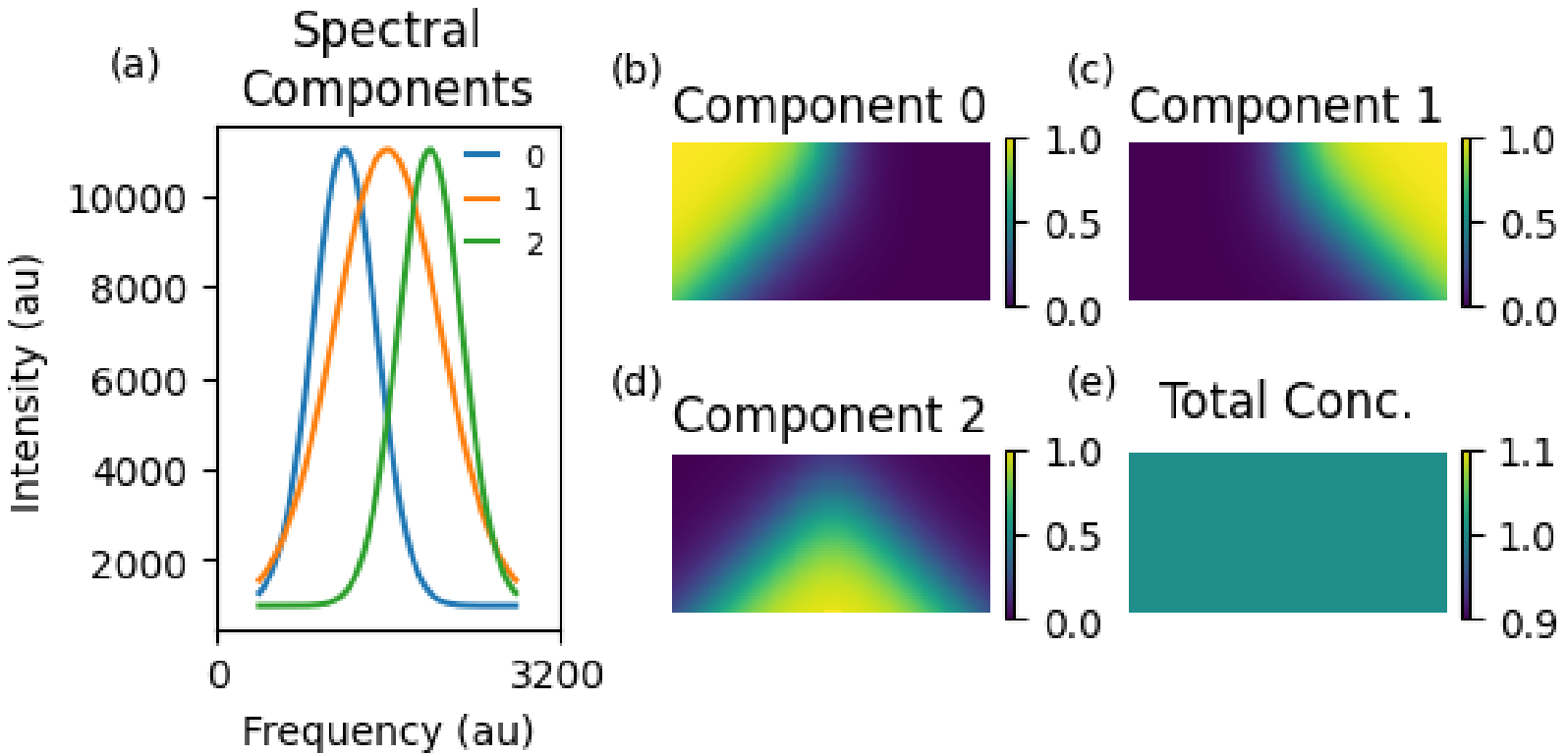
- Spectra with 3 unique components and concentration maps simulated using Gaussian waveform and functions.
- Spectra and concentration maps are combined to form a hyperspectral image (HSI)
- HSI image unfolded into a data matrix, D
- Spectral unmixing performed with MCR-ALS and MCR-NNLS

$$f(x) = a\, e^{\left(-\frac{(x-b)^2}{2c^2}\right)}$$

a : amplitude (height of the peak)

