A Data Science Approach to Automated Raman Spectra Analysis

Dominic Conricode

# Scope

* Define an algorithmic approach for fitting Raman spectra with a series of peaks
  + Extract parameters that correlate to graphite microstructural properties
* Use this technique to characterize graphite grades over a range of temperature values

# Background and Motivation

* Graphite is used in MSRs, FHRs, and HTGRs for various purposes (and non-nuclear applications)
  + Important to understand characteristics of different graphite grades to choose the most beneficial one for a given application
* Semi-automatic fitting does exist, but it requires a human to oversee and manually adjust when it gets “stuck”
* A fully automated approach would allow for analysis at a larger scale

# Approach

## Pre-Processing

* Input: CSV file(s) from Raman spectroscopy with “x” and “y” values

Chart, histogram

Description automatically generated

Raw Input Data

* Smooth data with Savitsky-Golay filter
  + Parameters for smoothing can be passed in by user
* Normalize spectrum (easier to make normalized initial guesses)
* Restrict “x” values to between 500 and 3000
  + The peaks we are interested in are known in the literature to not be outside this range

Chart, histogram

Description automatically generated

Savitsky-Golay filter applied to raw data with a window-length of 27 and a 2nd degree polynomial fitting

* Remove background noise by fitting baseline (done automatically)
* Iterative baseline algorithm:
  + Fit a polynomial to the entire spectrum
  + Remove data points above the fit line
  + Re-fit to new spectrum with less data points
  + Repeat until the error between fit and data is within threshold or less than a certain number of points is getting removed between iterations

Chart, histogram

Description automatically generated

Fitting baseline with a 2nd degree polynomial

## Fitting

* Uses scipy.optimize.minimize package to minimize residuals between reconstructed spectrum and real spectrum
* Initial Guesses: (Gaussian/Lorentzian peaks can be defined in terms of peak center, amplitude, and area) (User can select which peak type to use)
  + Center: peaks known to be in given range from literature, choose middle value of this range
  + Amplitude: 1 (subject to change?)
  + Area: 1 (subject to change?)
* Loss function (error):
  + Sum the curves at each “x” value
  + Distance between reconstructed curve and real curve
* Constraints
  + Area of G’\_3Da / G’\_3Db approximately 0.5
    - Known from literature
  + Center of G’\_3Db > G’\_2D (these peaks’ center range overlaps, other peaks don’t have this issue)

Table

Description automatically generated

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

MAPE 0.0088

# Results (In Progress)

* Run this program on many graphite spectra spanning a wide range of grades and temperatures

# Analysis (In Progress)

* What can the extracted parameters tell us about the graphite

# Summary of Research

* Fit spectra, extract parameters, characterize graphite

# Summary of Research Experience

* Learned optimization techniques
* …