Application of Extended Inverse Scatter Correction to Mid-Infrared Reflectance Spectra of Soil

NIR 2005, April 10-15, 2005, Auckland, New Zealand

Neal B. Gallagher
nealg@eigenvector.com
Eigenvector Research, Inc., Manson, WA USA
Thomas A. Blake, Paul L. Gassman
Pacific Northwest National Laboratories
Richland, WA USA

EIGENVECTOR RESEARCH INCORPORATED

Multiplicative Scatter Correction

- Preprocessing methods that attempt to reduce effects of scatter in reflectance spectroscopy
- Result is less signal related to scattering artifacts and more signal related to analyte(s) of interest
- Multiplicative scatter correction (MSC)
  - based on classical least squares (CLS) model
  - Geladi P, MacDougall D, Martens H., *Appl. Spectrosc.* 1985; **39**(3): 491-500.

EIGENVECTOR RESEARCH INCORPORATED

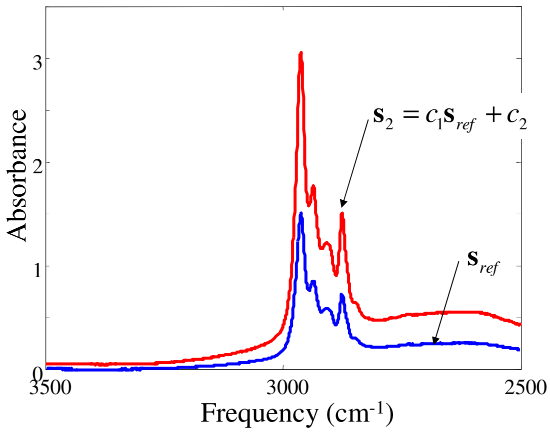
Multiplicative Scatter Correction

MSC and ISC model scatter with an offset and slope

$$\mathbf{s}_2^T = \mathbf{c}\begin{bmatrix}\mathbf{s}_{ref} & \mathbf{1}\end{bmatrix}^T \quad \mathbf{c}_{1x2} \quad \text{MSC}$$

$$\begin{bmatrix}\mathbf{s}_2 & \mathbf{1}\end{bmatrix}\mathbf{b} = \mathbf{s}_{ref} \quad \mathbf{b}_{2x1} \quad \text{ISC}$$

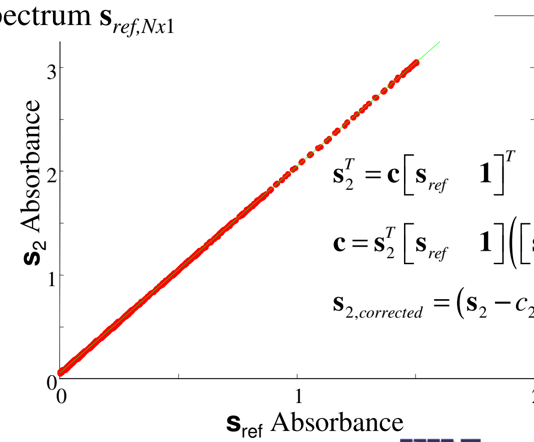
classical least squares (CLS)
inverse least squares (ILS)



EIGENVECTOR RESEARCH INCORPORATED

MSC

Regress new measured spectrum $\mathbf{s}_{2,Nx1}$ onto reference spectrum $\mathbf{s}_{ref,Nx1}$