b : center position of the peak

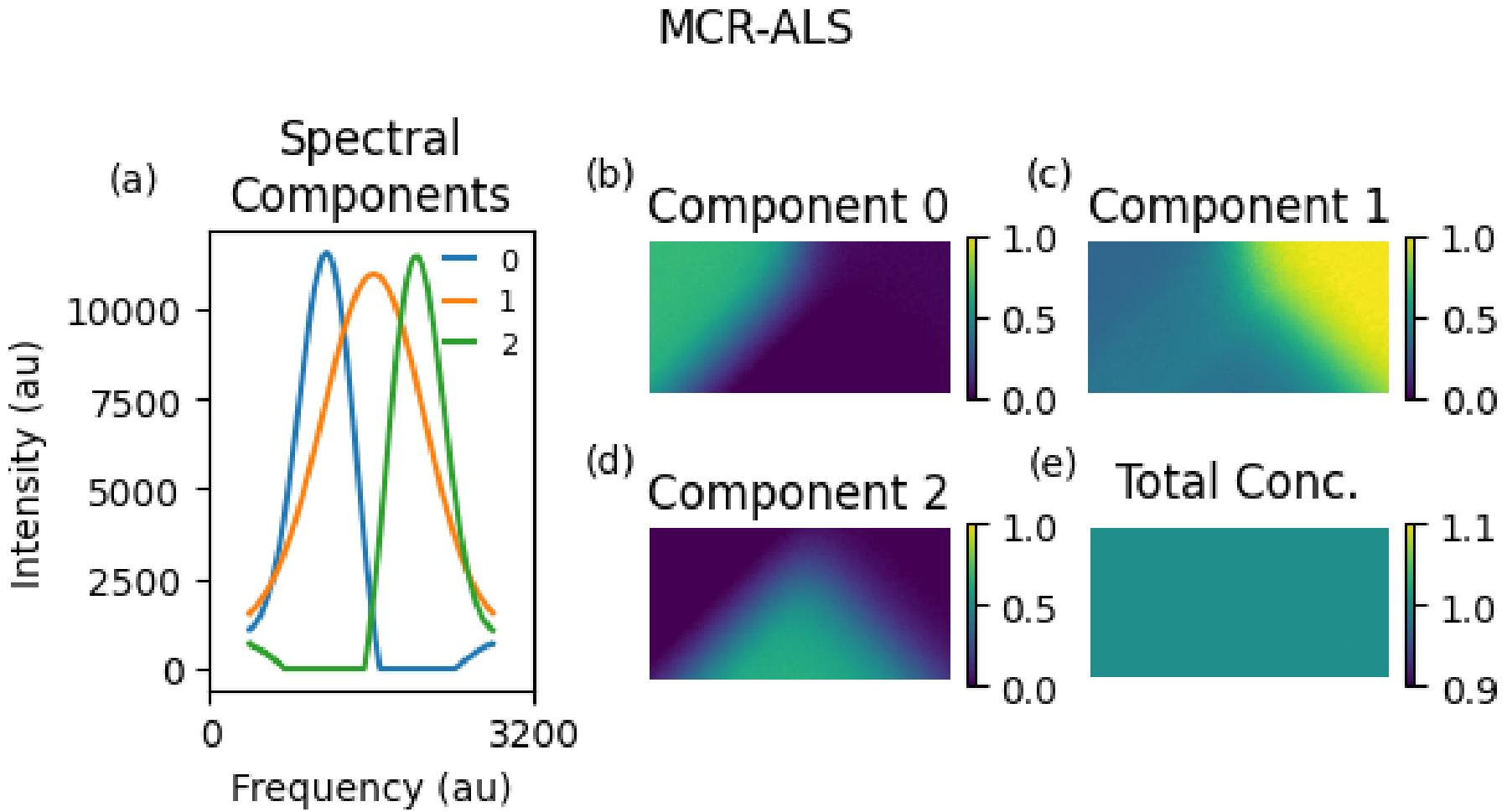
c : standard deviation



Simulated dataset with 3 unique components with their corresponding concentration maps

