

Open-source, Extensible Software for Advanced Spectroscopic Analysis

Zachery Crandall¹ and Paul Hinker²
Mathematics and Computer Science
South Dakota School of Mines and Technology
Rapid City, SD 57701
¹*zachcran@gmail.com*
²*Paul.Hinker@sdsmt.edu*

Abstract

Chemometrics, the data-driven extraction of information to analyze chemical systems, is a highly researched topic in forensics analysis. When partial least squares (PLS) chemometrics techniques are coupled with attenuated total internal reflectance Fourier transform infrared (ATR-FTIR) spectroscopy, researchers have successfully quantified cocaine in an unknown powder, measured caffeine in soft drinks, and classified fraudulent bovine meat. PLS chemometrics techniques are available through expensive, stand-alone software, higher levels of instrumentation software licenses, or packages in statistical computer languages like R. Interested institutions without the financial or programming resources are limited in their ability to participate in this research area.

An open-source, extensible spectroscopic analysis program was developed to facilitate the advancement of the chemometrics field. This software intends to remove the pay wall of advanced spectroscopic analysis methods and increase ease-of-use of chemometrics methods. It will provide a user-friendly and extensible platform on which scientists can freely develop and utilize spectroscopic analysis methods. This extensibility allows users to add new functionality such as instrumentation support, additional spectral file types, spectral processing options, and quantification methods. New program windows and graphics can also be added to support the added functionalities. The software currently supports ATR-FTIR data and can be extended to other instruments.

1 Introduction

Forensic laboratories have a large backlog of evidence to analyze because forensic methods are not fast enough to meet the demand of the industry. Rapid methodologies for quantifying controlled substance are needed to mitigate this issue.[1] With a lack of cost-effective software, institutions and researchers cannot always gain access to the necessary software for instrumental research and education necessary to meet this need.

By providing open-source, extensible spectral processing software, this disconnect can be corrected. New popular methods for analyzing pharmaceuticals and controlled substances are Fourier transform infrared (FTIR) and Raman spectroscopy.[2] FTIR is recommended by United Nation Office on Drugs and Crime and the Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) for controlled substance analysis.[3-5] Standard transmission FTIR methods require painstaking sample preparation and are destructive to the sample.[6] Attenuated total reflectance (ATR-FTIR) is semi-nondestructive and requires little to no sample preparation, allowing for a simple, rapid analysis method that can even be employed in the field through a handheld spectrometer.[7]

Transmission FTIR requires the infrared laser to pass through the sample pellet or salt plate. ATR-FTIR uses reflections through a crystal in direct contact with the sample.[7] This allows many interactions with the surface of the sample (Figure 1). This results in a spectrum, which is a plot of the % transmittance or absorbance versus the wavenumber of light. A wavenumber is the reciprocal of the wavelength of light, typically in the units cm^{-1} . [7] The spectrum these wavenumbers create is unique to the chemicals, and sometimes amount of the chemicals, in the sample. Since these reflections are random, the spectra can be inconsistent, requiring additional spectral processing before use.[8]

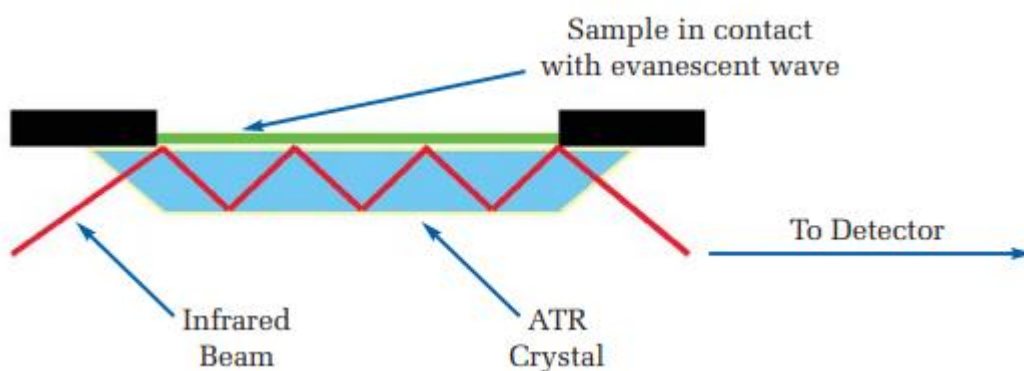


Figure 1: Diagram of the laser interacting with the sample is shown to illustrate the multiple reflections and surface penetration of ATR-FTIR.[9]