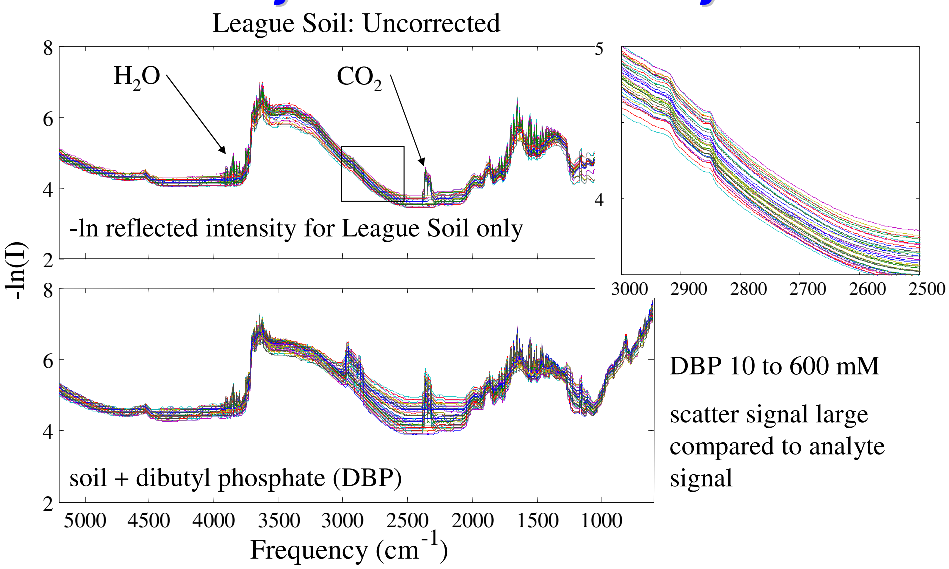
$$\mathbf{s}_2^T = \mathbf{c}\begin{bmatrix}\mathbf{s}_{ref} & \mathbf{1}\end{bmatrix}^T$$

$$\mathbf{c} = \mathbf{s}_2^T\begin{bmatrix}\mathbf{s}_{ref} & \mathbf{1}\end{bmatrix}\left(\begin{bmatrix}\mathbf{s}_{ref} & \mathbf{1}\end{bmatrix}^T\begin{bmatrix}\mathbf{s}_{ref} & \mathbf{1}\end{bmatrix}\right)^{-1}$$

$$\mathbf{s}_{2,corrected} = \left(\mathbf{s}_2 - c_2\mathbf{1}\right)/c_1$$

EIGENVECTOR RESEARCH INCORPORATED

Why is MSC Necessary?



Objective

- Remove
  - scattering artifacts
  - atmosphere analytes (H_2O and CO_2)
- Retain
  - reference soil signal
  - analyte signal [dibutyl phosphate (DBP) – model organophosphorous]
- However, scattering and atms are more complicated than simple offset and gain
  - extended least squares model

EIGENVECTOR RESEARCH INCORPORATED

Extended MSC & EISC

EISC attempts to correct for scatter that manifests in forms other than just linear.

$$\begin{bmatrix}\mathbf{s}_2 & \boldsymbol{\upsilon}^2 & \boldsymbol{\upsilon} & \mathbf{1}\end{bmatrix}\mathbf{b} = \mathbf{s}_{ref} \qquad \mathbf{Z}_{Nx4} = \begin{bmatrix}\mathbf{s}_2 & \boldsymbol{\upsilon}^2 & \boldsymbol{\upsilon} & \mathbf{1}\end{bmatrix}$$

$$\mathbf{b} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{s}_{ref}$$

$$\mathbf{s}_{2,corrected} = \mathbf{Zb}$$

b is 4x1

EIGENVECTOR RESEARCH INCORPORATED

Extended MSC & EISC

can add spectra of known target analyte $\mathbf{S}_{NxJ}$
polynomial can be of order K-1
can add spectra of known interference $\mathbf{Q}_{NxL}$.

$$\begin{bmatrix}\mathbf{s}_2 & \mathbf{S} & \mathbf{P} & \mathbf{Q}\end{bmatrix}\mathbf{b} = \mathbf{s}_{ref} \qquad \mathbf{P}_{NxK} = \begin{bmatrix}\cdots & \boldsymbol{\upsilon}^2 & \boldsymbol{\upsilon} & \mathbf{1}\end{bmatrix}$$

$$\mathbf{b} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{s}_{ref} \qquad \mathbf{Z}_{Nx(1+J+K+L)} = \begin{bmatrix}\mathbf{s}_2 & \mathbf{S}_A & \mathbf{P} & \mathbf{Q}\end{bmatrix}$$

$$\mathbf{s}_{2,corrected} = \begin{bmatrix}\mathbf{s}_2 & \mathbf{P} & \mathbf{Q}\end{bmatrix}\begin{bmatrix}b_R & \mathbf{b}_P^T & \mathbf{b}_Q^T\end{bmatrix}^T \qquad \mathbf{b} = \begin{bmatrix}b_R & \mathbf{b}_S & \mathbf{b}_P & \mathbf{b}_Q\end{bmatrix}$$

b is (1+J+K+L)x1

EIGENVECTOR RESEARCH INCORPORATED

How to get Q?

