

Automated Raman Spectra Analysis with Machine Learning Integration for Material Identification and Decomposition Monitoring

DIRECT
Data Intensive Research
Enabling Clean Technologies

Brandon Kern², Elizabeth Rasmussen^{1,3}, Parker Steichen²

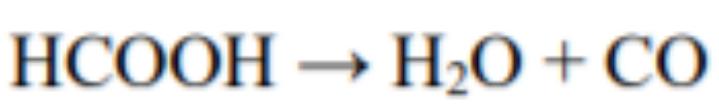
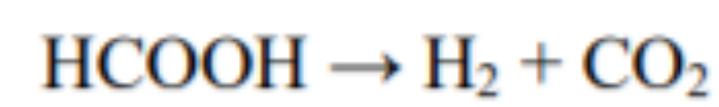
¹Department of Mechanical Engineering, ²Department of Material Science and Engineering, ³Clean Energy Institute
University of Washington, Seattle, WA, USA

CLEAN ENERGY INSTITUTE
UNIVERSITY of WASHINGTON

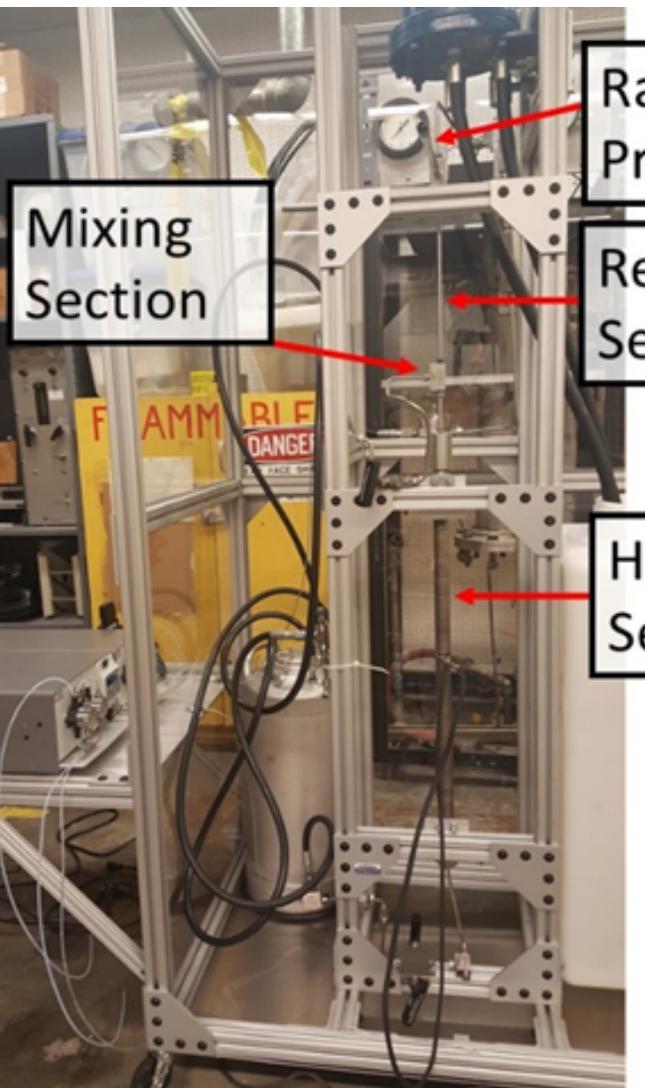
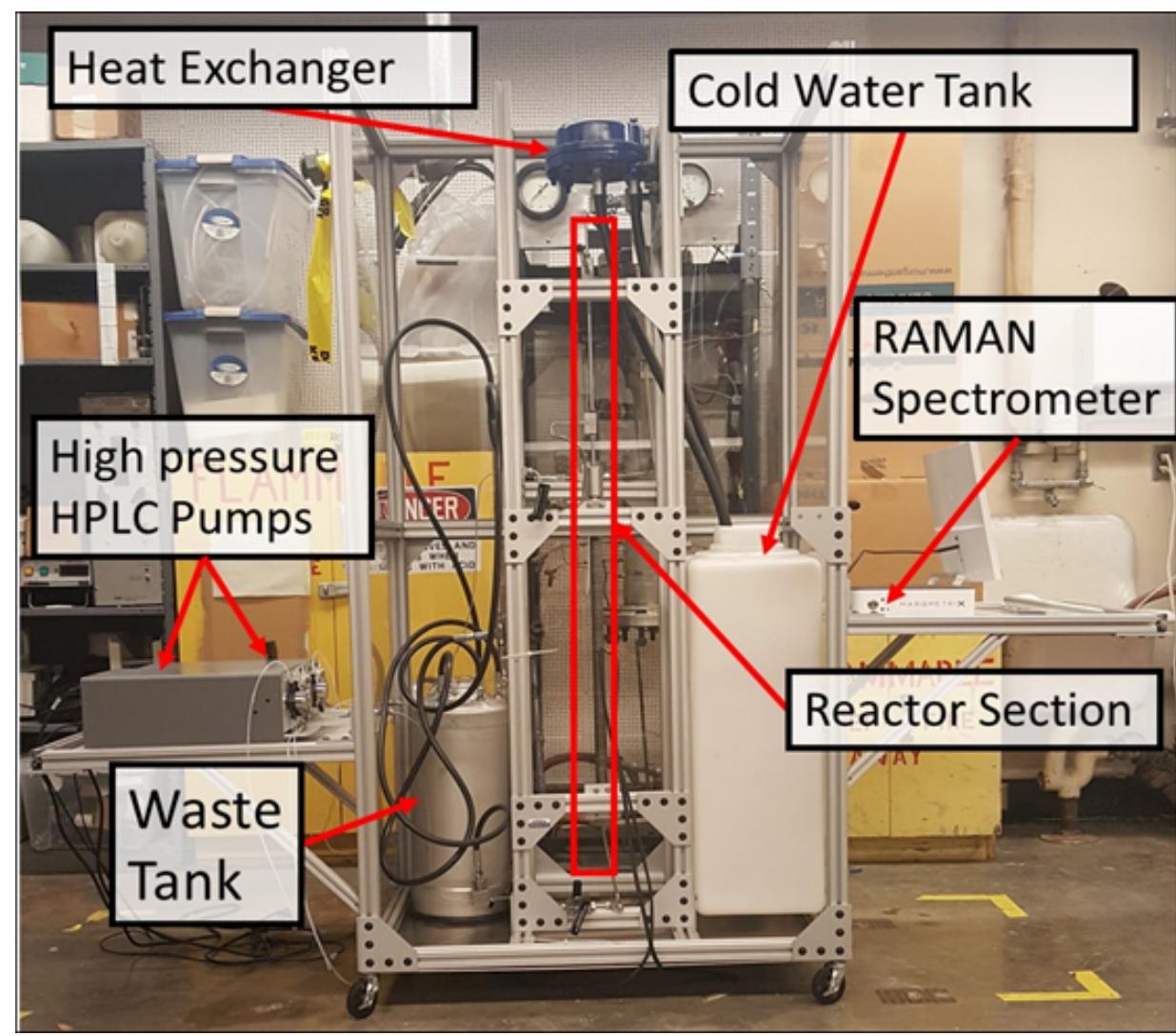
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Introduction

