RUTGERS
UNIVERSITY | NEW BRUNSWICK

MATLAB Based Applications As Accessible and Interactive Educational Modules to Advance Spectroscopic Understanding

Jakub Konkol, Justin Marlowe, and George Tsilomelekis

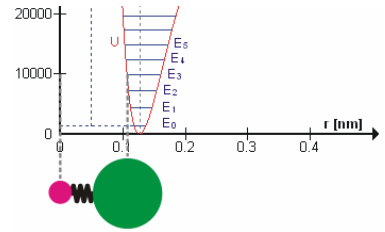
Chemical and Biochemical Engineering, Rutgers, The State University of New Jersey

2021 AIChE ANNUAL MEETING

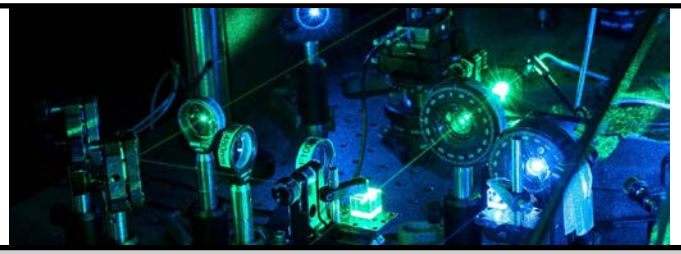
November 9, 2021

Spectroscopic Pedagogy

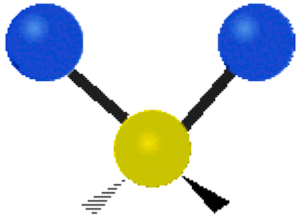
Theory



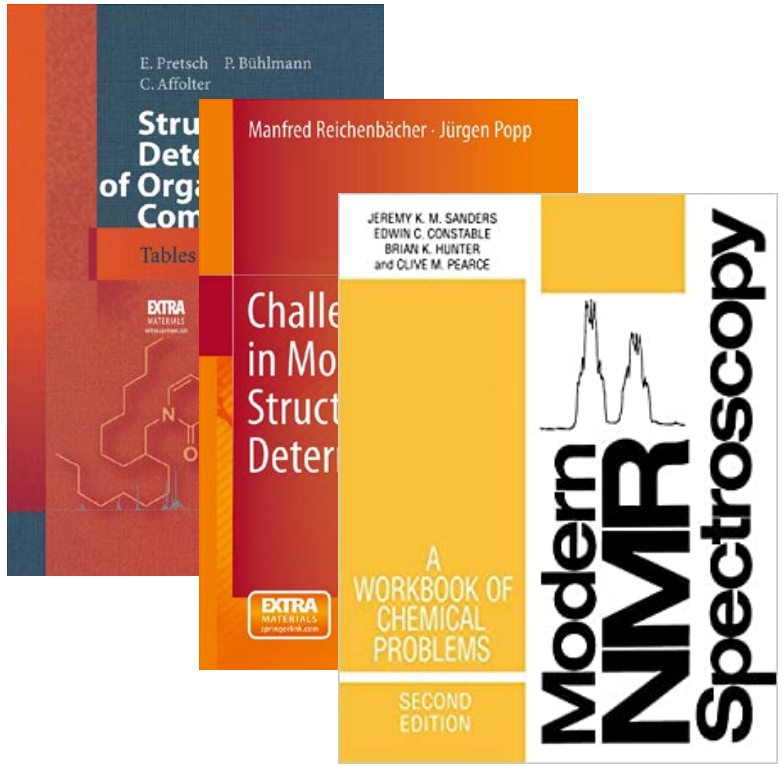
Practice



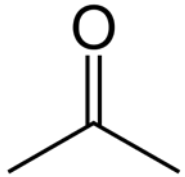
Interpretation

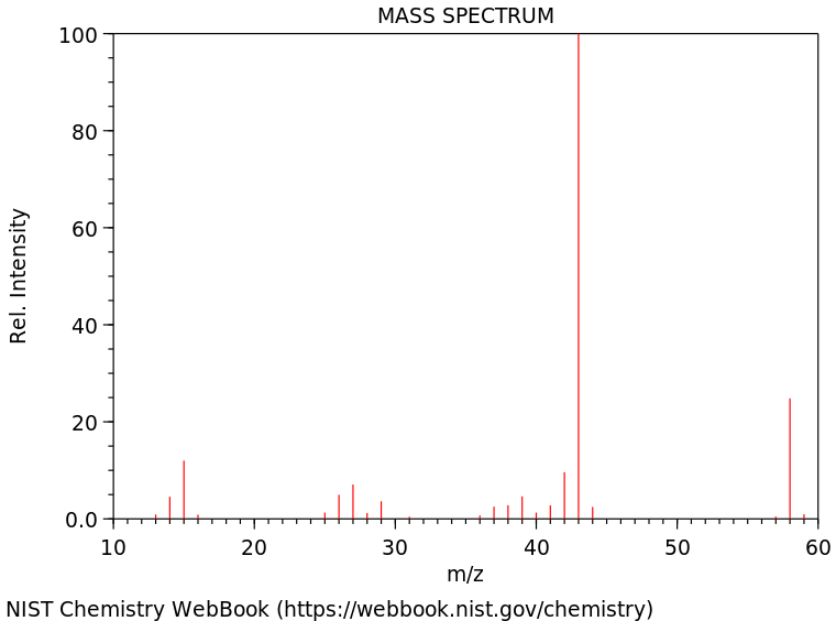


Theory and Interpretation

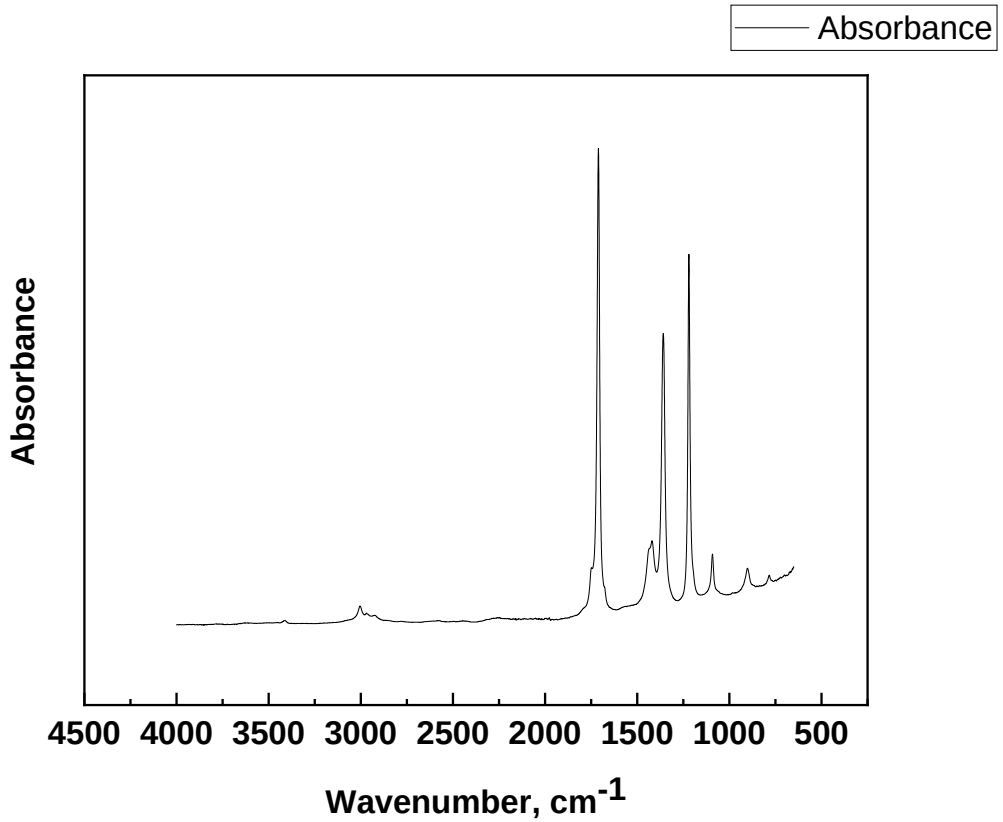


Theory and Interpretation





NIST Chemistry WebBook (https://webbook.nist.gov/chemistry)

