JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc

Technology Report

Porchlight: An Accessible and Interactive Aid in Preprocessing of Spectral Data

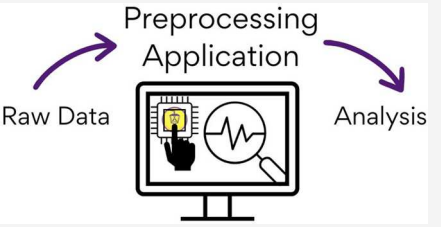
Jakub A. Konkol and George Tsilomelekis*

Cite This: *J. Chem. Educ.* 2023, 100, 1326−1332

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Preprocessing is a critical step in the analysis pipeline of spectroscopic data. However, students are rarely introduced to preprocessing when learning spectral techniques in laboratory courses which in turn may affect and delay their progress in the field. Despite its undoubtable importance, students will be mainly performing spectroscopic analysis in the context of a research project where preprocessing is encountered as part of a routine or "recipe" to follow. In this work, a Python-based application has been developed that allows facile application of common spectral preprocessing techniques with instantaneous results to support student learning. The developed application, i.e. Porchlight, and supplied Jupyter notebooks can substitute costly commercial software and make spectroscopic analysis widely available to students, trainees, and users in general.



KEYWORDS: *Spectroscopy, Computer-Based Learning, Distance Learning/Self-Instruction, Graduate, Chemical Engineering, Laboratory Instruction, Physical Chemistry*

INTRODUCTION

Computer programs are commonplace as an educational aid, especially in chemical engineering education. Within chemical engineering, process simulation software is now indispensable for teaching process design, unit operations, and control.[1−3] Kinetics courses will often have students programming in Python or MATLAB to predict concentration profiles in complex reactor systems. Despite concerns of students treating such applications as black boxes, applications and virtual laboratories have been shown to increase conceptual understanding in students.[4,5] Such incorporation of computer-based and programming lessons is fortunate, as programming is a desirable skill for chemical engineers in industry.[6] However, the benefits are not just limited to chemical engineering. The American Chemistry Society recently published a book on the ways educators are incorporating computer applications and programming to strengthen chemistry education.[7] With these successes in chemical engineering and chemistry, we can expect value in supplementing spectroscopic education with computer applications and tutorials.

Computer-based teaching aids are being developed for spectroscopic education. Some intend to aid understanding the theory behind spectroscopies, such as how proton relaxation leads to spectral peaks in nuclear magnetic resonance (NMR) spectroscopy, or how fundamental molecular vibrations lead to peaks in infrared (IR) spectroscopy.[8,9] Another category of applications intends to teach the practice of collecting spectra, with applications for NMR, IR, UV−vis, fluorescence, and most recently Raman spectroscopies.[10−13] These programs allow students to consider real world ramifications of instrumental parameters in data collection, or at least practice the concepts of data collection. Others have focused on the computer interpretation of spectra, where they developed computer applications or programming lessons for chemometrics.[14−17] These many examples provide evidence of the community's interest in theory and practice of spectroscopy; however the aspect of preprocessing raw data before any spectroscopic analysis is still neglected, despite its ubiquity.

Preprocessing spectra has been described by Rogers et al. as mathematical techniques to reduce noise and systematic variations and to enhance spectral features.[18] One such circumstance would be correcting for scattering differences between catalyst powder samples in Raman spectroscopy. Preprocessing is standard practice in spectroscopy, with protocols for Raman and Fourier transform IR (FTIR) spectroscopy of biological material recommending preprocessing data by default.[19,20] It is known that the preprocessing techniques applied to data affect downstream analysis which in turn potentially affects the interpretation and validation of data.[21] A sizable portion of literature has been dedicated toward consolidating and describing available preprocessing methods.[18,22−26] Others have made attempts to optimize the usage and combination of preprocessing methods.[27−31] However, preprocessing methods are not created equally. A common but susceptible goal is baseline fitting, where Liland