- **Q** is a sub-space that spans scatter
  - measure multiple reflectance spectra of soil samples that do not contain analyte → $\mathbf{X}_Q$
  - define reference spectra as mean of $\mathbf{X}_Q$ → $\mathbf{s}_{ref}$
  - center $\mathbf{X}_Q$ to $\mathbf{s}_{ref}$ → $\mathbf{X}_{Qm}$
  - perform PCA on centered data → $\mathbf{X}_{Qm}=\mathbf{TQ}^T + \mathbf{E}$
  - use the loadings $\mathbf{Q}_{NxL}$ to characterize scatter

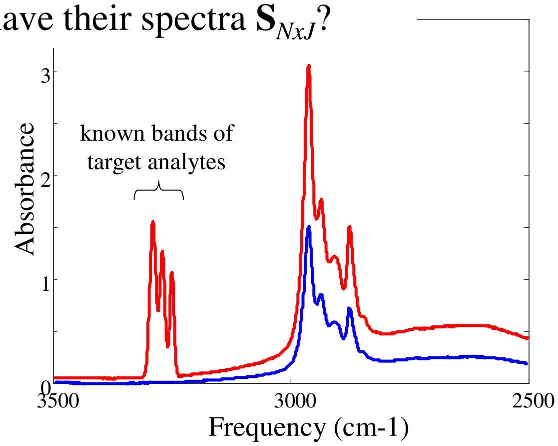
EIGENVECTOR RESEARCH INCORPORATED

Extended MSC

What if we know channels that should have target analytes but we don't have their spectra $\mathbf{S}_{NxJ}$?

If we do nothing the target spectra will bias the regression and provide poor correction.

Weighted least squares can be used to de-weight these channels in the regression step.



EIGENVECTOR RESEARCH INCORPORATED

Weighted EISC

- use a diagonal weighting matrix $\mathbf{W}_{NxN}$ to de-weight channels where analyte is present
  - weights are 0 where analyte is present

$$[\mathbf{s}_2 \quad \mathbf{S} \quad \mathbf{P} \quad \mathbf{Q}]\mathbf{b} = \mathbf{s}_{ref}$$

$$\mathbf{b} = \left(\mathbf{Z}^T\mathbf{W}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{W}\mathbf{s}_{ref}$$

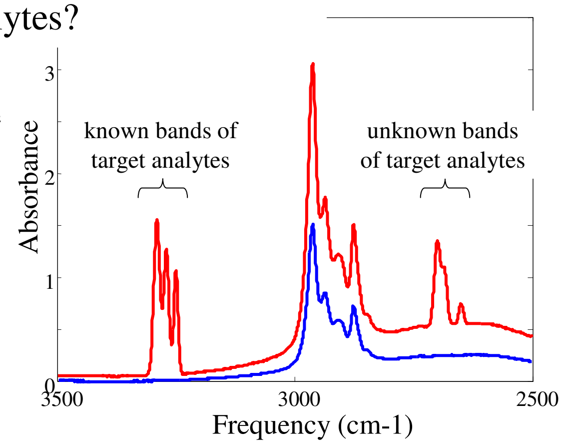
EIGENVECTOR RESEARCH INCORPORATED

Robust EISC

What if we don't know all the channels that should have target analytes?

If we do nothing the target spectra will bias the regression and provide poor correction.

Robust least squares (treats these points like "outliers") can be used to de-weight these channels in the regression step.