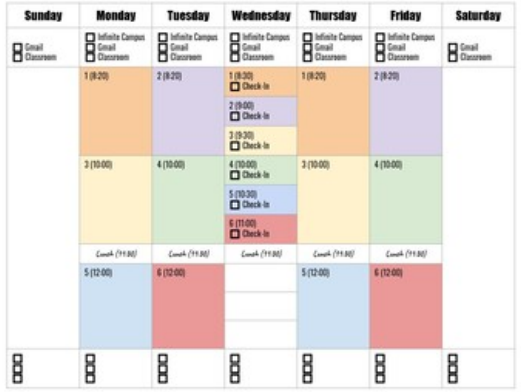


Teaching Spectroscopic Practice

Labs are challenging to run

Expensive

Time consuming



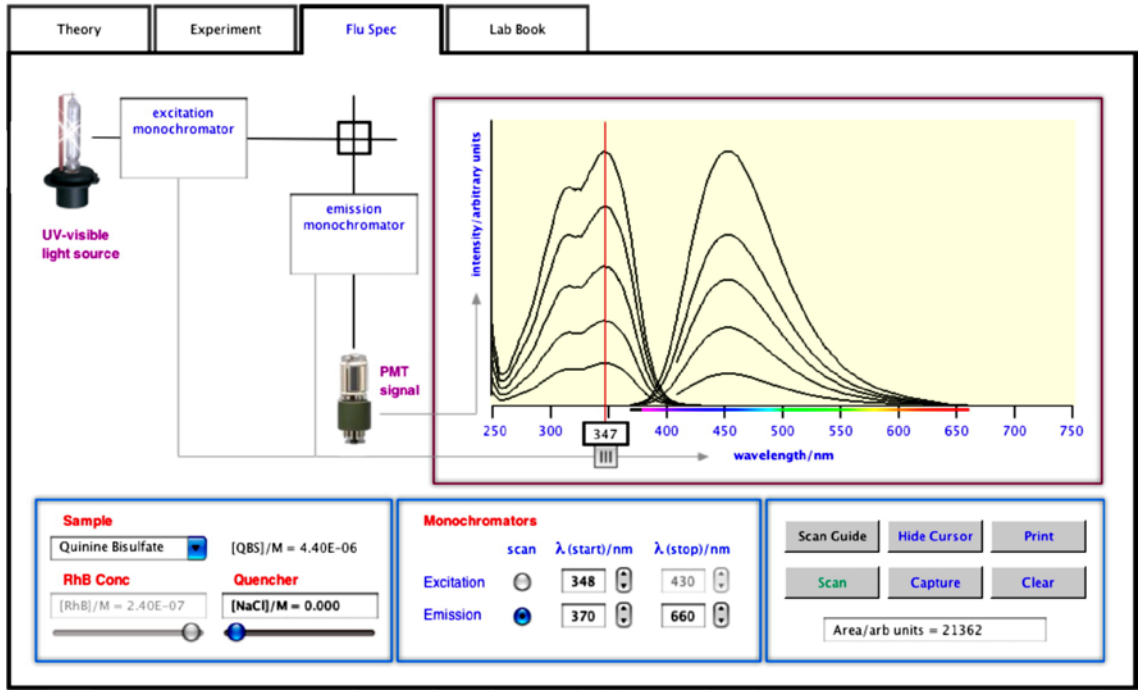
Training

COVID-19 has changed the way we teach

Kobayashi, R., et. al, "Virtual Computational Chemistry Teaching Laboratories—Hands-On at a Distance," **98** (10), pp. 3163–3171 (2021).; **Villanueva, M. E., et al**, "Teaching Instrumental Analytical Chemistry during COVID-19 Times in a Developing Country: Asynchronous versus Synchronous Communication," **97** (9), pp. 2719–2722 (2020).