Received: August 18, 2022
Revised: February 3, 2023
Published: February 14, 2023

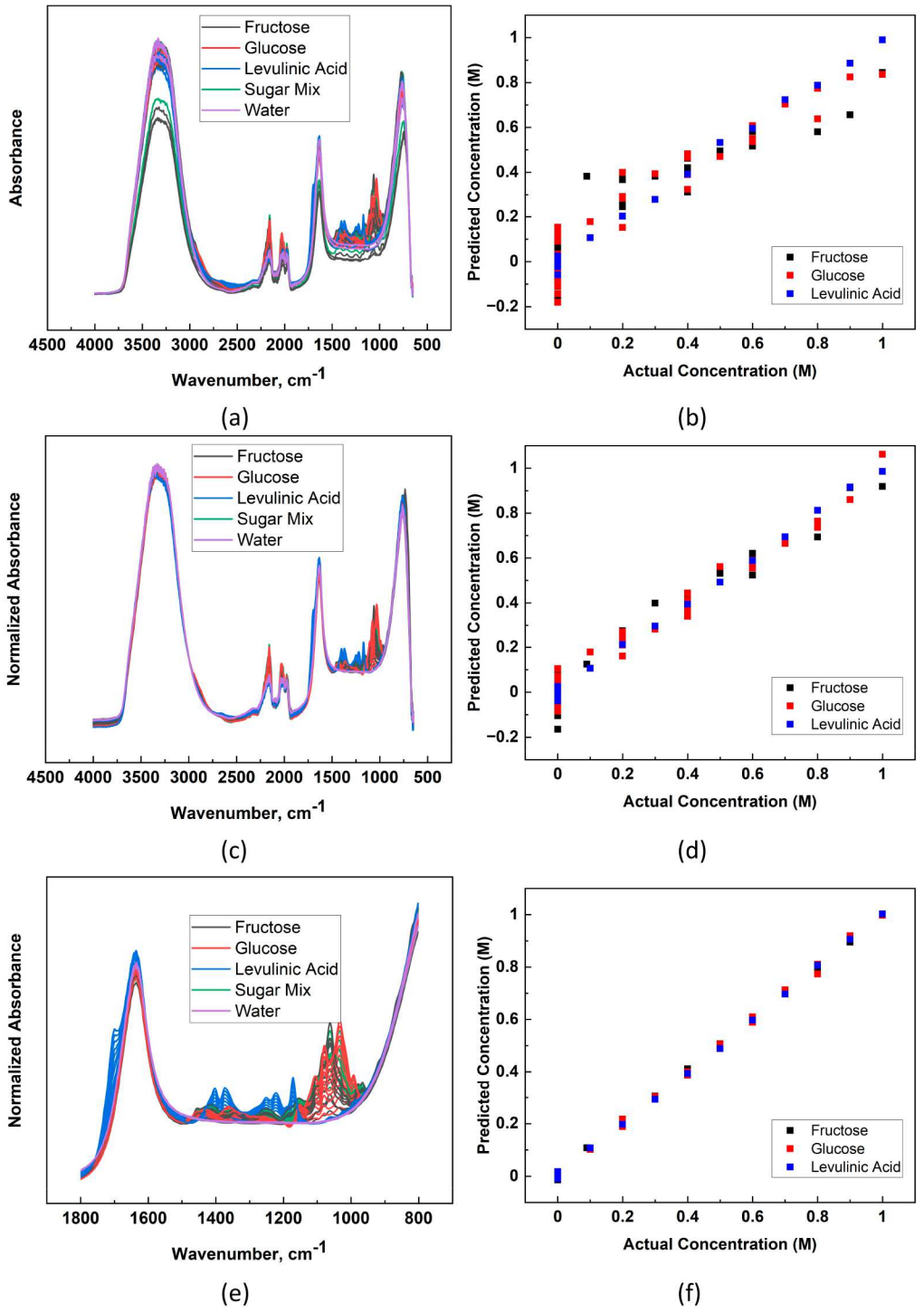


Figure 1. ATR-FTIR spectra of fructose (black), glucose (red), levulinic acid (blue) binary solutions in water, glucose and fructose ternary solutions in water (green), and water (violet) alongside parity plots from a four component PLS model. In (a) are raw spectra, (b) the corresponding concentration predictions, (c) Standard Normal Variate (SNV) processed spectra and (d) corresponding concentration predictions, and (e) SNV then trimmed spectra and (f) corresponding concentration predictions.

et al. found that the techniques which improved Raman calibration models could be deleterious in MALDI-TOF calibration models.[28] Especially in chemometrics applications, a method is generally considered better if the resulting model achieves a better fit but rarely do researchers delve into why it turned out better.[32] In fact, it is seen that researchers may not explore the preprocessing methods available, but instead they tend to assume familiar methods are sufficient.[22] While the number of preprocessing methods continues to grow, the understanding by practitioners does not grow with it. For this reason, upon introduction of spectroscopic techniques, instructors should also introduce the preprocessing methods along with their mathematical formulation and applicability.

https://doi.org/10.1021/acs.jchemed.2c00812
J. Chem. Educ. 2023, 100, 1326−1332

An interactive computer-based module on this direction will be instrumental to support teaching spectral preprocessing.