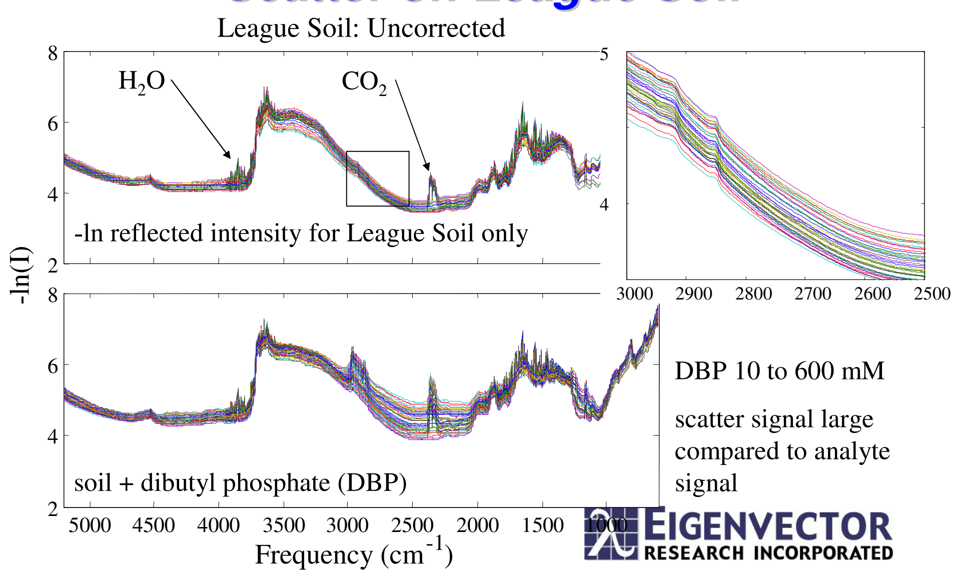
EIGENVECTOR RESEARCH INCORPORATED

Soil/Analyte Samples

- League Soil (44% clay, 42% silt, 14% sand)
- Quincy Soil (7% clay, 17% silt, 76% sand)
- Analyte: Dibutyl phosphate in 2-Methyl Butane
  - 0, 10-600 mM dripped onto soil sample
  - 2 MB highly volatile, evaporates quickly
  - measure spectra w/ and w/o dry-N_2 purge
  - sample (DBP concentration) randomized