Computer-based spectroscopy practice

FluSpec

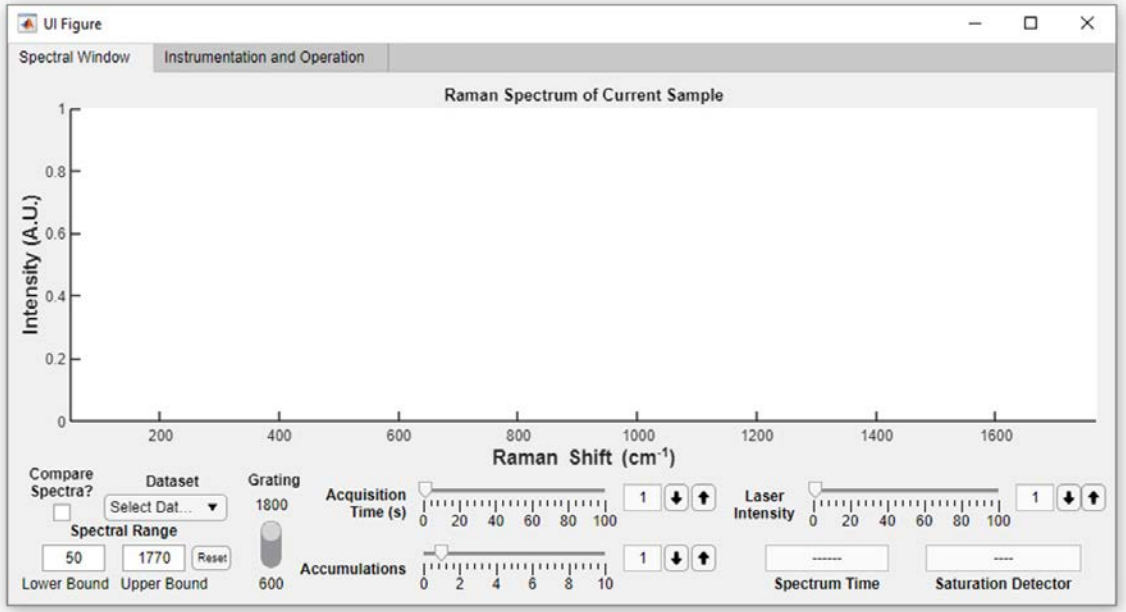


Online UV-Vis experiment

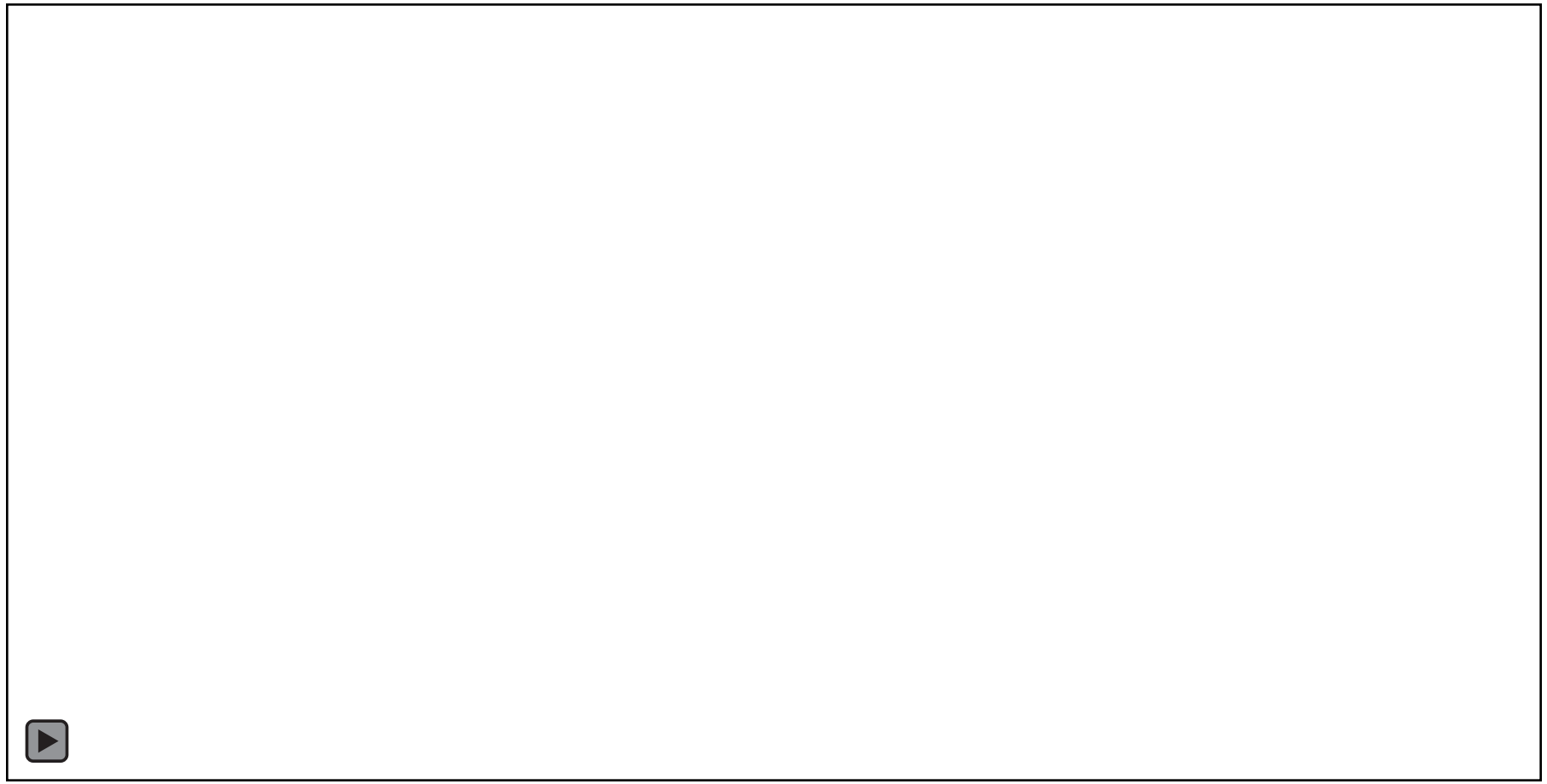
[1] **Bigger, S. W., A. S. Bigger, and K. P. Ghiggino**, "FluSpec: A Simulated Experiment in Fluorescence Spectroscopy," *Journal of Chemical Education*, ***91*** *(7), pp. 1081–1083 (2014).*
[2] https://web.mst.edu/~gbert/Color_Lg/color.html?455

Raman Experiment Simulator

Developed by Justin Marlowe
Developed in MATLAB App Designer
Available at gtsilomelekis.com



Key features



Lab ideas

Standalone

- Students explore parameters
- Answer questions about important topics

Pre-lab

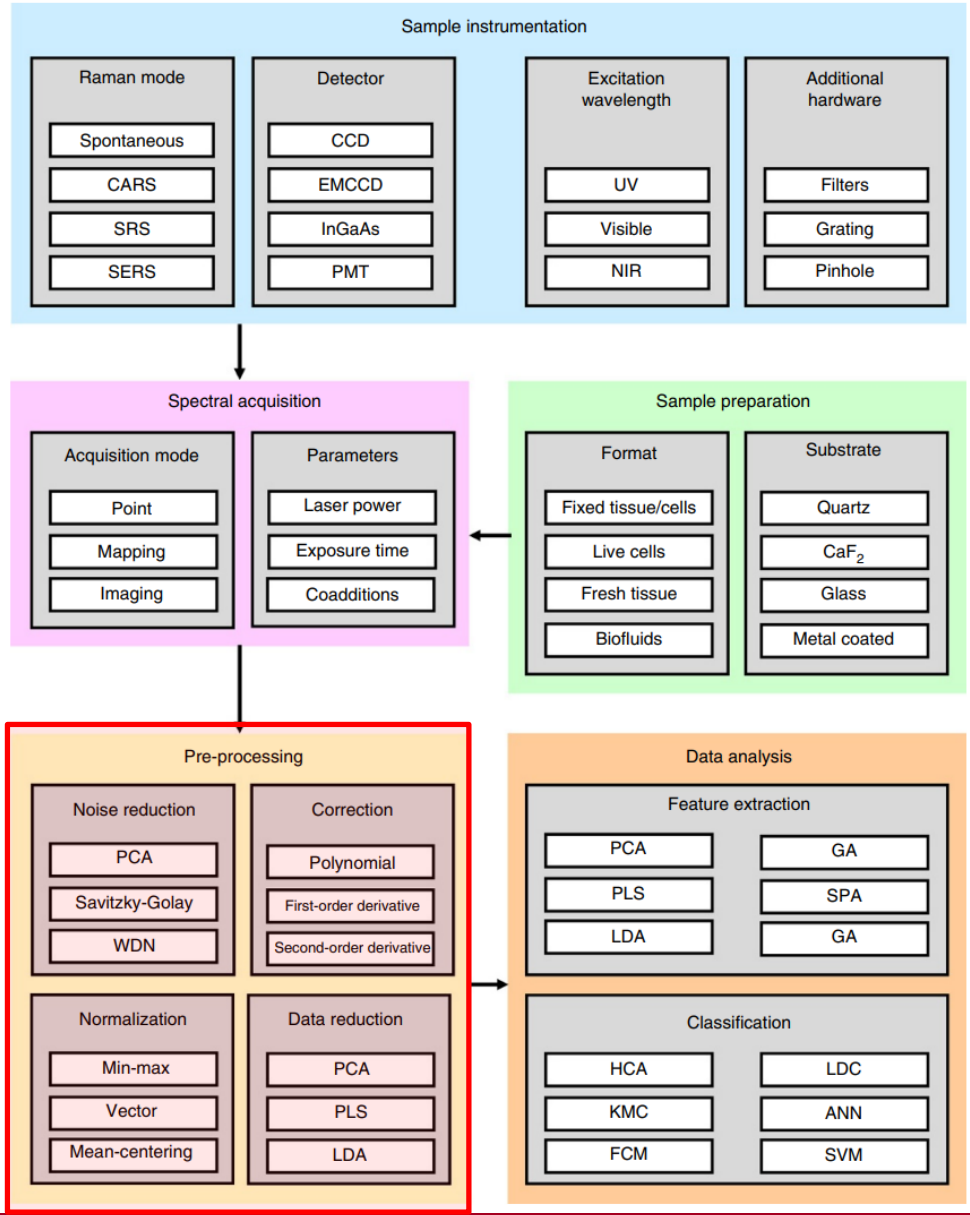
- Students explore parameters
- Decide on best ones that fit criteria
- Collect spectra and reflect