- Using data from a custom built supercritical gasification reactor on campus to analyze formic acid Raman spectra.
- Decomposition of formic acid constitutes the combination of two pathways:



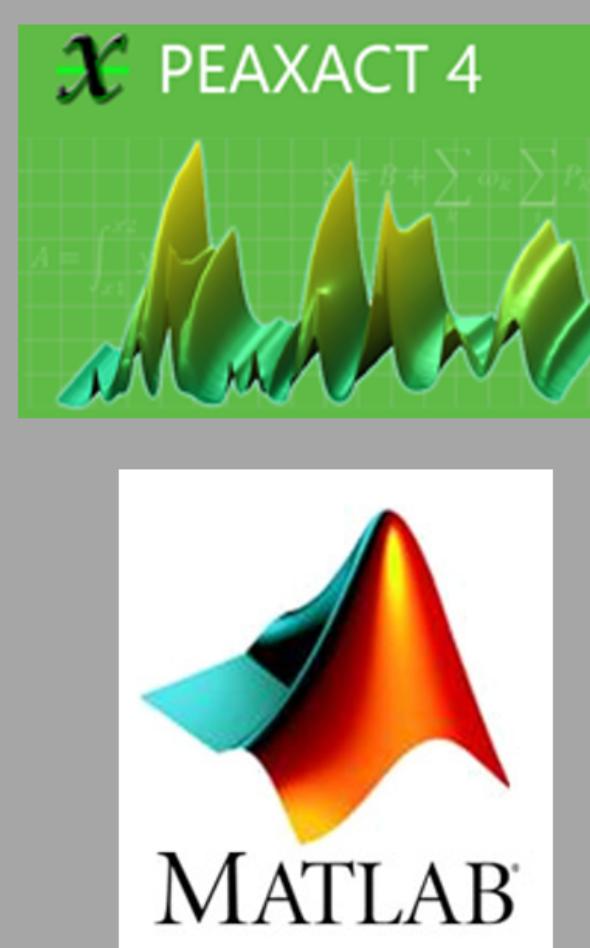
Photographs of UW Gasification Reactor with In-Situ Raman Probe



Motivation

Prior Analysis Method

- Commercial Software suite with yearly license cost of ~ \$3,000
- Not Open Source → Research backbone can be out of date
- Very tedious manual processes



