COMPARISON OF EXPERIMENTAL AND SIMULATED RAMAN SPECTRA THROUGH REVERSE SELF MODELING CURVE RESOLUTION FOR REGRESSION-BASED MACHINE LEARNING

NICOLE NORTH, ABIGAIL A ENDERS, HEATHER C. ALLEN, *Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA*.

Raman spectroscopy utilizes inelastic scattering to provide information about the vibrational environment of bonds within molecules. The intensity of vibrations can be cautiously used to determine relative concentrations of compounds. Machine learning methods are used to find patterns within datasets and commonly work better with large datasets, such as those from Raman spectral acquisitions; however, the generation of these large datasets is time consuming and the data manipulation is cumbersome. This work seeks to circumvent these issues by determining if experimental data can be fortified or substituted with simulated data. To initiate this process, Raman spectra were collected on a variety of different molecules in solutions made of 2-3 chemical species. These spectra were then used to simulate data with the same concentrations using Reverse Self Modeling Curve Resolution (RSMCR). The experimental and the RSMCR simulated data were used to train regression-type machine learning models. Models were then validated using previously withheld experimental Raman spectral data to determine how well each dataset worked as a generalizable basis for the regression model. We found that the RSMCR simulated data closely represents the experimental Raman spectral data with an approximate 2% error in the relative intensity.