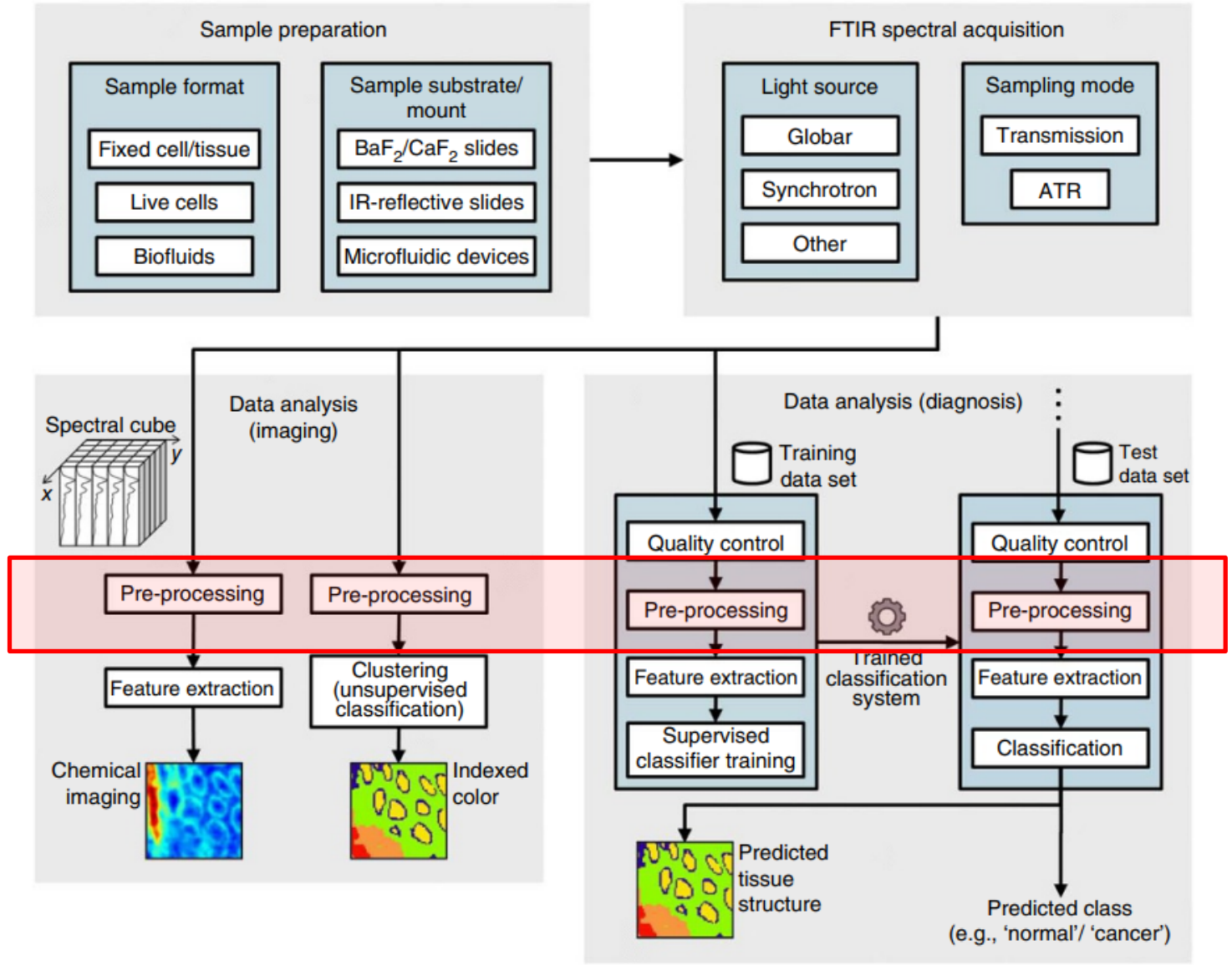
Training

- Familiarize with instrument parameters
- Faster time to meaningful data

We have collected our spectra… Now what?