Recent literature has analyzed ATR-FTIR as a quantification method by using chemometrics techniques. A particularly popular method employed is partial least squares (PLS) analysis.[10]-[16] PLS is advantageous over linear models, like Beer's Law plots, since it can include multiple variables in the model. PLS is a type of inverse least squares method, which is a class of models characterized by the general equation

$c = Rb$, where c is a vector of concentrations, R is a matrix of the spectra, and b is a vector containing the model coefficients. Using training spectra for R , with known concentrations of your analyte c , you can find b . This b can be used on new unknown spectra R , predicting the concentration of the analyte in each spectrum. Inverse models differ in the way they solve for b and predict unknowns. PLS, unlike other models, incorporates the covariance of the concentration and the spectra.[17]

Partial least squares methods are versatile for spectral analysis, as the researcher can choose to use the full spectrum or only a specific portion of the spectrum. Noise modelling and compensation are also built into PLS and when multiple spectra of the same sample are given the multivariate effect, a signal averaging of the spectra, will occur. There are drawbacks to this technique, however. PLS models are sensitive to unusual compounds, so unreliable results can occur if proper care is not taken during calibration and analysis. Recent literature has made an attempt to reduce this issue by reviewing various data preprocessing techniques and attempting to provide recommendations of when to apply each method.[8]

Despite the popularity of partial least squares in spectroscopic analysis, there exists no free software for this application. To create an open-source, free replication of this method for researchers, the statistical computing language, R, was used.[18] R provides the ability for users to make their own packages, and is under the GNU Public License, allowing for easy access to the *pls* package.[19]

With a lack of cost-effective software, institutions and researchers cannot always gain access to the necessary software for instrumental research and education. By providing open-source, extensible spectral processing software, this disconnect can be corrected. An open-source implementation of partial least squares regression on FTIR spectra is provided in the base software, tested using unknowns in a single blind study. Further instrumentation support, spectral processing techniques, unknown prediction techniques, windows, and graphics can be added as well.

2 Software Layout

2.1 Layout and Implementation

This software was developed according to a Model-View-Controller (MVC) architectural pattern in a C# Windows Form Application. Using an Object-Oriented Programming (OOP) approach, this tool was designed to be extensible through the addition of new method classes, file classes, and page classes. Reflection is used to detect and compile all classes, to which the software will extend as necessary to incorporate this information.

2.2 Workflow

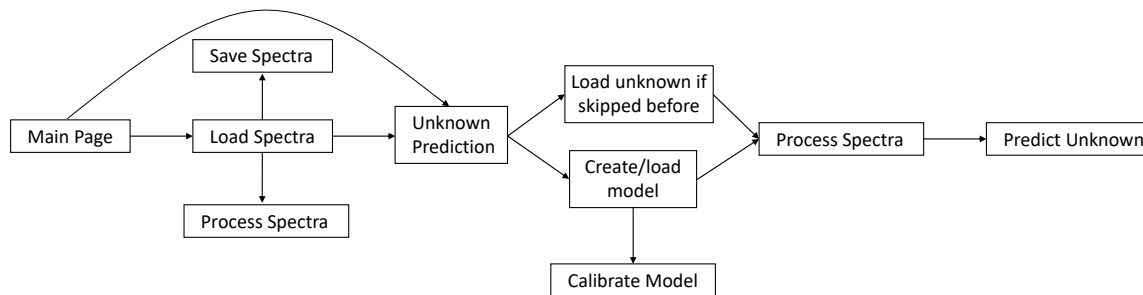


Figure 2: Work flow of the software.

The workflow of the program is shown in Figure 2. On start, the user will be navigated to the main page with a spectral display. On the main page, loading, saving, and basic processing of spectra is possible. The user can choose to navigate to the unknown prediction page. An unknown spectrum, along with a model, can be loaded from files. If a premade model does not exist, one can be created, calibrated, and diagnosed. Once the files are loaded, basic spectral processing can be applied to both the unknown and model. The concentration of analyte in the unknown can be predicted once the user has tuned the model to their needs.

3 Software Interfaces

Specific C# class interfaces and abstract classes have been designed which newly added classes must inherit from to be added to the software. The interfaces are the ISpectralFile for spectral data file types, IProcessor for spectral processing routines, and IUnknown for unknown prediction routines. An abstract class is provided for all new pages to build upon to ensure they are properly integrated into the software. Following are the required functions and properties for each interface or abstract class.