Spectral unmixing hands-on tutorial in Python

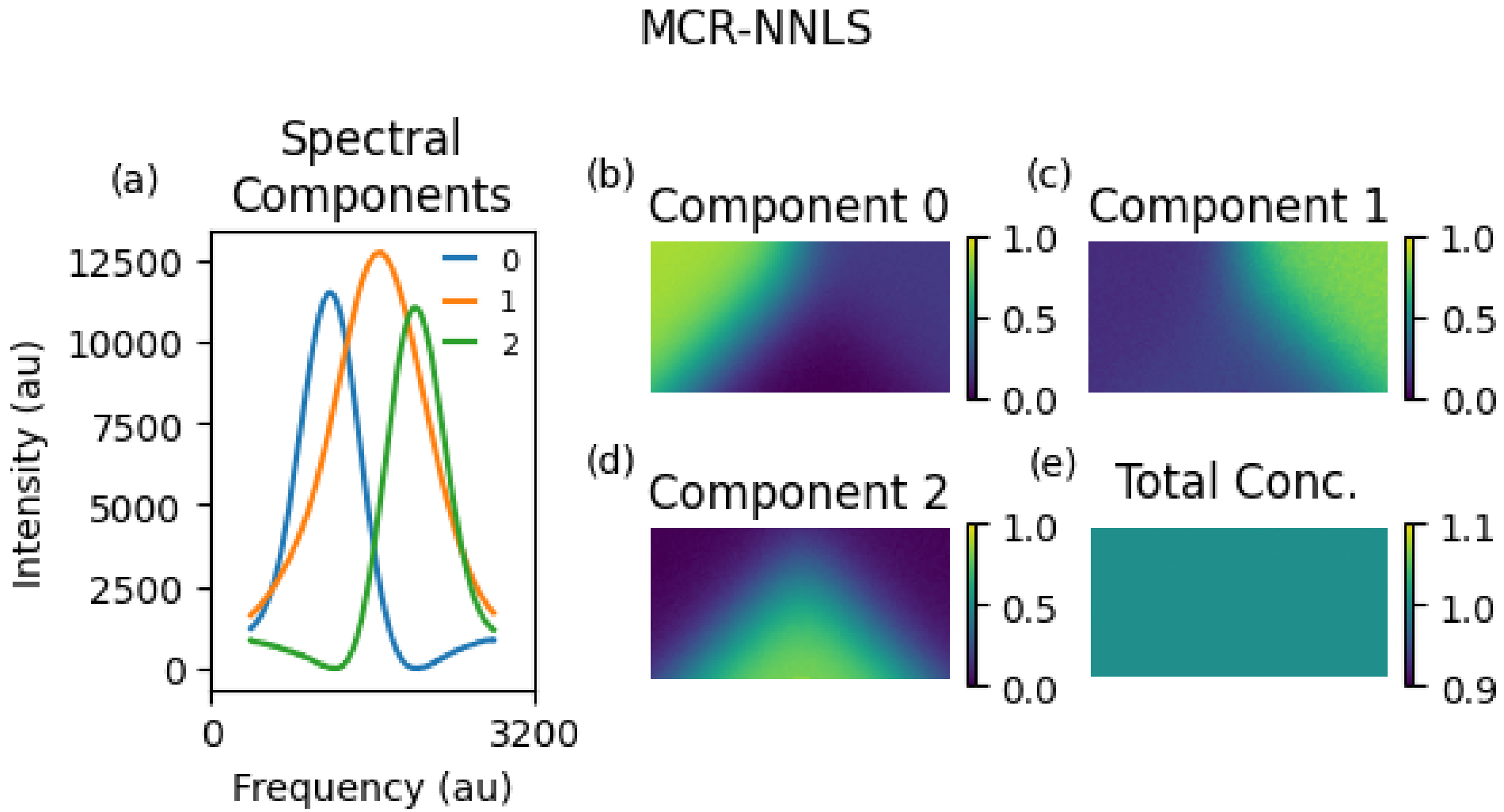
MCR-ALS results



Retrieved results using spectral unmixing with MCR-ALS

Spectral unmixing hands-on tutorial in Python

MCR-NNLS results

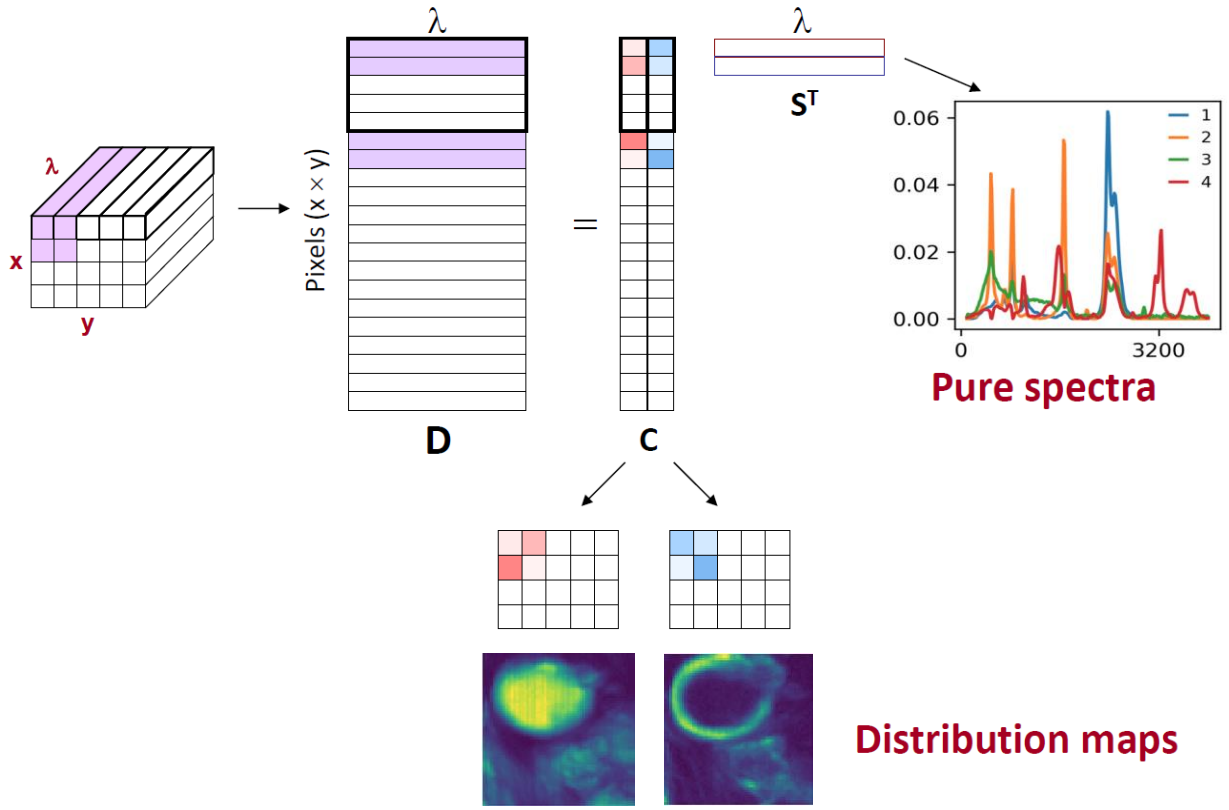


Retrieved results using spectral unmixing with MCR-NNLS