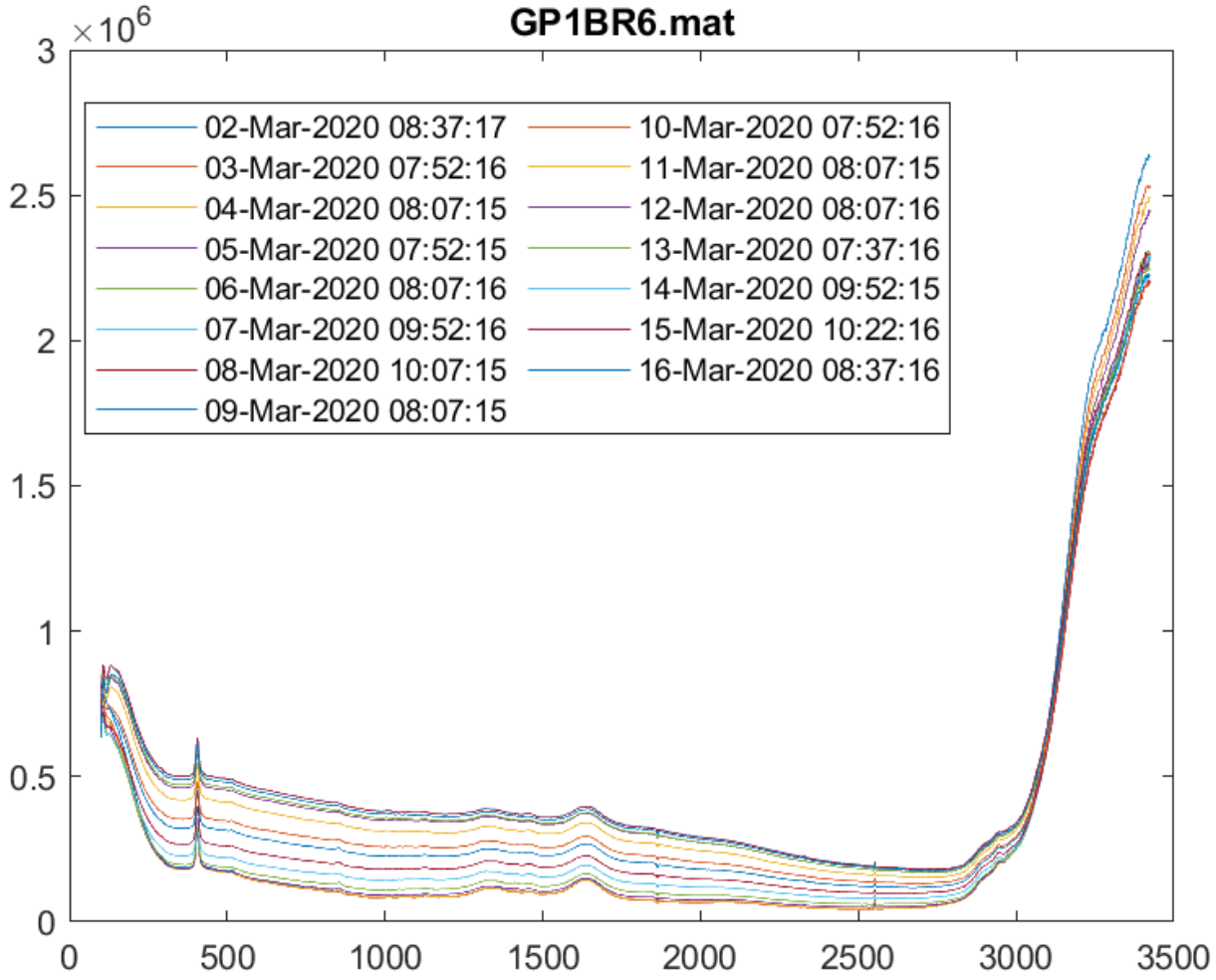
Preprocessing spectra





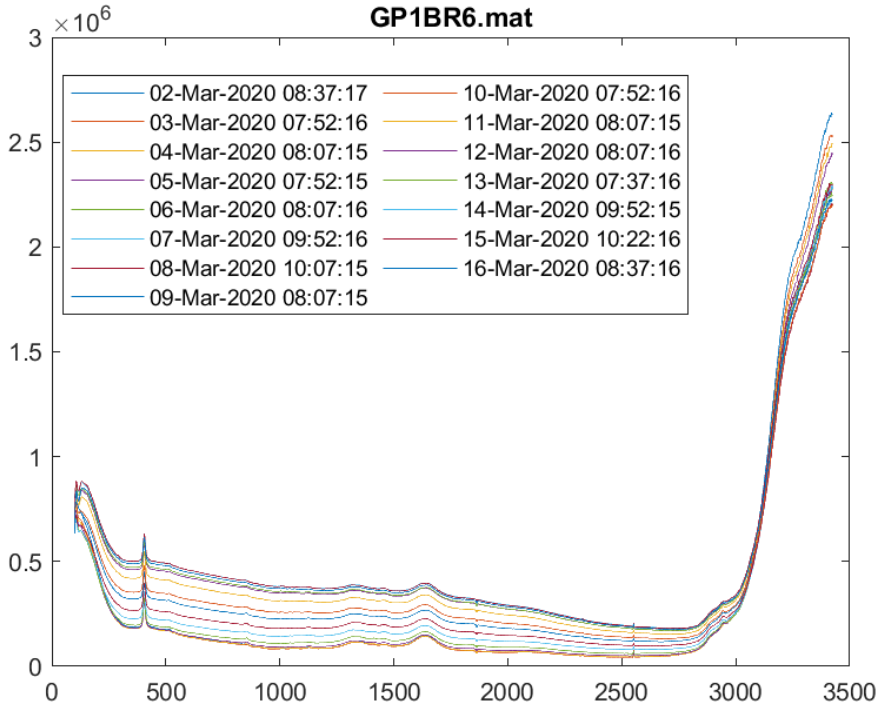
Butler, H. J., et al, "Using Raman spectroscopy to characterize biological materials," *Nature Protocols*, ***11*** *(4), pp. 664–687 (2016).*
Baker, M. J., et al, "Using Fourier transform IR spectroscopy to analyze biological materials," *Nature Protocols*, ***9*** *(8), pp. 1771–1791 (2014).*

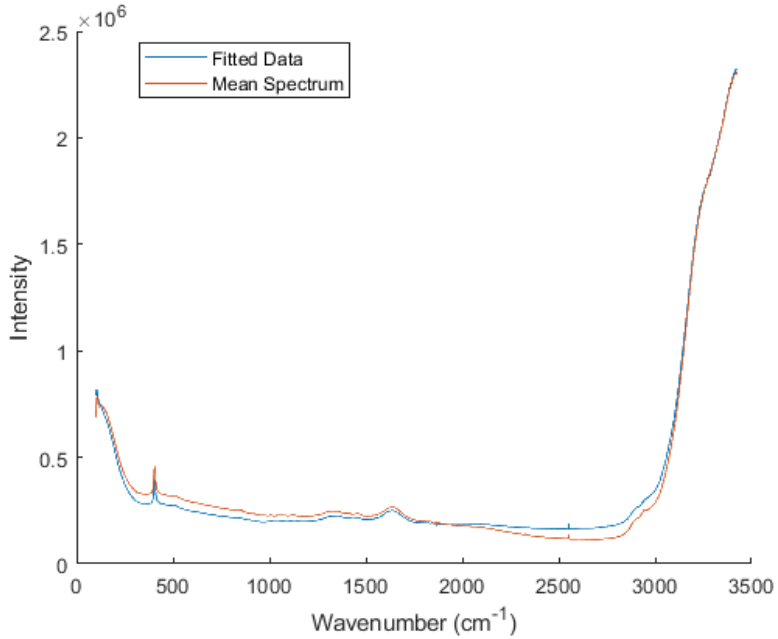
What is preprocessing?



What is preprocessing?

Multiplicative Scattering Correction (MSC)



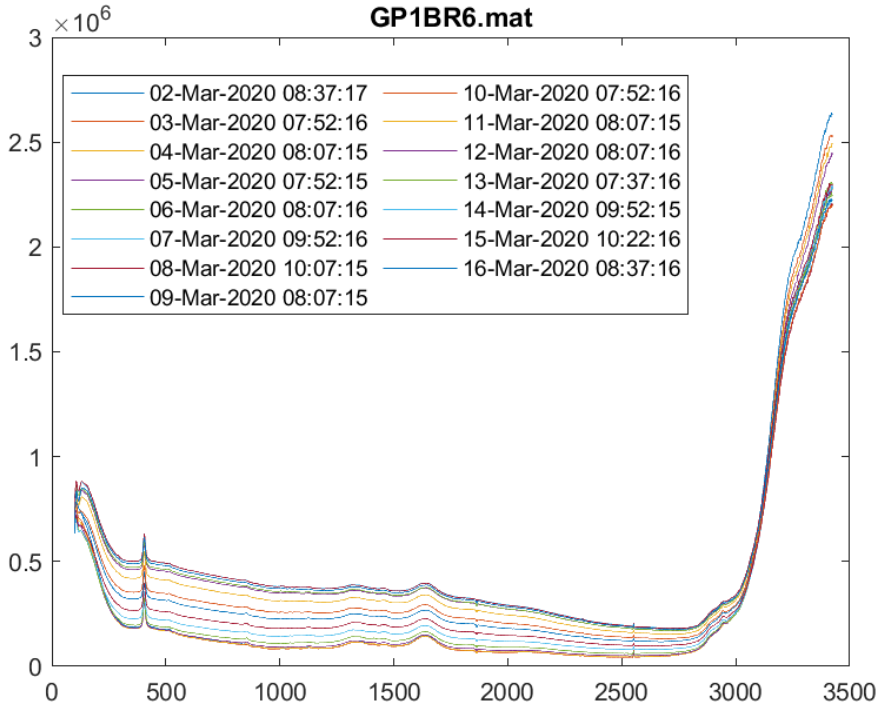


$$\text{Fit } y = a + b \times y_{ideal} + E$$

$$y_{MSC} = \frac{y_{real} - a}{b}$$

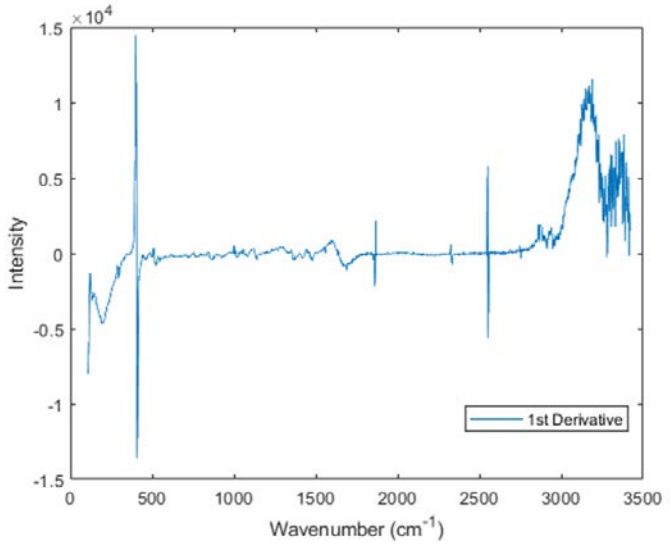
What is preprocessing?