EIGENVECTOR RESEARCH INCORPORATED

Scatter on League Soil

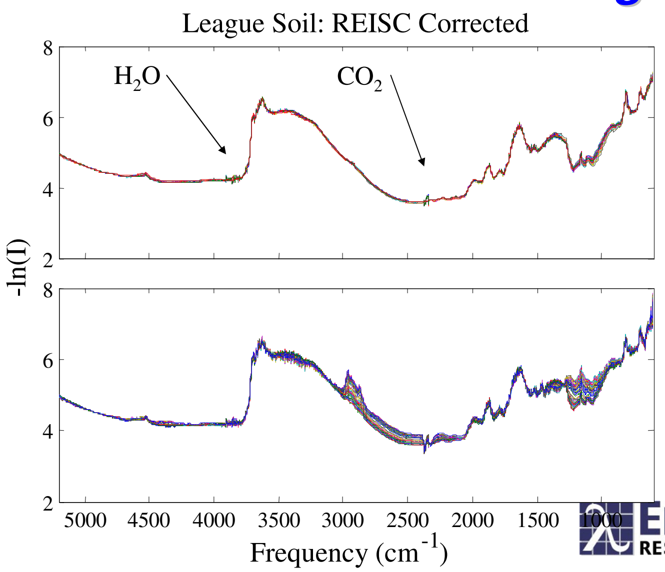


DBP 10 to 600 mM

scatter signal large compared to analyte signal

EIGENVECTOR RESEARCH INCORPORATED

Robust EISC for League Soil

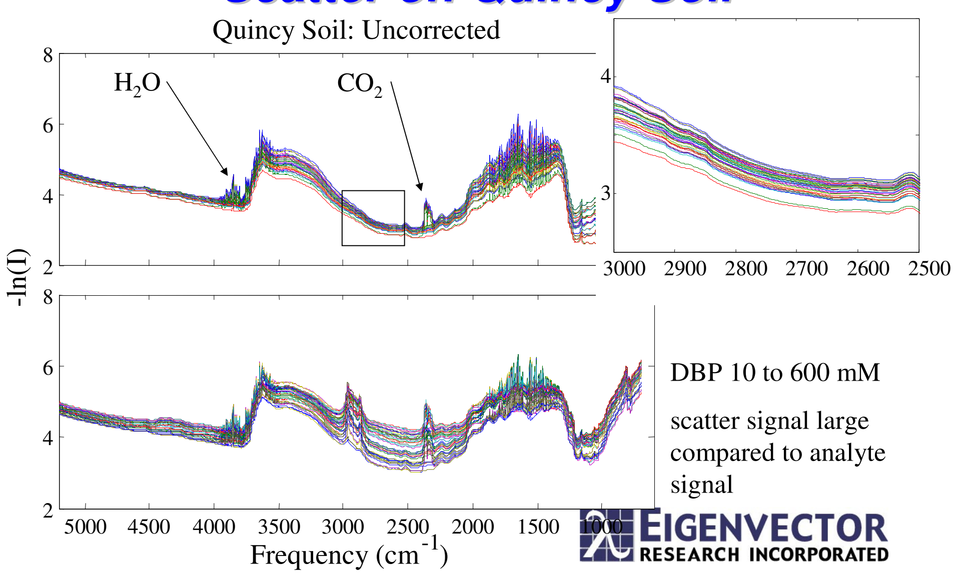


DBP 10 to 600 mM

scatter signal much smaller compared to analyte signal

EIGENVECTOR RESEARCH INCORPORATED

Scatter on Quincy Soil

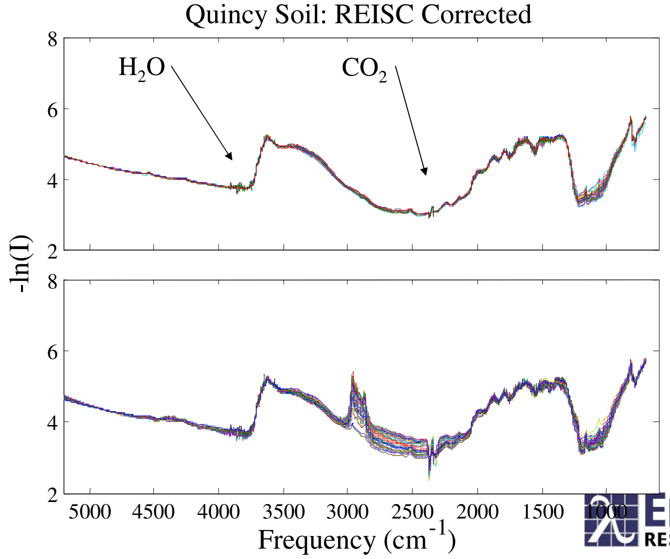


DBP 10 to 600 mM

scatter signal large compared to analyte signal

EIGENVECTOR RESEARCH INCORPORATED

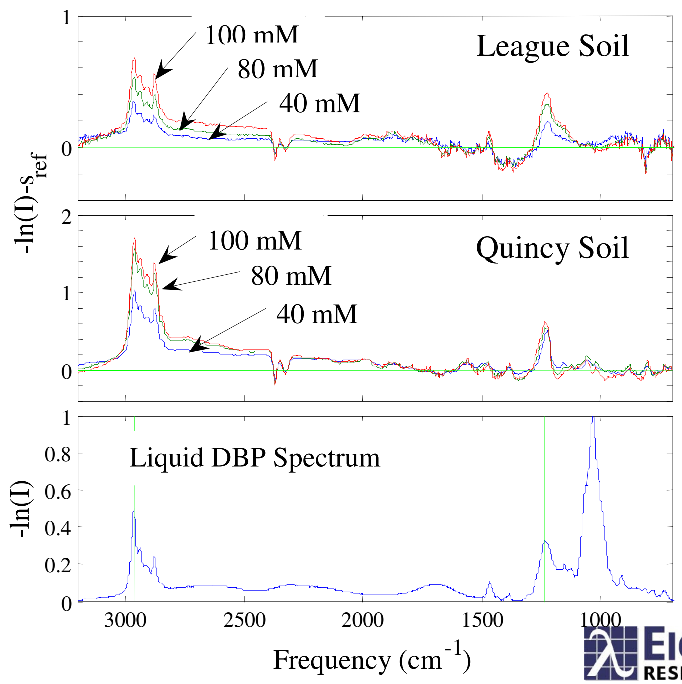
Robust EISC for Quincy Soil



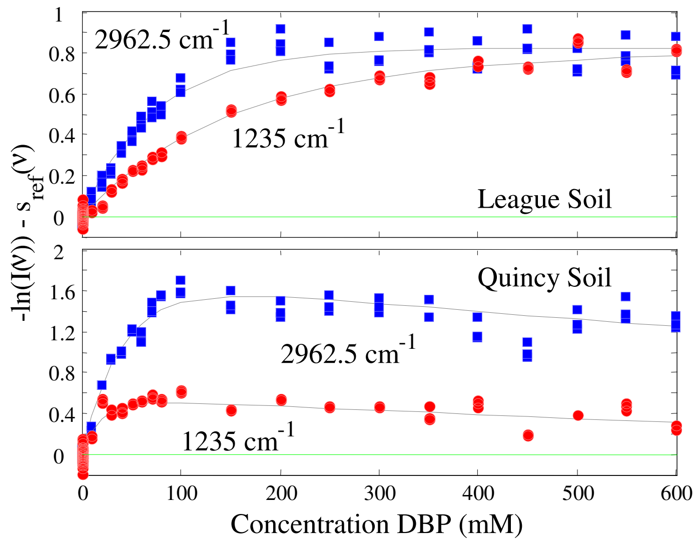
DBP 10 to 600 mM

scatter signal much smaller compared to analyte signal

Corrected Spectra - S_{ref}



Corrected Spectra vs DBP Conc.



Regression Analysis

- Examine effect of scatter correction using PLS regression for quantification of DBP
  - Calibration set: 0, 10, 30, 50, 70, and 100 mM
  - Test set: 0, 20, 40, 60, and 80 mM
  - ~avoids non-linearity due to observed “saturation”

EIGENVECTOR RESEARCH INCORPORATED

Regression Results League