To the authors' knowledge, there are no standalone programs/modules that specifically focus on preprocessing spectra. Many commercial spectroscopy packages contain some preprocessing capability, such as CytoSpec,[33] Aspen Unscrambler,[34] and InfoMetrix Pirouette.[35] However, there is a licensing expense associated with these programs limiting their use to well-funded research groups and large businesses. Spectragryph[36] is a full spectroscopic suite offering preprocessing, plotting, and analysis of spectra. PyChem[37] is an open-source chemometrics package developed in Python 2 which includes preprocessing, but development of PyChem has ceased. Finally, there is RamanLIGHT, which is domain specific in Raman hyperspectral imaging.[38] With the availability of such programs, one can argue, why not just use one package that does everything? In teaching chemometrics and spectral analysis, a student should learn the technique and the underlying mathematics, not an application. The other extreme is where researchers implement the preprocessing functions themselves in their preferred programming language. While the need for custom scripts that fit one's unique workflow is unquestionable, such an approach risks opening the data to errors as well as slowing down researcher time to investigate and validate each algorithm. Programming has been identified as a valuable skill for chemists and chemical engineers. Therefore, there is value for educational applications to achieve a middle ground with simple to use applications with a programming interface.

In this work, Python is chosen for such a preprocessing program for its open-source nature, for its ability to run on all operating systems, and for providing the flexibility to develop a GUI program and a tool to be called in scripts. Additionally, development focused on using freely available and standard libraries so no functionality is locked behind paywalls.

A SHORT REVIEW OF SPECTRAL PREPROCESSING

A thorough review of preprocessing methods is outside the scope of this report, but some discussion is necessary to show the many applications of preprocessing. Lasch describes preprocessing through eight categories in his review article,[26] but for brevity, this article will repeat five of them: exclusion, normalization, filtering, transformations, and miscellaneous.

- **Exclusion** is the process of removing spectra from the set, such as outlier spectra, or removing unnecessary abscissa variables, such as trimming.
- **Normalization** is the mathematical manipulation of spectra to improve comparison. Despite spectroscopists' best efforts, spectra cannot always be collected under identical conditions. Examples include the following: Temperature and humidity of the room vary throughout the day, catalyst powders can vary in morphology leading to different scatter profiles, or increasing total cell concentration changes solution scatter. Each normalization technique will attempt to overcome specific issues in the data and homogenize a dataset.
- **Filtering** techniques often appear as smoothing of data to decrease noise. Some techniques, such as Raman and near-infrared spectroscopies, can have a low signal-to-noise ratio because of the rarity of those signal events. While the best time to minimize noise is through appropriate instrumental parameters during collection, smoothing can be helpful after data collection to increase the signal-to-noise ratio, although at the cost of deformation of signal peaks.
- **Transformations** are given the distinction where they are grounded in a physical model. Examples include converting between absorbance and transmittance, attenuated total reflection (ATR) correction, and converting reflectance to Kebulka-Munk. Each of these transform the data according to well described systems.
- **Miscellaneous** are the ones that do not quite fit into the other groups but tackle a specific defect in data, such as baseline fitting, cosmic ray correction, etc.

A common goal in preprocessing is to homogenize a dataset in order to compare spectral data under normalized conditions. An example is shown in Figure 1; ATR-FTIR spectra of solutions of fructose, glucose, levulinic acid, or a mixture of sugars in water are presented. The objective of this example is to use these spectra as a calibration dataset in a model that will predict sugar concentration in biomass reaction solutions containing fructose, glucose, and levulinic acid. This goal will be achieved using a partial least-squares (PLS) regression model, a form of multivariate analysis commonly used in spectroscopy and chemometrics. In univariate modeling, one would use individual variables to predict the concentration but that does not work well when the variables are convoluted, such as in the spectra of compounds with overlapping peaks. In multivariate modeling such as PLS, a representation of all variables is found that represents a hidden pattern in the data, a latent variable, which is then used to estimate the *y*-variable. A model is described by the number of latent variables, or components, used for making predictions. For a more thorough introduction of multivariate modeling, one may be interested in starting with principal component analysis[39] or reading more about PLS.[40]