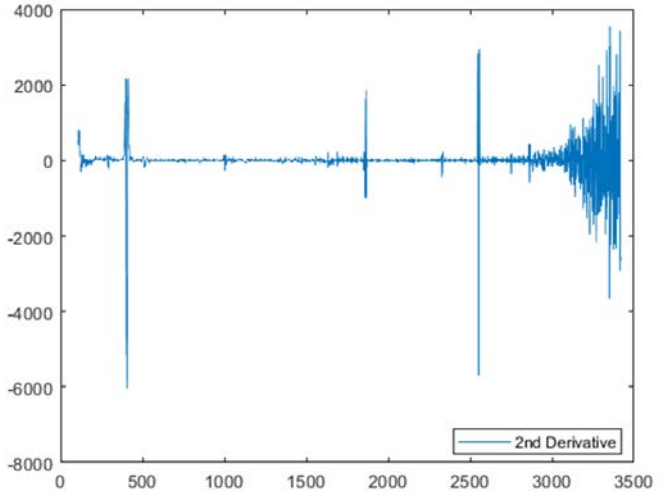
Derivative Spectroscopy



1st derivative

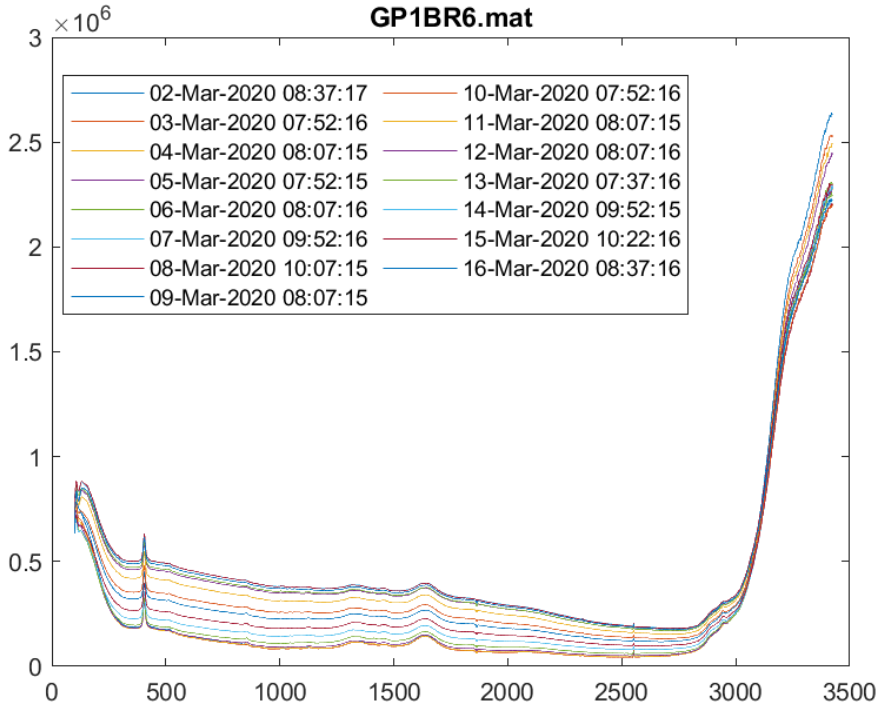
2nd Derivative

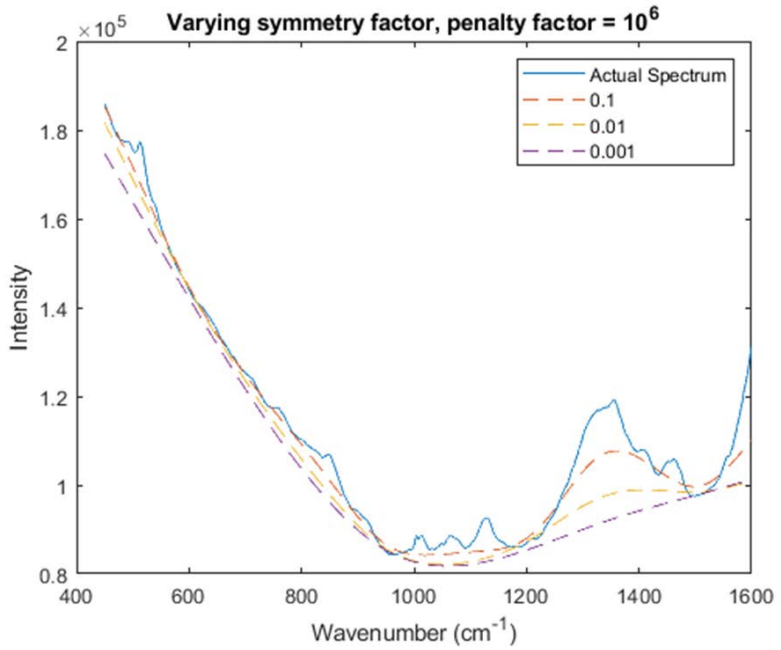




What is preprocessing?

Asymmetric Baseline Fitting





Current preprocessing practice

- Existing preprocessing techniques bordering malpractice[1]
  - Failure to compare preprocessing techniques
  - Inconsistent application
  - Metric of "better"[2]

Chemometrics and Intelligent Laboratory Systems 163 (2017) 64–75

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum

Loong Chuen Lee[a,b], Choong-Yeun Liong[b,*], Abdul Aziz Jemain[b]

[a] Forensic Science Program, FSK, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300 Kuala Lumpur, Malaysia
[b] School of Mathematical Sciences, FST, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ARTICLE INFO

Keywords:
IR spectrum
ATR-FTIR spectroscopy
Chemometrics
Modeling
Data preprocessing

ABSTRACT

ATR-FTIR spectroscopy in the combination with chemometrics has been practiced over the past decades. Works presented in numerous disciplines provide ample empirical evidence in support for the coupling relationship. However, Data Pre-processing (DP) which constitutes the first step in chemometric analysis pipelines, is seldom given reasonable attentions. The aim of this paper is two-fold: (a) to review contemporary DP practice strategy by ATR-FTIR user, and (b) to critically discuss the rationales that could have been nurturing such practices. In the first part, basic concepts of chemometrics and ATR-FTIR spectroscopy are described. Then, the status quo of DP practice strategy is outlined and critically discussed on whether the contemporary practice has been malpractice or best practice. Finally, rationales that could have possibly contributed to some of the malpractices are discussed.