Spectral unmixing hands-on tutorial in Python

Spectral unmixing of an oil emulsion

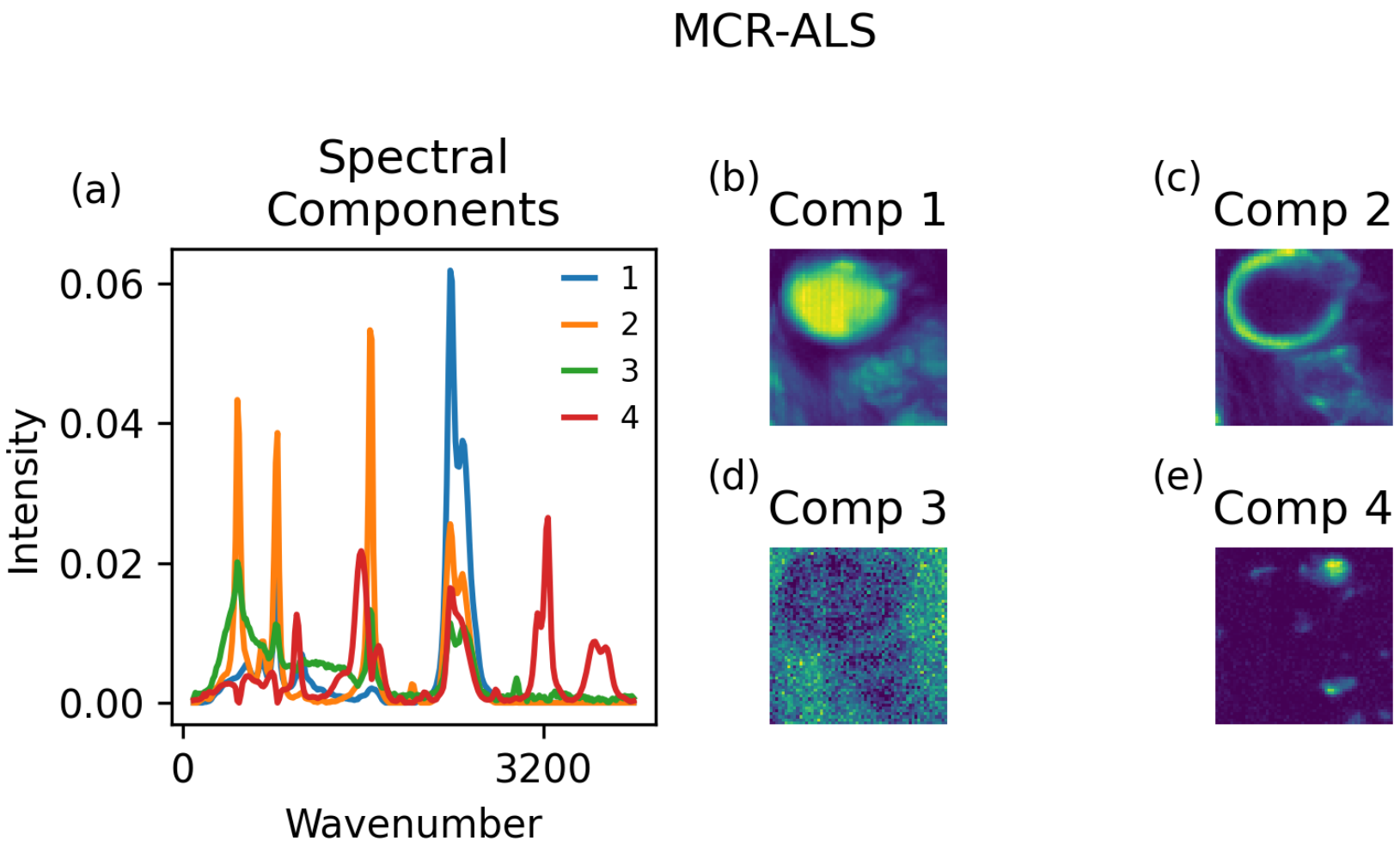
- Downloaded hyperspectral Raman image of an oil emulsion, sized (60 x 60 x 253).
- Oil emulsion is a mixture of oil, water, and an emulsifying agent
- Hypercube unfolded into a data matrix, D
- Performed spectral unmixing using a custom-developed software program in Python
- Number of components (4)
- Initial estimates of spectral profile determined using SIMPLISMA, a pure variable selection method.
- Constraints: non-negativity on C and S matrix
- Used 1000 iterations during optimization



