

Supplementary Materials:

Quantifying Soil Inorganic Matter: Integrating SERDS with Machine Learning for Enhanced Analysis of Carbonates

Sara Poursorkh¹, Natalia Solomatova², Edward Grant¹

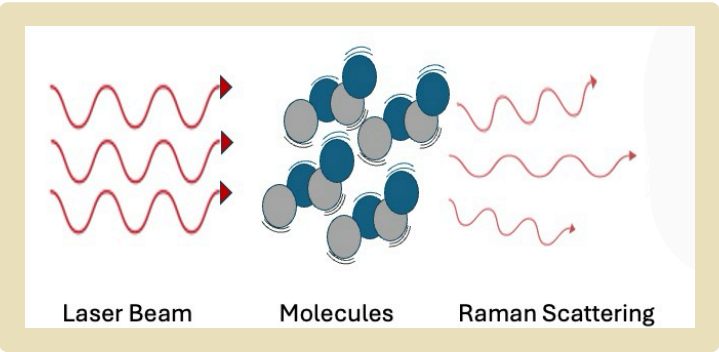
¹ Department of Chemistry, University of British Columbia, Vancouver, Canada ²Miraterra Technologies Inc, Vancouver, Canada



Raman Spectroscopy

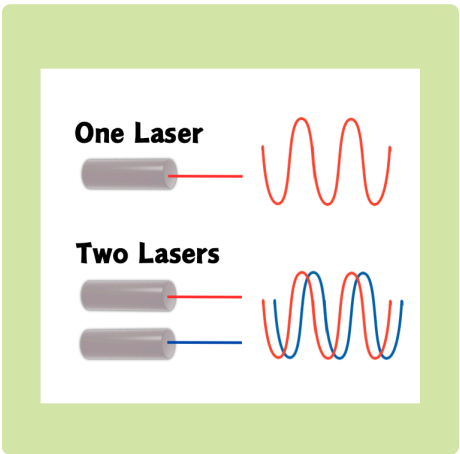
Raman spectroscopy is a powerful analytical technique used to identify and characterize the molecular composition of a material by detecting vibrational modes of its molecules. It works by shining a monochromatic laser light onto a sample and analyzing the scattered light. While most light scatters elastically (Rayleigh scattering), a small portion scatters inelastically, shifting in energy due to interactions with molecular vibrations—this is called Raman scattering. The resulting spectrum acts as a molecular fingerprint, with peaks corresponding to specific bond vibrations. Raman spectroscopy is widely used in chemistry, materials science, biology, and environmental studies because it is non-destructive, requires minimal sample preparation, and can provide detailed information about chemical structure and molecular interactions.

Conventional Raman Spectroscopy



Shifted Excitation Raman Difference Spectroscopy (SERDS)

Shifted Excitation Raman Difference Spectroscopy (SERDS) is an advanced Raman technique designed to eliminate fluorescence interference, which often masks weak Raman signals—especially in complex or organic-rich materials like soil, food, or biological samples. In SERDS, two slightly different laser excitation wavelengths are used to collect two Raman spectra from the same sample. Because fluorescence is broad and largely insensitive to small wavelength shifts, it remains nearly unchanged between the two spectra, while the Raman peaks shift with the excitation. By subtracting the two spectra, the fluorescence background is effectively canceled out, and the true Raman signal—representing the sample’s molecular structure—is revealed with higher clarity. This makes SERDS particularly valuable in applications where conventional Raman spectroscopy struggles due to strong background noise.



Machine Learning Models: PLSR and XGBoost

Partial Least Squares Regression (PLSR) is a statistical method used to find the relationship between two datasets, typically predictors (like spectra) and responses (like chemical concentrations, %C, etc.). It works by projecting both datasets into a new space using latent variables that capture the most relevant information for predicting the response. Unlike traditional regression, PLSR is especially useful when the predictors are highly collinear (correlated) or when there are more predictors than observations. It simplifies complex data, improves prediction accuracy, and is widely used in fields like chemometrics, spectroscopy, and machine learning.

Machine Learning Models: XGBoost

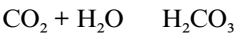
XGBoost is a powerful machine learning algorithm that builds a series of decision trees to make highly accurate predictions. It handles complex, non-linear relationships in data and is especially effective for large, noisy datasets like Raman spectra. In our work, XGBoost outperformed traditional models, providing faster and more accurate carbonate quantification.

Importance of Carbonates in soil

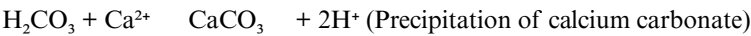
Soil carbonates are naturally occurring minerals that play a crucial role in pH buffering, helping stabilize soil acidity for healthy plant growth. More importantly, they are a major form of inorganic carbon storage in soils.

Carbonates act as a long-term carbon sink, locking away atmospheric CO₂ through weathering and mineralization processes. When CO₂ dissolves in water, it forms carbonic acid, which reacts with minerals in the soil to form stable carbonates like calcium carbonate. These compounds can remain in the soil for thousands of years, making them vital for climate regulation and carbon sequestration efforts.

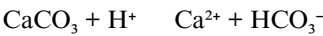
CO₂ Dissolution: Carbon dioxide forms carbonic acid in water.



Carbonate Formation (Mineralization)



pH Buffering: Neutralization of acid to stabilize soil pH



| Carbonates' Types | Description |
|--|---|
| Calcite <i>CaCO₃</i> | The most abundant and widespread form of carbonate in soils, especially in arid and semiarid regions. It plays a significant role in buffering soil pH. |
| Dolomite <i>CaMg(CO₃)₂</i> | A carbonate containing both calcium and magnesium, found in soils formed from dolomitic limestone or magnesium-rich parent materials. |
| Siderite <i>FeCO₃</i> | An iron carbonate found in waterlogged or reducing conditions, where iron is present in its reduced form. Siderite is less common in well-drained soils but can be important in hydromorphic soils. |
| Magnesite <i>MgCO₃</i> | Found in soils rich in magnesium, typically in regions with ultramafic rocks. Magnesite often occurs in smaller quantities compared to calcite and dolomite. |
| Aragonite <i>CaCO₃</i> | A polymorph of calcite, aragonite is less stable and typically forms under different temperature and pressure conditions than calcite. It may occur in soils influenced by marine deposits or shells. |
| Hydromagnesite <i>Mg₅(CO₃)₄(OH)₂ · 4H₂O</i> | A hydrated form of magnesium carbonate, it can form in highly alkaline environments, especially in soils with high magnesium and low calcium contents. |