1. Introduction

Over the past two decades, technological knowledges have been evolving so rapidly and contributing to production of High Dimensionality (HD) data in various knowledge disciplines [1–5]. Technological advancements have made collection of analytical data from a tiny sample possible and feasible within such a short period of time [6–9]. Nonetheless, the technological advancements resemble a two-bladed knife, at the same time, such cutting-edge analytical instruments tend to produce data which cannot be readily analyzed and interpreted so to achieve the targeted goal of analysis [10,11]. Data preprocessing (DP) which is also known as data pre-treatment methods are used to remove or reduce unwanted signals from the HD data prior to modeling analysis. As such, DP step is always located right after data collection or acquisition steps in the chemometric pipeline for analytical data. An improper selection of DP methods may negatively affecting the model accuracy and interpretability [12,13]. The vital roles of DP methods have been discussed by numerous sources of books and references that are available in the literature [10–22].

Vibrational spectroscopy instruments including Raman, NIR and MIR spectroscopy, have been coupling with chemometric algorithms in accomplishing different analytical tasks [5,8,9,15–17]. Recently, ATR-FTIR spectroscopy is preferred over transmission FTIR spectroscopy, in diverse field of application [20–31]. The replacement is credited to its non-destructiveness, ease of application and relatively low analysis cost as well as rapid analysis time [6]. Following that, plenty of papers have been published in diverse application fields with the aim to "develop methods to *class or differentiate or identify* a particular samples by using *ATR-FTIR spectra combined with chemometrics*" [23–31]. However, most of these papers has not allocated considerable efforts to systematically select and assess DP methods, prior to modeling. The importance of proper selection of DP methods have been ignored that the user tends to just follows conventional choices of DP methods or shortlisted a few DP methods intuitively. We shall discuss on this matter more in the following section.

To date, a few reports have been reviewed on the application impacts of DP methods in HD data [10,14,15], but only one is devoted to DP evaluation tools [12].To the best of our knowledge, no paper is discussing on the DP practice strategy. On the other hand, most review works or tutorials related to DP methods has always been using NIR data [e.g.14] or Raman [e.g. 22,32] data to demonstrate its practical aspects. It is hardly found any work which is addressing the impacts of

Abbreviations: AS, autoscaling; ATR, attenuated total reflectance; BC, baseline correction; CART, classification and regression tree; DP, data preprocessing; Drv, derivative; FTIR, Fourier

[1] Lee, L. C., C.-Y. Liong, and A. A. Jemain, "A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum," *Chemometrics and Intelligent Laboratory Systems*, ***163***, *pp. 64–75 (2017)*
[2] Engel, J., J. Gerretzen, E. Szymańska, J. J. Jansen, G. Downey, L. Blanchet, and L. M. C. Buydens, "Breaking with trends in pre-processing?," *TrAC Trends in Analytical Chemistry*, ***50***, *pp. 96–106 (2013).*

The pursuit of optimal

“For every digital filter there is a criterion so that this filter is optimal”[1]

“… a robust, quantitative approach to establishing the optimal pre-processing strategy [is] of the utmost importance”[2]

[1] **Bromba, M. U. A., and H. Ziegler**, “Digital smoothing of noisy spectra,” *Analytical Chemistry, **55** (4), pp. 648–653 (1983).*
[2] **Engel, J., J. Gerretzen, E. Szymańska, J. J. Jansen, G. Downey, L. Blanchet, and L. M. C. Buydens**, “Breaking with trends in pre-processing?,” *TrAC Trends in Analytical Chemistry, **50**, pp. 96–106 (2013).*

Porchlight

Interactive preprocessing of data

Supports many filetypes

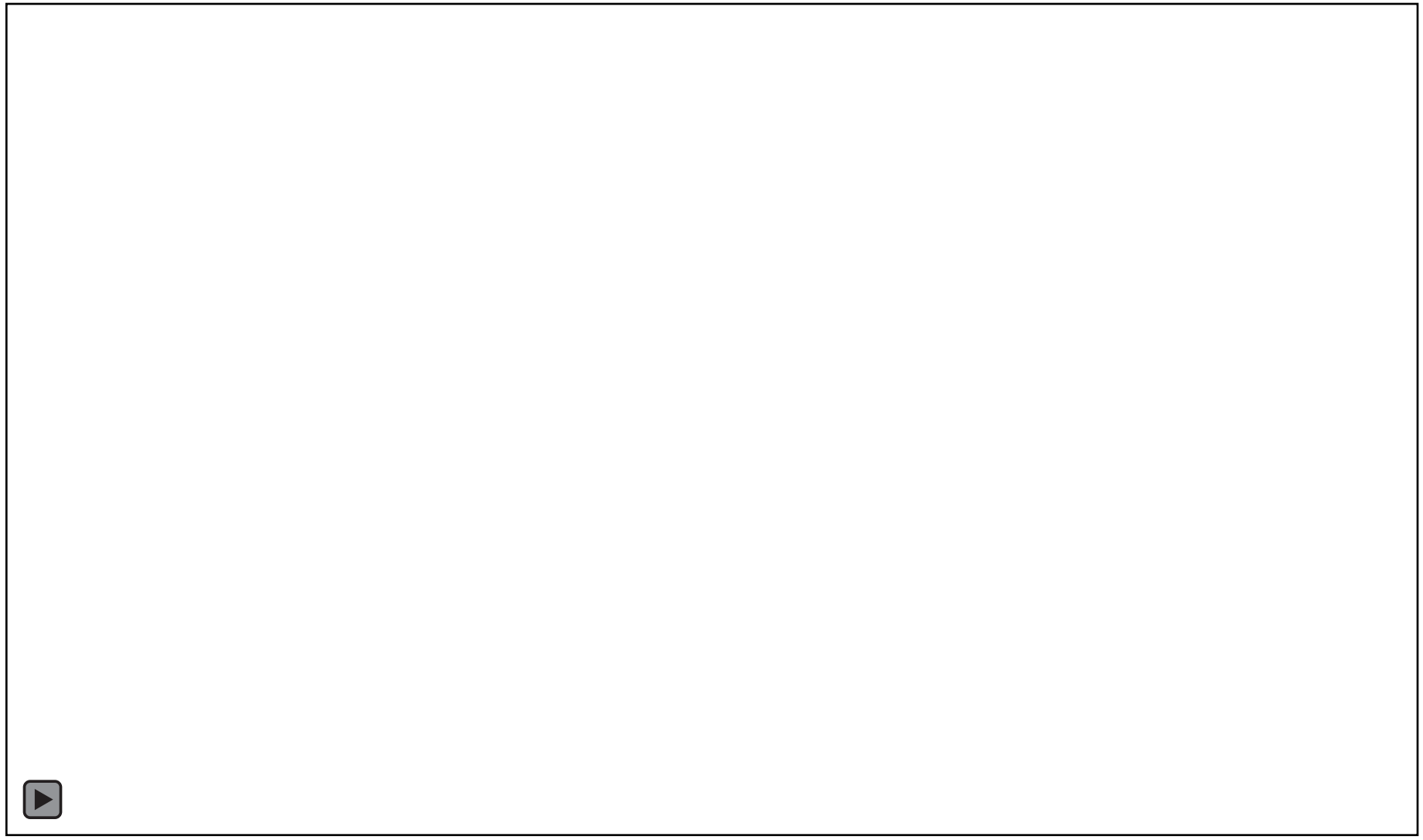
Developed in MATLAB App Designer

Immediately visualize impact

Works with MATLAB 2D-COS program[1]