The dataset contains 40 spectra. There are nine spectra each of fructose, glucose, and levulinic acid binary solutions in water ranging from 0.1−1 M, ten spectra of fructose and glucose ternary mixtures where the total sugar content is 1 M, and finally three spectra of water. Figure 1a shows the raw spectra of the solutions, where several defects are present. The first major defect is that three spectra do not match the intensity of the others. The intensity mismatch is most noticeable in the OH region, 3750−3000 cm^{-1}, where two fructose and one sugar mix spectra are lower in intensity than the rest, or in the 1500−1000 cm^{-1} region where those same spectra appear as if they had a lowered baseline. The remainder of the spectra have roughly the same overall intensity profile, especially the other fructose and sugar mix solutions, which is why we were able to identify these three as outliers. There is no anticipated variation in the baseline, so baseline correction is unnecessary. The noise is in the fingerprint region is low so filtering is would do more harm than good. Finally, without prior knowledge of the solution refractive indices, a transformation such as ATR-correction cannot be performed. Figure 1b shows a parity plot for predictions from a four component PLS regression model. A parity plot allows us to quickly see how well the model works. By plotting the predicted value against the known value, a perfect model will result in parity where all points are on the diagonal—the prediction matches the known value. A poor model will deviate from the diagonal. In Figure 1b, the model shows weak predictions for the sugars. This is especially seen around 0, where even in solutions without glucose, the