Regression results for DBP in League Soil.

		% Variance in First PLS Factor		Regression Error (mM) RMSE		
Correction	Num Factors	Spectra	Conc	Calib	CV	Pred
Uncorrected	5	97.72	72.60	2.11	3.66	5.52
EISC	3	89.28	75.83	2.65	3.05	5.55
WEISC	4	91.81	99.02	1.34	2.68	3.88
REISC	4	91.66	98.63	1.17	2.58	3.24
2nd Deriviative	2	91.45	49.64	4.08	4.17	5.05

Regression Results Quincy

Regression results for DBP in Quincy Soil.

		% Variance in First PLS Factor		Regression Error (mM) RMSE		
Correction	Num Factors	Spectra	Conc	Calib	CV	Pred
Uncorrected	3	92.70	90.62	4.80	6.88	7.48
EISC	4	85.32	93.93	4.58	7.66	12.34
WEISC	1	97.35	97.46	7.50	7.47	5.50
REISC	3	93.12	97.49	4.36	6.06	5.37
2nd Deriviative	3	83.08	65.10	6.50	9.25	5.08

Conclusions

- The extended mixture model can be used to account for complicated scatter in reflectance
  - Best to have many analyte-free soil measurements to characterize scatter
- Weighted and Robust regression are useful for estimating the model coefficients
- EISC shows promise
  - spectra interpretable, relevant/predictive variance brought to top (scatter artifacts removed)

EIGENVECTOR RESEARCH INCORPORATED