Raman Noodles

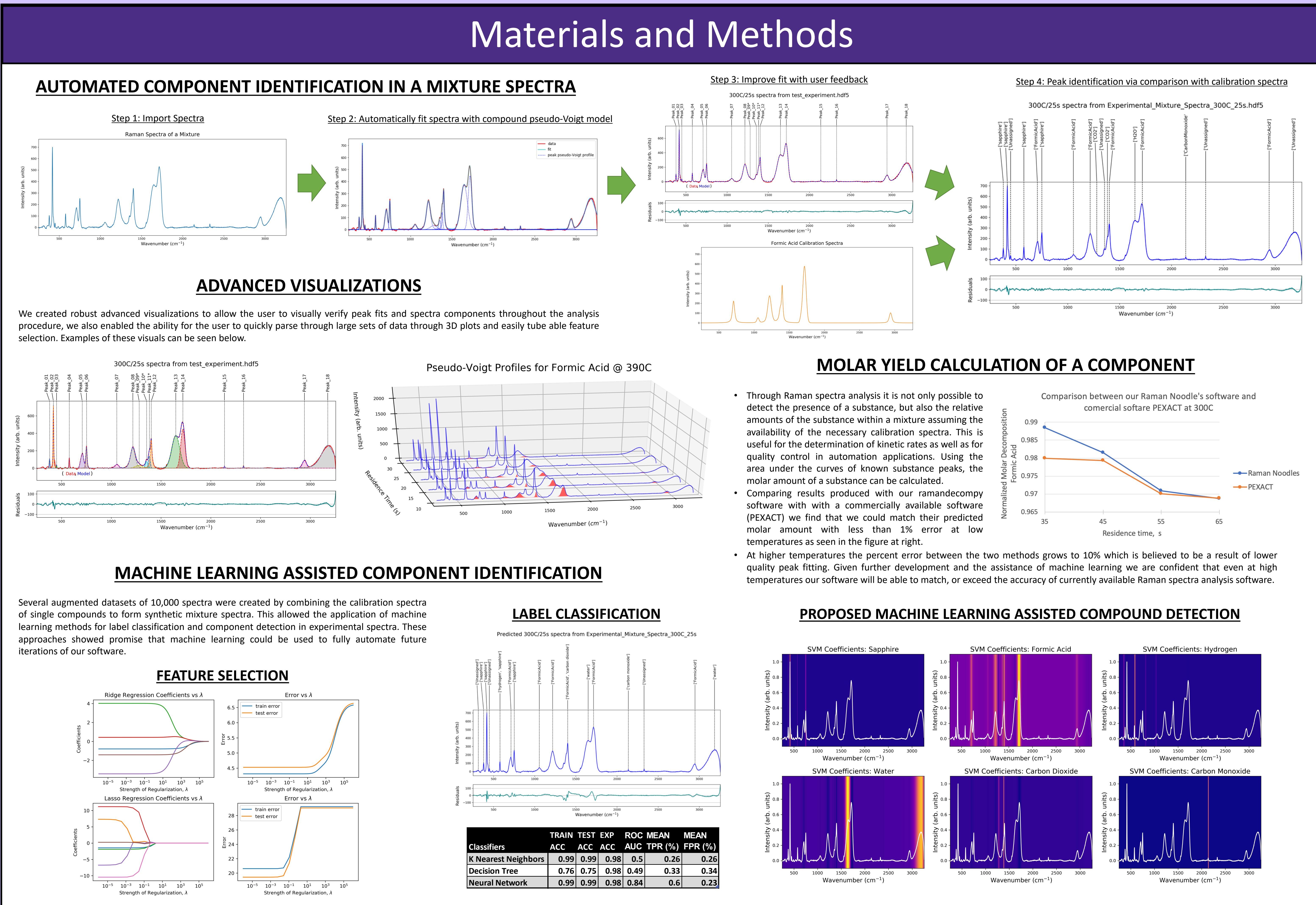
New Method!

- Free
- Open Source + Customizable → Research backbone easily updated by anyone
- Automated processes



Goals

- Feature Rich Peak Identification Tools for Raman Spectra:
 - Automatic peak detection, compound pseudo-Voigt model generation, and compound identification
- Data Visualization:
 - Suite of tools for visualizing results and tracking progress of analysis
- Molar Decomposition Calculations:
 - For quality control analysis and future formation calculations
- Machine Learning:
 - Demonstrate the potential for the application of ML approaches to improve spectra fitting and fully automate molar yield calculations



Conclusions and Future Work

In conclusion, our team successfully created a platform code base for researchers to visualize and analyze their Raman spectra in a fast, automated manner – reducing post-processing time by days and enabling future work to continue on a solid base of open source tools.

Future Work: Next steps would include automatic baseline subtraction of spectra to decrease pre-processing time, molar formation calculations to predict reaction pathways, and increased robustness of machine learning for component selection including unsupervised methods.

Acknowledgements

- Igor Novoselov, Dave Beck, and Kelly Thornton
- NSF DIRECT Program, Clean Energy Institute, USA DETRA HDTRA1-17-1-0001
- Data sets from publicly available from Mendeley Data, "Raman Spectra of Formic Acid Gasification Products in Subcritical and Supercritical Water"
- Only open source packages were used in this work, documentation of all packages used can be found at our GitHub at:
 - <https://github.com/raman-noodles/ramandecompy>