https://doi.org/10.1021/acs.jchemed.2c00812
J. Chem. Educ. 2023, 100, 1326−1332

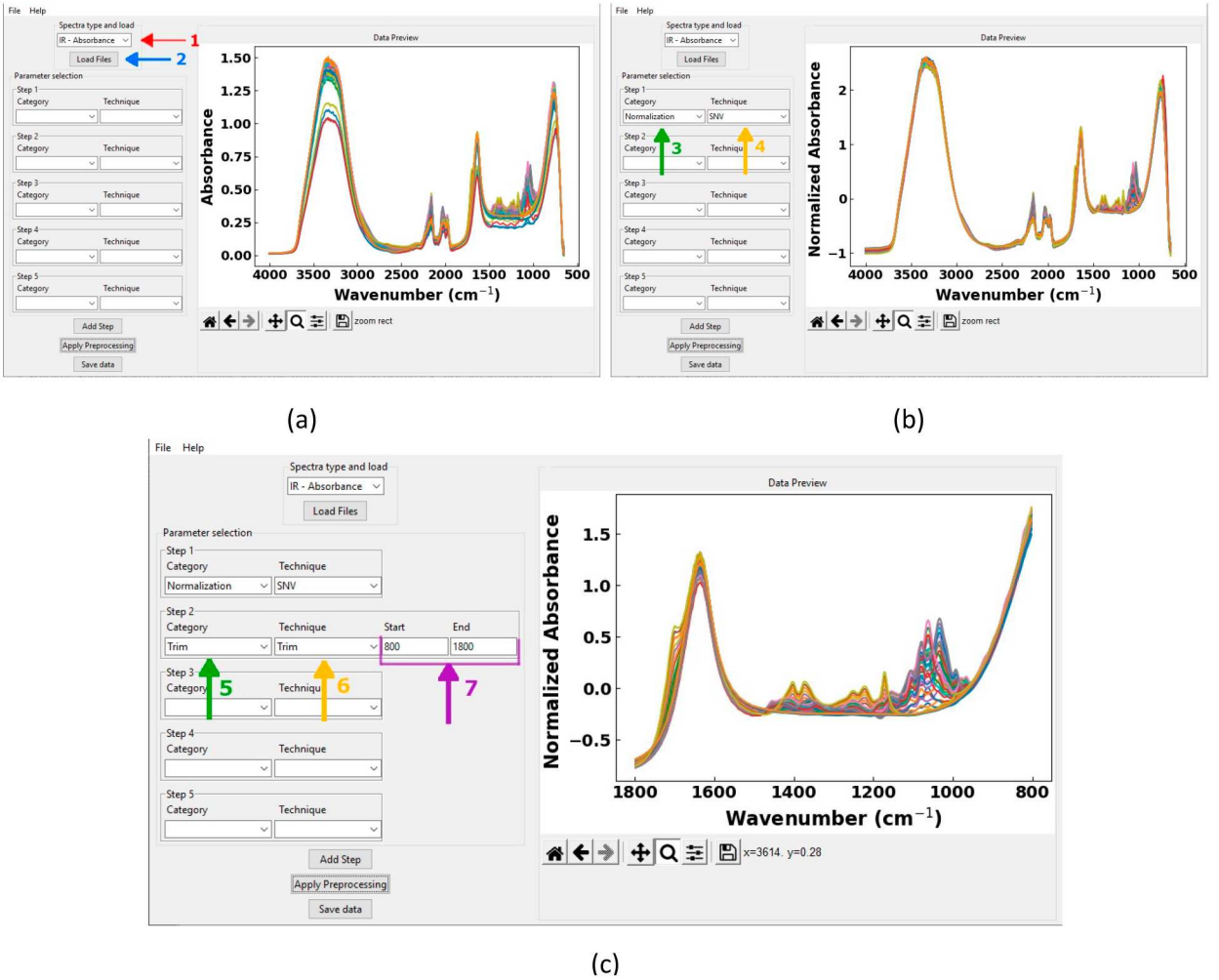


Figure 2. Applying the preprocessing steps on the sugar solutions in Porchlight using the GUI. (a) shows the spectra after selecting the type of spectroscopy (arrow 1, red) and loading the files (arrow 2, blue). (b) shows the selection of "Normalization" in Step 1 Category (arrow 3, green), and then "SNV" in Technique (arrow 4, amber). (c) shows the trimming step cutting down to the range 800−1800 cm^{-1} by selecting the category "Trim" (arrow 5, green), Technique "Trim" (arrow 6, amber), and then placing bounds in the parameter boxes that appeared (arrow 7, violet).

presence of fructose had the model estimate between apparent negative concentrations to 0.2 M of glucose. Levulinic acid is already adequately estimated as it is very different in structure and spectra. In order to contend with the intensity defect, one can normalize the spectra by Standard Normal Variate (SNV). The SNV normalized spectra in Figure 1c shows, at least visually, that the intensity defect has been corrected. Figure 1d shows a model trained on the SNV corrected data is improved, with data points not as spread out.

The next major defect to fix is the diamond ATR region, from 2300−1900 cm^{-1}. This region is a result of diamond being a strong IR absorber in this range, which makes it susceptible to minor changes resulting in high noise. Spurious correlations from this region can degrade predictive models. Additionally, there is not much information to be gained from the OH region either as the absorbance from the water is already so high that the small changes from the addition of sugars or levulinic acid will be minor. We will try to improve the model by focusing only on the fingerprint region, as this is where the most important signal variance is found. Trimming down to the fingerprint region reveals the spectra in Figure 1e. The parity plot in Figure 1f shows we can greatly improve the PLS model's ability to identify the variable correlations and their impact by focusing in on this region. This preprocessing can be performed in the GUI of Porchlight, as shown in Figure 2. Porchlight can also be incorporated in scripts such as the one used for the PLS regression and in Jupyter notebooks. With a guiding hand, future spectroscopists will be better prepared for their work.

■ PYTHON APPLICATION

Python has been among the most popular programming languages in the past ten years,[41] is one of the most commonly used among chemical engineers,[6] and has been used in many classes within chemistry and chemical engineering to teach programming.[42−45] Python is therefore an excellent choice as the base language for our application. Specifically, Porchlight is developed using Python 3.9, NumPy 1.21.5,[46] Pandas 1.4.2,[47] Matplotlib 3.5.1,[48] and Tkinter as standard libraries. The installation files can be found at https://github.com/Jakub-Konkol/Porchlight. The complete guide is also provided in the Supporting Information.

The main application window is split into two parts, as seen in Figure 2. The left panel is the user control side where users can locate files and apply preprocessing methods. The right panel is where the preprocessed data are shown. The use of the

https://doi.org/10.1021/acs.jchemed.2c00812
J. Chem. Educ. 2023, 100, 1326−1332

program begins with selection of the type of spectroscopy. As a teaching aid, we took special care to have the plot labels accurately reflect the contents as well as the impact of preprocessing techniques. Next, the student selects their spectral files. Porchlight currently supports .txt, .csv, Microsoft Excel, and .spc files, with mixed single or multispectral files. If data files do not contain the same amount of data points, Porchlight will do its best to match up the closest ones. Next, the student can select a category of preprocessing techniques. Doing so populates the Technique dropdown with appropriate methods. Upon selecting a technique, Porchlight will dynamically prompt the user for relevant parameters. A label above each entry box describes what the student must provide, but full descriptions of these parameters and some recommendations are available inside the manual. A list of all methods available is presented in Table 1. The

Table 1. Preprocessing Methods Available in Porchlight

Preprocessing Method Type	Method	Source
Trim	Trim	
	Inverse Trim	
Baseline Correction	Polyfit	49
	Asymmetric Least Squares	50
Smoothing	Savitzky-Golay	51
	Moving Window	
Normalize	Standard Normal Variate (SNV)	52
	Area	25
	Vector Normalization	25
	Min-Max	
	Multiplicative Scattering Correction (MSC)	53
	Peak normalization	25
Derivative	Savitzky-Golay	51
Centering	Last Point	
	At point	
	Mean	
Dataset Operations	Ignore	
	Subtract	
	Reset	

preprocessing steps are then performed by pressing the button labeled "Apply preprocessing." If the student wants more steps they can add more using the "Add step" button. Finally, students can export the data using "Save data," where they can save to .csv or Excel files. On the right panel, the plot will automatically replot the data after calculations, resetting bounds and dynamically renaming axis titles where appropriate. Using the standard plotting widgets of Matplotlib, students can pan, zoom, and save the plot to an image.

