Detection of Organic Analyte on Soils Using IR Reflectance Spectra

Paul L. Gassman1
1Eigenvector Research, Inc., Manson, WA, USA
2Pacific Northwest National Laboratory, Richland, WA, USA

Research Objective
Detect the presence of organic analyte in previously analyze-free soil using multiplicative scatter correction spectroscopy. Results are from a laboratory study conducted to benchmark and guide future studies for active standoff detection of analytes of interest. Technology for environmental monitoring and remediation, and precision agriculture applications.

This study examined different signal processing and detection statistics for detection of the presence of strong C-H bands in the 2855 to 2966 cm⁻¹ region. This region was selected because C-H signal is strong and atmospheric interference signal from carbon dioxide and water vapor is low. Posterior detection of C-H initiates search for the presence of additional absorbance bands that may indicate the type of analyte present.

Related Work


Soils Studied
Quincy Soil (Washington, USA)
85% silt, mean diameter 152 µm, total organic carbon 0.03%

League Soil (Texas, USA)
40% non-siltstone, mean diameter 9.9 µm, total organic carbon 0.5%

Needs to be continued for the results of Leech, Kieth, et al. (2000), and Gallagher, Blake, Slavery, et al. (2006). Note: The data for Quincy Soil was used for further testing. The data for League Soil was used for initial testing.

Model Calibration and Test Procedures
Calibrate the detection model using preprocessed spectra from analyte-free soil: Use spectra from soils spiked with a single compound per soil and purged with dry-N₂. The calibration data are used to i) calculate any parameters used for signal processing and ii) to determine any parameters used for the detection model.

Test A)
Calibrate on 40 samples collected on one day. Test the same samples after coating with tributyl phosphate, dodecane, or decanal. Solutions were evaporated without dry-N₂ purge. Soils were then coated drop-wise with a solution of analyte in 2-methyl butane. The solvent was allowed to evaporate prior to measuring analyte.

Test B)
Calibrate on 200 samples collected on three separate days. Test on different soils: Both analyte-free and coated with analyte. Test all data used in Test A.

Signal Processing Studied
All preprocessing was applied to spectra measured in the 5000 to 2000 cm⁻¹ region and was followed by centering to the calibration set mean.

(a) None: No Preprocessing
(b) SG2: Savitzky-Golay Second Derivative
(c) EMSC: Extended Multiplicative Scatter Correction
(d) EMSC+SG2: EMSC followed by SG2
(e) EMSCwt: Extended Multiplicative Scatter Correction with de-correlation of weighted residuals
(f) SG2-EMSC: SG2 followed by EMSC

Extended Multiplicative Scatter Correction
EMSC can be considered a signal filter where the signal is modeled as containing target signal (known and unknown), clutter, and multiplicative (e.g. path length) effects. The reference (calibration set mean) measured spectrum can be used to filter out measured spectra to filter out target signal from noise (H₂O and CO₂ spectra)

EMSC with De-correlation: EMSCwt
EMSC with de-correlation is a more flexible, but can be time consuming and require selectivity for clutter and target in each window. De-correlation of the clutter covariance matrix is an alternative that allows more flexibility in the clutter basis.

For C, the first step is determining V to be mean-center the calibration data X, and calculate the covariance matrix C = (X - µ)T (X - µ) and calculate the eigenvectors of C. For EMSCwt, a symmetric, banded diagonal weighting matrix was used with weighted residuals for the diagonal of the covariance matrix. In EMSCwt, the eigenvectors of C were used for scatter-correlated clutter.

Detection Statistics Studied
Six different statistics were used for detection where subsets W_i = I, D corresponding to matrices B, D, and C:

- Sensitivity = Number of true positives detected
- Specificity = Number of true negatives not detected (Fraction false alarm)
- False alarm rate = Number of false positives detected
- Gain = Ratio of detected to observed signals
- AUC: Area under the receiver operating characteristic curve
- Chi squared (see statistical appendix for definition)

Test A Results
Shows perfect performance for all preprocessing except ‘None’ with Q₃ performance best (no expected). EMSCwt has largest distance from the 99.5% line for test samples. However, the results are likely over-optimistic because test data consist only for the same soils used for calibration. ‘Performance is about 10% to 20%’ for monitoring a single soil over time.

Test B Results
Results are likely more typical of performance because different soils are used to calibrate and test the models. In this case, Q₃ tends to false alarm (likely due to overfitting) although this could be damped Q₃ appeared to perform best as not much better than Q₃ SG2 is simple and performed the best at 99% CL, while EMSCSG2, EMSCSG2, and EMSCwt performed best. EMSCwt was better than EMSC suggesting that additional flexibility was required for EMSC.

Conclusions and Future Work
Second derivatives were simple to apply and worked well for both detection and removal of analyte. Both EMSCwt and Q₃ show promise methodologies for increasing sensitivity, however EMSC-based approaches require good characterizations of the clutter. Strategies will be applied to new field data.

Statistical Appendix
For W, V is a simple scale factor and it is a 3X1 vector of ones.

W = [F W] / ||W||
V = e = (||W||^2)^{-1/2}

Detection Statistics Studied
Six different statistics were used for detection where subsets W_i = I, D corresponding to matrices B, D, and C:

- Sensitivity = Number of true positives detected
- Specificity = Number of true negatives not detected (Fraction false alarm)
- False alarm rate = Number of false positives detected
- Gain = Ratio of detected to observed signals
- AUC: Area under the receiver operating characteristic curve
- Chi squared (see statistical appendix for definition)

Limitations of EMSCwt and Q₃ were determined using T-distributions, and for Q₃ using the method by Jackson and Mudholkar, Technometrics, 25(3), 1979.