3.1 ISpectralFile

The required functions and properties for ISpectralFile are contained in Table 1.

Function or Property	Description
Composition	Look-up for molecular or elemental compositions
Extension	File extension to apply this class to
Filename	Name of the loaded file
Intensities	Intensities at each wavenumber (parallel to Wavenumbers)
IntensityUnits	Units of Intensities
Read	Read from the file
Wavenumbers	Wavenumbers of the spectrum
WavenumberUnits	Units of Wavenumbers
Write	Write spectrum to a file of the class type

Table 1: Required functions and properties of ISpectralFile.

3.2 IProcessor

IProcessor requires a single function be implemented, process. This function will apply the given spectral processing routine to a given spectrum.

3.3 IUnknown

The required functions for IUnknown are contained in Table 2.

Function	Description
calibrate	Calibrate the prediction model using the given spectra
predict	Apply the unknown prediction routine to the given spectrum

Table 2: Required functions of IUnknown.

3.4 PageBase

The required functions and properties for PageBase are contained in Table 3.

Function or Property	Description
analyte	Stores the analyte spectrum for page transitions
fileManager	The file manager class used to load and save spectra
getSpectrum	Accessor for analyte. Used in page transitions.
setFileManager	Sets the file manager to be used by the page
setSpectrum	Sets analyte. Used in page transitions.
showSpectrum	Display routine for the analyte spectrum
toolStrip	The ToolStrip menu located at the top of each page

Table 3: Required functions and properties of PageBase.

4 Current Functionalities

Currently, this software supports basic processing techniques of ATR-FTIR spectral data along with an advanced unknown prediction method. The advanced unknown prediction method implemented is partial least squares regression. Additionally, all spectra can be processed using a myriad of spectral data processing routines.

4.1 Partial Least Squares Regression

A custom open-source PLS regression routine is used for unknown prediction. This routine applies the R *pls* package to perform the necessary calculations.[19] A third-party assembly, R.NET, is included with the software to allow the use of R inside C#.[20] The PLS regression routine used was tested in a single blind study to verify the quality of the prediction.

In the test study, mixtures of caffeine and gluten-free, lactose-free Nestle Coffe-Mate Coffee Creamer: The Original were used to create a PLS model and. Mixtures from 10% weight by weight (w/w) caffeine to 90%w/w caffeine were then created and labelled with letters. An analyst then performed the ATR-FTIR readings and PLS predictions for the content of caffeine in these lettered unknowns. All unknowns below 60%w/w caffeine were correctly predicted within 5%w/w caffeine. After analysis, it was concluded that the discrepancies at higher concentrations of caffeine were caused by saturation of the peaks in the FTIR spectrum, not the PLS routine.

4.2 Spectral Data Processing

Support currently exists for spectral normalization, scatter correction (standard normal variate and multiplicative scatter correction), mean centering, autoscaling, and baseline correction and derivative algorithms.

ATR-FTIR spectra often vary due to sample size or quantity when measuring samples, which spectral normalization is commonly used to resolve. Intrinsic limitations or flaws of an ATR-FTIR instrument may also cause variations in a spectrum, which can be mitigated by scatter correction techniques such as standard normal variate and multiplicative scatter correction. Mean centering and autoscaling are typically applied to remove variations caused by different mean or standard deviations. An ATR-FTIR also typically experiences an elevated baseline, which can be corrected for using spectral baseline correction or derivative algorithms, removing irrelevant variations, offset baselines, and resolving overlapped peaks.

5 Conclusions

Through this open-source software, spectroscopists will have easy, free access to spectral processing and PLS regression analysis techniques. This will allow more rapid research by removing the licensing and pay barriers typically impeding progress in less-endowed institutions. One impact of particular interest to forensic research is reduced turnaround times on controlled substance analysis in forensic laboratories. In turn, faster analysis in a case could reduce persecution times of drug offenders including drug dealers, hopefully reducing the availability of harmful controlled substances currently at large. This software, being open and free to use, will also be accessible to teaching institutions to allow students to learn spectral processing without the need for proprietary software that only comes with an instrument.

6 Acknowledgements

The authors would like to express their sincerest thank you to the U.S. Department of Homeland Security (DHS) Science and Technology Directorate Office of University Programs for sponsoring the summer internship program that lead to this research, Oak Ridge Institute for Science and Education (ORISE) for administering the summer program, and, through the management of ORISE, Oak Ridge Associated Universities and the Department of Energy. Additionally, they would like to thank the employees at the CBP LSSD Springfield Laboratory for the knowledge and assistance that was provided during the summer internship project. The authors would also like to thank the South Dakota School of Mines and Technology's Mathematics and Computer Science department for their continued support of the graduate students of the department.