Spectral unmixing hands-on tutorial in Python

MCR-ALS results

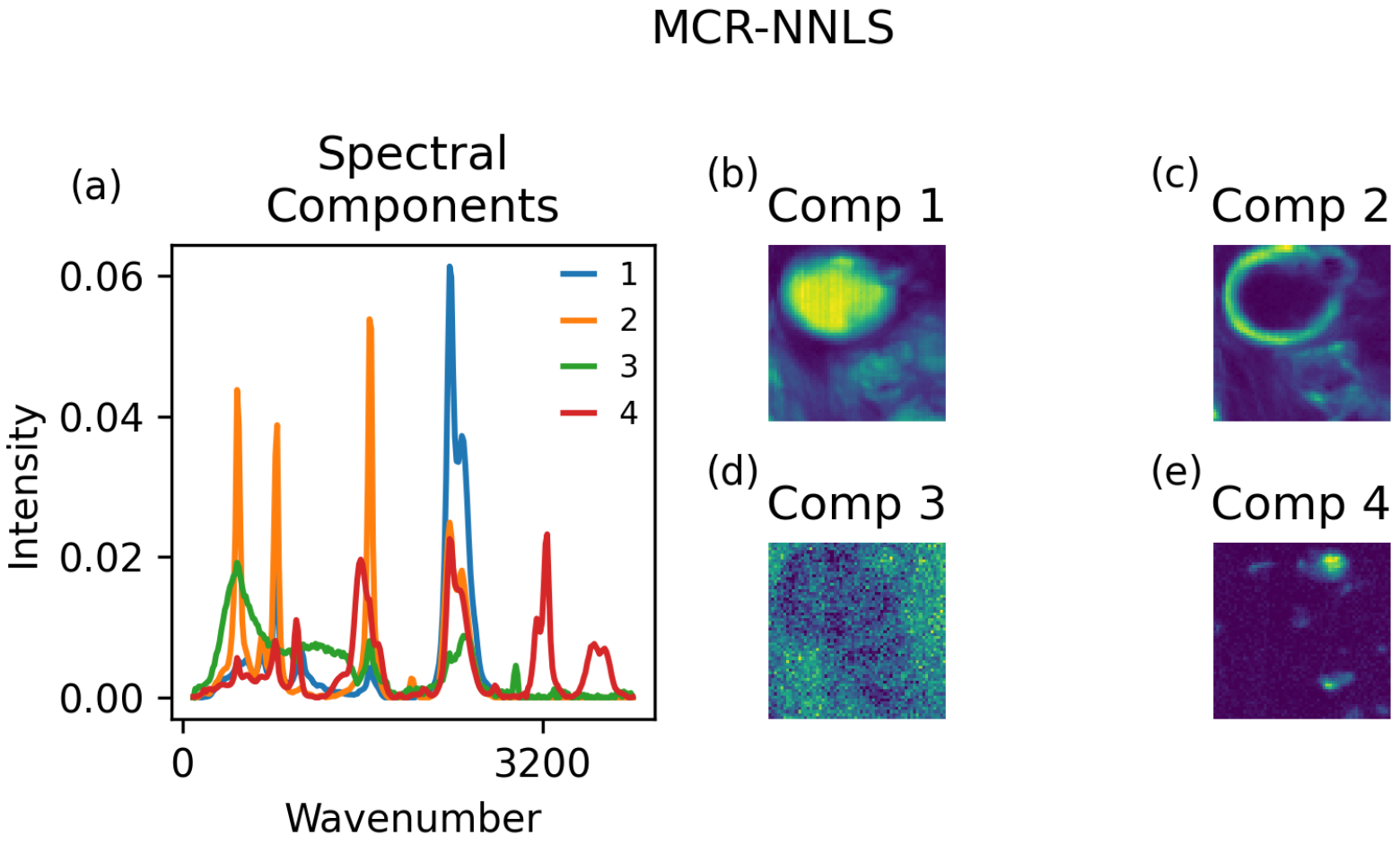
- Pure spectra and distribution maps reflect the surface pattern of an emulsion
- An oily phase represented by two components (a big drop and the interface around), the outer aqueous phase and a small patch, which is an emulsion additive.



Retrieved results using spectral unmixing with MCR-ALS

Spectral unmixing hands-on tutorial in Python

MCR-NNLS results



Retrieved results using spectral unmixing with MCR-NNLS

Conclusions

- Spectral unmixing is effective in chemometrics for mixing problems
- Hands-on practice for spectral unmixing in python was explored
- Research opportunities using spectral unmixing in chemometrics still available
- Flexibility during implementation of multivariate curve resolution alternating regression (MCR-AR) such as various regressors, and constraints can be applied during the optimization process
- With this introduction and hands-on workshop, this approach can be adapted to participants' research
- Looking forward to various collaborations on spectral unmixing

Thank You

References

- Badaró, A. T., Amigo, J. M., Blasco, J., Aleixos, N., Ferreira, A. R., Clerici, M. T. P. S., & Barbin, D. F. (2021). Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fiber distribution in enriched pasta. *Food Chemistry*, *343*, 128517. https://doi.org/https://doi.org/10.1016/j.foodchem.2020.128517
- De Juan, A., Jaumot, J., & Tauler, R. (2014). Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Analytical Methods*, *6*(14), 4964–4976. https://doi.org/10.1039/C4AY00571F
- Olmos, V., Benítez, L., Marro, M., Loza-Alvarez, P., Piña, B., Tauler, R., & de Juan, A. (2017). Relevant aspects of unmixing/resolution analysis for the interpretation of biological vibrational hyperspectral images. *TrAC Trends in Analytical Chemistry*, *94*, 130–140. https://doi.org/https://doi.org/10.1016/j.trac.2017.07.004

References

- Guo, X., Ahlawat, Y. K., Liu, T., & Zare, A. (2022). Evaluation of postharvest senescence of Broccoli via hyperspectral imaging. Plant Phenomics. https://doi.org/10.34133/2022/9761095
- Forchetti, D. A. P., & Poppi, R. J. (2017). Use of NIR hyperspectral imaging and multivariate curve resolution (MCR) for detection and quantification of adulterants in milk powder. LWT - Food Science and Technology, 76, 337–343. https://doi.org/https://doi.org/10.1016/j.lwt.2016.06.046