FUTURE OPPORTUNITIES AND LEARNING METHODS

Porchlight would be best used as part of laboratory-based spectroscopy or chemometrics courses, where simple or advanced spectral analysis would showcase the benefits or detriments certain techniques may have. Graduates, upper-level undergraduates in specialty electives, and those working with spectroscopy research would benefit the most from such lessons. Regardless, a number of possibilities exist that would allow students to explore the importance of preprocessing with a few ideas presented here. The first lesson plan would be as a wholly *in silico* lab, where the students are presented with a dataset with a known defect. Some example lectures are provided in the GitHub of the project as Jupyter notebooks with sample data, primarily showcasing vibrational spectra in regression and classification objectives. The instructor would present relevant preprocessing techniques, how they work, and the type of defect they correct, and provide literature where appropriate as additional learning material. Then the students would proceed to use the lessons from the lecture to preprocess data and, for example, compare techniques in multivariate calibration. They should find that some techniques yield superior results in modeling than others and should be able to explain qualitatively why the chosen technique works. This would fit nicely into a chemometrics lesson, where students can see the challenges that are present in real data and begin associating challenges with solutions.

Alternatively, instructors may incorporate our preprocessing application into a larger lab-based module, possibly even with another computer-based module such as RamanCAT.[13] In this case, students collect the spectra themselves and attempt to tackle any shortcomings with preprocessing techniques. This would allow students to see the whole spectroscopic data collection pipeline: parameter selection, collection, preprocessing, and finally analysis. It is worth nothing that a key factor in such a lab or course would be to clearly underscore to students that preprocessing is not a magic cure—bad data are bad data. The onus is on the spectroscopist to collect the best possible data they can and then use preprocessing to improve the interpretation.

Finally, a third lab concept is oriented toward graduate students and students in research. One of the challenges of preprocessing is that, done incorrectly, can lead to (a) "beautified" bad data, (b) altered interpretation of the data and, (c) wrong predictive models.

We can introduce students to these concerns by challenging them to alter data. A simple case is to explore the impact of smoothing, and how too much smoothing can look nicer but the information contained in the spectra is diminished. Alternatively, one can present a region that is meaningless but may look like it has information, such as the diamond ATR region, and try to preprocess it until it successfully predicts a set of *y*-variables. By seeing firsthand the impact that poor preprocessing has, students will learn to scrutinize the literature.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.2c00812.

Manual describing installation and use of Porchlight (PDF)

AUTHOR INFORMATION

Corresponding Author

George Tsilomelekis − *Department of Chemical and Biochemical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-0435-8216; Email: g.tsilo@rutgers.edu

https://doi.org/10.1021/acs.jchemed.2c00812
J. Chem. Educ. 2023, 100, 1326−1332

Author

Jakub A. Konkol − *Department of Chemical and Biochemical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-6488-2167

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.2c00812

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This material is based upon work supported in part by Rutgers, The State University of New Jersey, and the National Science Foundation Awards 1751683 and 1751683.

■ REFERENCES

(1) Wankat, P. C. Integrating the Use of Commercial Simulators into Lecture Courses. *Journal of Engineering Education* **2002**, *91*, 19−23.

(2) Roman, C.; Delgado, M. A.; García-Morales, M. Using process simulators in Chemical Engineering education: Is it possible to minimize the "black box" effect? *Computer Applications in Engineering Education* **2020**, *28*, 1369−1385.

(3) Moodley, K. Improvement of the learning and assessment of the practical component of a Process Dynamics and Control course for fourth year chemical engineering students. *Education for Chemical Engineers* **2020**, *31*, 1−10.

(4) Psycharis, S. Examining the effect of the computational models on learning performance, scientific reasoning, epistemic beliefs and argumentation: An implication for the STEM agenda. *Computers & Education* **2013**, *68*, 253−265.

(5) Koretsky, M. D. An interactive virtual laboratory addressing student difficulty in differentiating between chemical reaction kinetics and equilibrium. *Computer Applications in Engineering Education* **2020**, *28*, 105−116.

(6) Albrecht, J. Step Into the Digital Age with Python. *Chemical Engineering Progress* **2021**, *117*, 47−53.