[1] **Lasch, P., and I. Noda**, "Two-Dimensional Correlation Spectroscopy (2D-COS) for Analysis of Spatially Resolved Vibrational Spectra," *Applied Spectroscopy*, ***73*** *(4), pp. 359–379 (2019).*

Features



Lab ideas

Existing Data — Purely on the computer

- Data with known defects (real or artificial)
- Explore preprocessing effects
- Impact on chemometrics

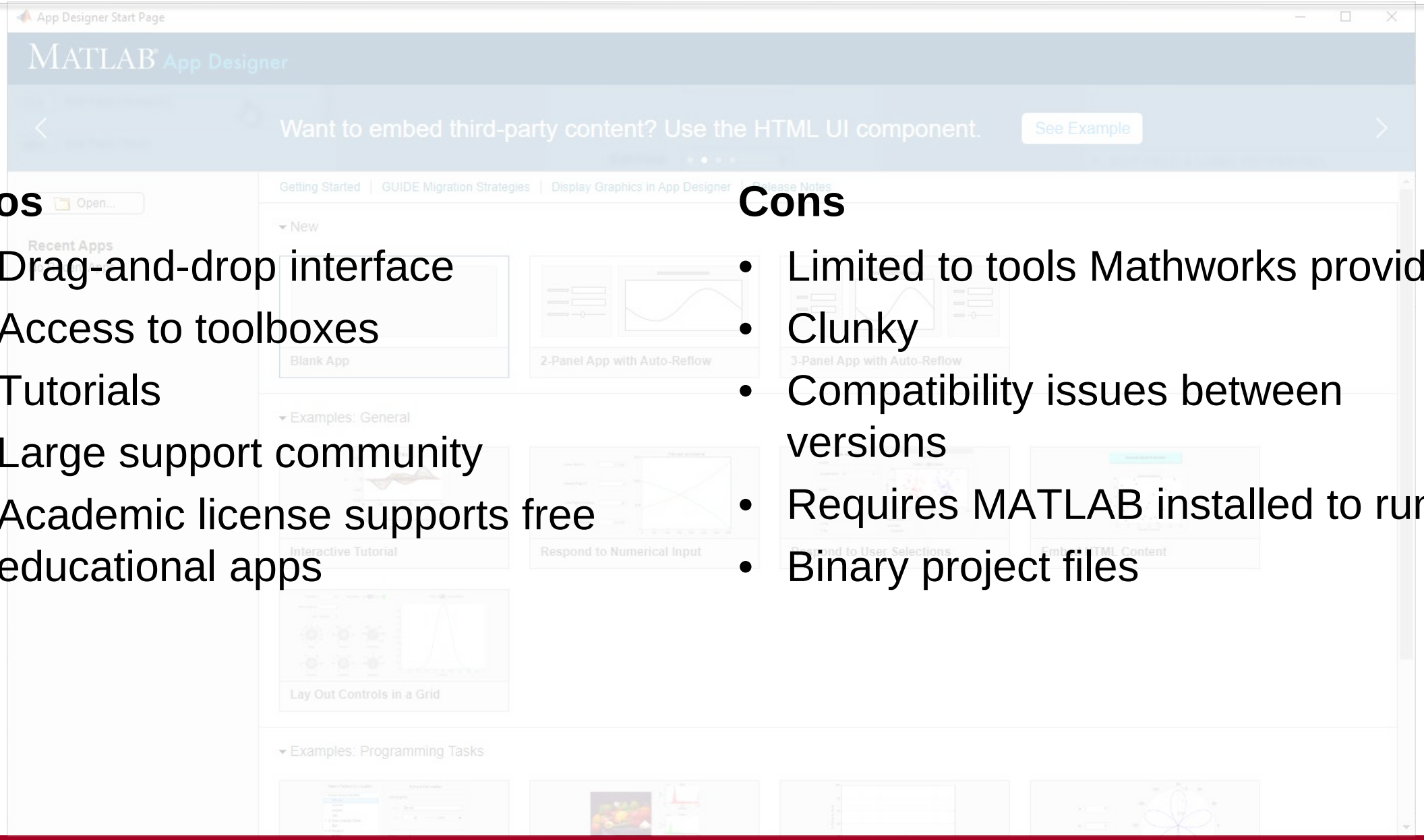
Students' Data — Combined lab practice

- Students collect data of difficult samples
- Identify issues and apply preprocessing
- Apply 2D-COS to explore interactions

Explore Malpractice — Healthy skepticism

- Provide data with understood interpretation
- Task students to tell a different story

Developing GUI applications using MATLAB



Pros

- Drag-and-drop interface
- Access to toolboxes
- Tutorials
- Large support community
- Academic license supports free educational apps

Cons

- Limited to tools Mathworks provides
- Clunky
- Compatibility issues between versions
- Requires MATLAB installed to run
- Binary project files

Take-aways

Jakub.konkol@rutgers.edu
gt241@soe.rutgers.edu

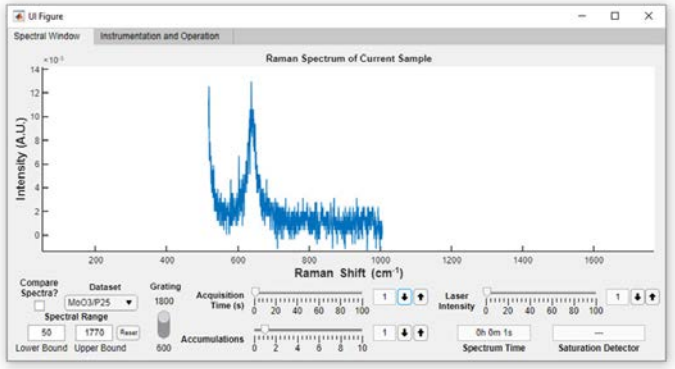
Computer based Spectroscopic Teaching Aids

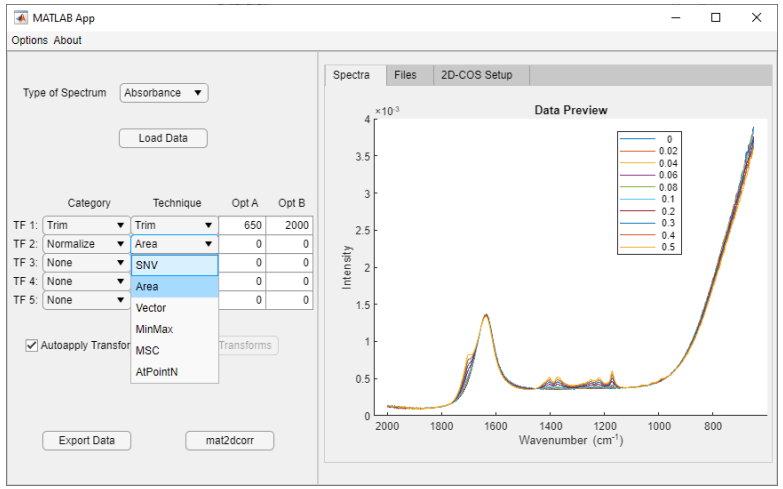
Spectroscopic Practice

Preprocessing

Raman App

Porchlight





Acknowledgements

Rutgers University

Thu Nguyen
Adam Zuber

Award#1751683: "CAREER: Revolutionizing Propylene Production via Oxidative Dehydrogenation and CO_2 Dissociation in Tandem"

RUTGERS
UNIVERSITY | NEW BRUNSWICK

This material is based upon work supported in part by Rutgers, The State University of New Jersey