References

- [1] P. J. Speaker, "Opioids and the Drain on Laboratory Resources: Financial Modeling & Planning for Resource Allocation," in *Online Seminar from West Virginia University*, 2017.
- [2] S. Harkai and M. Pütz, "Comparison of rapid detecting optical techniques for the identification of New Psychoactive Substances in 'Legal High' preparations," *Toxichem Krimtech*, vol. 82, no. (Special Issue), pp. 229-238, 2015.
- [3] *Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) Recommendations*, U. S. D. o. J. D. E. Administration, 2016.
- [4] *Recommended methods for Identification and Analysis of Synthetic Cannabinoid Receptor Agonists in Seized Materials*, 2013.
- [5] *Recommended Methods for Identification and Analysis of Synthetic Cathinones in Seized Materials*, 2015.
- [6] R. M. Silverstein, F. X. Webster, D. J. Kiemle, and D. L. Bryce, "Infrared Spectroscopy," in *Spectrometric Identification of Organic Compounds* 8 ed.: Wiley, 2015.
- [7] S. Bell, "Instrumentation," in *Forensic Chemistry* 2 ed.: Pearson, 2013.
- [8] L. C. Lee, 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*, vol. 163, pp. 64-75, 2017.
- [9] "FT-IR Spectroscopy Attenuated Total Reflectance (ATR) Technical Note," PerkinElmer Life and Analytical Sciences 2005, Available: https://shop.perkinelmer.com/content/TechnicalInfo/TCH_FTIRATR.pdf.
- [10] N. V. S. Rodrigues, E. M. Cardoso, M. V. O. Andrade, C. L. Donnici, and M. M. Sena, "Analysis of Seized Cocaine Samples by using Chemometric Methods and FTIR Spectroscopy," *J. Braz. Chem. Soc.*, vol. 24, no. 3, pp. 507-517, 2013.
- [11] T. S. Grobério, J. J. Zacca, M. Talhavini, and J. W. B. Braga, "Quantification of Cocaine Hydrochloride in Seized Drug Samples by Infrared Spectroscopy and PLSR," *J. Braz. Chem. Soc.*, vol. 25, no. 9, pp. 1696-1703, 2014.
- [12] T. S. Grobério, J. J. Zacca, É. D. Botelho, M. Talhavini, and J. W. B. Braga, "Discrimination and quantification of cocaine and adulterants in seized drug samples by infrared spectroscopy and PLSR," *Forensic Science International*, vol. 257, pp. 297-306, 2015.
- [13] M. C. A. Marcelo, K. C. Mariotti, M. R. Ferrão, and R. S. Ortiz, "Profiling cocaine by ATR-FTIR," *Forensic Science International*, vol. 246, pp. 65-71, 2015.
- [14] D. Mainali and J. Seelenbinder, "Automated Fast Screening Method for Cocaine Identification in Seized Drug Samples Using a Portable Fourier Transform Infrared (FT-IR) Instrument," *Applied Spectroscopy*, vol. 70, no. 5, pp. 916-922, 2016.
- [15] K. M. Nunes, M. V. O. Andrade, A. M. P. S. Filho, M. C. Lasmar, and M. M. Sena, "Detection and characterisation of frauds in bovine meat in natura by non-meat ingredient additions using data fusion of chemical parameters and ATR-FIR spectroscopy," *Food Chemistry*, vol. 205, pp. 14-22, 2016.

- [16] S. Materazzi, A. Gregori, L. Ripani, A. Apriceno, and R. Risoluti, "Cocaine profiling: Implementation of a predictive model by ATR-FTIR coupled with chemometrics in forensic chemistry," *Talanta*, vol. 166, pp. 328-335, 2017.
- [17] K. R. Beebe, R. J. Pell, and M. B. Seasholtz, *Chemometrics: A Practical Guide*. John Wiley & Sons, Inc., 1998.
- [18] *What is R?* Available: <https://www.r-project.org/about.html>
- [19] B.-H. Mevik, R. Wehrens, and K. H. Liland, "Partial Least Squares and Principal Component Regression," 2.6-0 ed. CRAN Repository, 2016.
- [20] jperraud *et al.*, "R.NET," ed, 2015.