(7) *Teaching Programming across the Chemistry Curriculum*; McDonald, A. R., Nash, J. A., Eds.; American Chemical Society, 2021.

(8) Sengupta, I. Illustrating Elementary NMR Concepts through Simple Interactive Python Programs. *J. Chem. Educ.* **2021**, *98*, 1673−1680.

(9) Jablonka, K. M.; Patiny, L.; Smit, B. Making Molecules Vibrate: Interactive Web Environment for the Teaching of Infrared Spectroscopy. *J. Chem. Educ.* **2022**, *99*, 561−569.

(10) Pfister, K. L. *Teaching spectroscopy through the use of computer simulation*. PhD thesis, Middle Tennessee State University, Murfreesboro, TN, 1991.

(11) Bertrand, G. L. Spectrophotometry. https://web.mst.edu/~gbert/Color_Lg/color.html?455, University of Missouri—Rolla, 2002.

(12) Bigger, S. W.; Bigger, A. S.; Ghiggino, K. P. FluSpec: A Simulated Experiment in Fluorescence Spectroscopy. *J. Chem. Educ.* **2014**, *91*, 1081−1083.

(13) Marlowe, J.; Tsilomelekis, G. Accessible and Interactive Learning of Spectroscopic Parameterization through Computer-Aided Training. *J. Chem. Educ.* **2020**, *97*, 4527−4532.

(14) Antonelli, T. M.; Olivieri, A. C. Developing and Implementing an R Shiny Application to Introduce Multivariate Calibration to Advanced Undergraduate Students. *J. Chem. Educ.* **2020**, *97*, 1176−1180.

(15) Öberg, T. Introducing Chemometrics to Graduate Students. *J. Chem. Educ.* **2006**, *83*, 1178.

(16) Wanke, R.; Stauffer, J. An Advanced Undergraduate Chemistry Laboratory Experiment Exploring NIR Spectroscopy and Chemometrics. *J. Chem. Educ.* **2007**, *84*, 1171.

(17) Chau, F. T.; Chung, W. H. Using Matlab to Assist Undergraduates in Learning Chemometrics. *J. Chem. Educ.* **1995**, *72*, A84.

(18) Roger, J.-M.; Boulet, J.-C.; Zeaiter, M.; Rutledge, D. N. *Comprehensive Chemometrics*; Elsevier, 2020; pp 1−75.

(19) Baker, M. J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H. J.; Dorling, K. M.; Fielden, P. R.; Fogarty, S. W.; Fullwood, N. J.; Heys, K. A.; Hughes, C.; Lasch, P.; Martin-Hirsch, P. L.; Obinaju, B.; Sockalingum, G. D.; Sulé-Suso, J.; Strong, R. J.; Walsh, M. J.; Wood, B. R.; Gardner, P.; Martin, F. L. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* **2014**, *9*, 1771−1791.

(20) Butler, H. J.; Ashton, L.; Bird, B.; Cinque, G.; Curtis, K.; Dorney, J.; EsmondeWhite, K.; Fullwood, N. J.; Gardner, B.; Martin-Hirsch, P. L.; Walsh, M. J.; McAinsh, M. R.; Stone, N.; Martin, F. L. Using Raman spectroscopy to characterize biological materials. *Nat. Protoc.* **2016**, *11*, 664−687.

(21) Gaigneaux, A.; Ruysschaert, J. M.; Goormaghtigh, E. Cell Discrimination by Attenuated Total Reflection—Fourier Transform Infrared Spectroscopy: The Impact of Preprocessing of Spectra. *Appl. Spectrosc.* **2006**, *60*, 1022−1028.

(22) Lee, L. C.; Liong, C.-Y.; Jemain, A. A. A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum. *Chemometrics and Intelligent Laboratory Systems* **2017**, *163*, 64−75.

(23) Liland, K. H.; Kohler, A.; Afseth, N. K. Model-based pre-processing in Raman spectroscopy of biological samples. *J. Raman Spectrosc.* **2016**, *47*, 643−650.

(24) Rinnan, A. Pre-processing in vibrational spectroscopy - when, why and how. *Anal. Methods* **2014**, *6*, 7124−7129.

(25) Gautam, R.; Vanga, S.; Ariese, F.; Umapathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ. Techniques and Instrumentation* **2015**, *2*, 8.

(26) Lasch, P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemometrics and Intelligent Laboratory Systems* **2012**, *117*, 100−114.

(27) Gabrielsson, J.; Jonsson, H.; Airiau, C.; Schmidt, B.; Escott, R.; Trygg, J. OPLS methodology for analysis of pre-processing effects on spectroscopic data. *Chemometrics and Intelligent Laboratory Systems* **2006**, *84*, 153−158.

(28) Liland, K. H.; Almøy, T.; Mevik, B.-H. Optimal Choice of Baseline Correction for Multivariate Calibration of Spectra. *Appl. Spectrosc.* **2010**, *64*, 1007−1016.

(29) Rato, T. J.; Reis, M. S. SS-DAC: A systematic framework for selecting the best modeling approach and pre-processing for spectroscopic data. *Comput. Chem. Eng.* **2019**, *128*, 437−449.

(30) Bian, X.; Wang, K.; Tan, E.; Diwu, P.; Zhang, F.; Guo, Y. A selective ensemble preprocessing strategy for near-infrared spectral quantitative analysis of complex samples. *Chemometrics and Intelligent Laboratory Systems* **2020**, *197*, 103916.

(31) Guezenoc, J.; Gallet-Budynek, A.; Bousquet, B. Critical review and advices on spectralbased normalization methods for LIBS quantitative analysis. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2019**, *160*, 105688.

(32) Rinnan, A.; Berg, F. v. d.; Engelsen, S. B. Review of the most common preprocessing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry* **2009**, *28*, 1201−1222.

(33) Lasch, P. Cytospec. https://www.cytospec.com/index.php, accessed 2022-07-15.

(34) Aspen Technology Inc. Aspen Uncsrambler. https://www.aspentech.com/en/products/msc/aspen-unscrambler, accessed: 2022-07-15.

(35) Infometrix. Pirouette. https://infometrix.com/pirouette/, accessed 2022-07-15.

(36) Menges, F. Spectragryph—optical spectroscopy software. http://www.effemm2.de/spectragryph/, accessed 2022-07-15.

(37) Jarvis, R. M.; Broadhurst, D.; Johnson, H.; O'Boyle, N. M.; Goodacre, R. PYCHEM: a multivariate analysis package for python. *Bioinformatics* **2006**, *22*, 2565−2566.
(38) Schmidt, R. W.; Woutersen, S.; Ariese, F. RamanLIGHT—a graphical user-friendly tool for pre-processing and unmixing hyperspectral Raman spectroscopy images. *Journal of Optics* **2022**, *24*, 064011.
(39) Beattie, J. R.; Esmonde-White, F. W. L. Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra. *Appl. Spectrosc.* **2021**, *75*, 361−375.
(40) Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **2001**, *58*, 109−130.
(41) TIOBE. TIOBE Index. https://www.tiobe.com/tiobe-index/, accessed, 2022-12-07.
(42) van Staveren, M. Integrating Python into a Physical Chemistry Lab. *J. Chem. Educ.* **2022**, *99*, 2604−2609.
(43) Menke, E. J. Series of Jupyter Notebooks Using Python for an Analytical Chemistry Course. *J. Chem. Educ.* **2020**, *97*, 3899−3903.
(44) Weiss, C. J. Scientific Computing for Chemists: An Undergraduate Course in Simulations, Data Processing, and Visualization. *J. Chem. Educ.* **2017**, *94*, 592−597.
(45) Domínguez, J.; Alonso, M.; González, E.; Guijarro, M.; Miranda, R.; Oliet, M.; Rigual, V.; Toledo, J.; Villar-Chavero, M.; Yustos, P. Teaching chemical engineering using Jupyter notebook: Problem generators and lecturing tools. *Education for Chemical Engineers* **2021**, *37*, 1−10.
(46) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array programming with NumPy. *Nature* **2020**, *585*, 357−362.
(47) McKinney, W. Data Structures for Statistical Computing in Python. *Proceedings of the Python in Science Conference*; 2010.
(48) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9*, 90−95.
(49) Lieber, C. A.; Mahadevan-Jansen, A. Automated Method for Subtraction of Fluorescence from Biological Raman Spectra. *Appl. Spectrosc.* **2003**, *57*, 1363−1367.
(50) Eilers, P. H. C.; Boelens, H. F. M. *Baseline Correction with Asymmetric Least Squares Smoothing*; Techreport, 2005.
(51) Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627−1639.
(52) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* **1989**, *43*, 772−777.
(53) Geladi, P.; MacDougall, D.; Martens, H. Linearization and Scatter-Correction for NearInfrared Reflectance Spectra of Meat. *Appl. Spectrosc.* **1985**, *39*, 491−500.

https://doi.org/10.1021/acs.jchemed.2c00812
J. Chem. Educ. 2023, 100, 1326−1332