RDECOM
TECHNOLOGY DRIVEN. WARFIGHTER FOCUSED.

EDGEWOOD CHEMICAL BIOLOGICAL CENTER
1917 100th Anniversary 2017
A U.S. Army RDECOM Laboratory

Edgewood Chemical Biological Center

Maximum Discrimination Approach for Classification of Nearly Identical Signatures

6 December, 2017

Darren K. Emge
Electrical Engineer

Dr. Steven Kay
University of Rhode Island

“Fieldable” Analytical Instruments

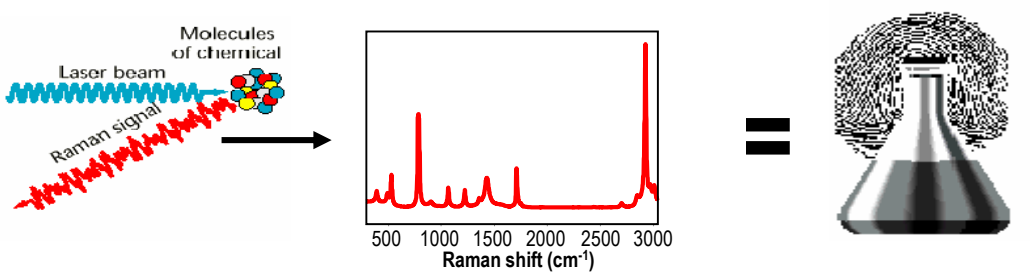
Sensitivity - Specificity

Mobile Laboratory

Handheld

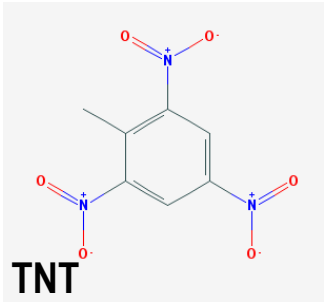
Raman Spectroscopy

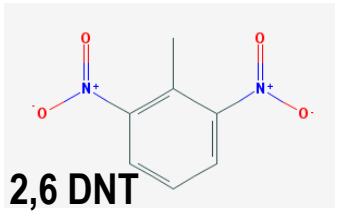
Raman = molecular light scattering phenomenon
Molecular size (nms) << wavelength of light (200 - 1064 nm)

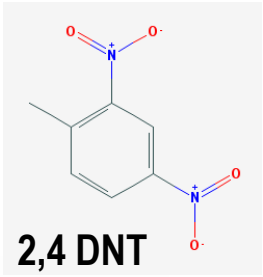


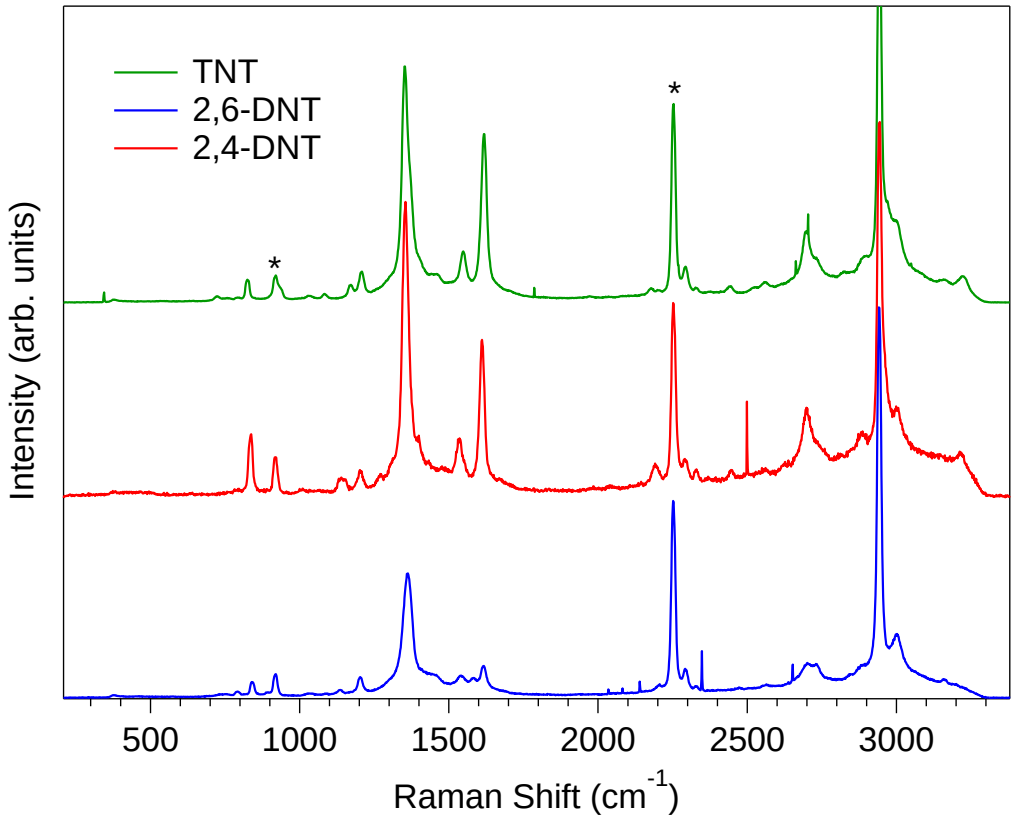
- Vibrational spectroscopy (high selectivity and spectral fingerprinting)
- Water (liquid or vapor) does not obscure the spectral fingerprint
- Future applications are driven by development of miniature UV laser sources (< 250 nm) and more efficient NIR detectors (> 1024 nm)

Specificity – Nearly Identical Signatures







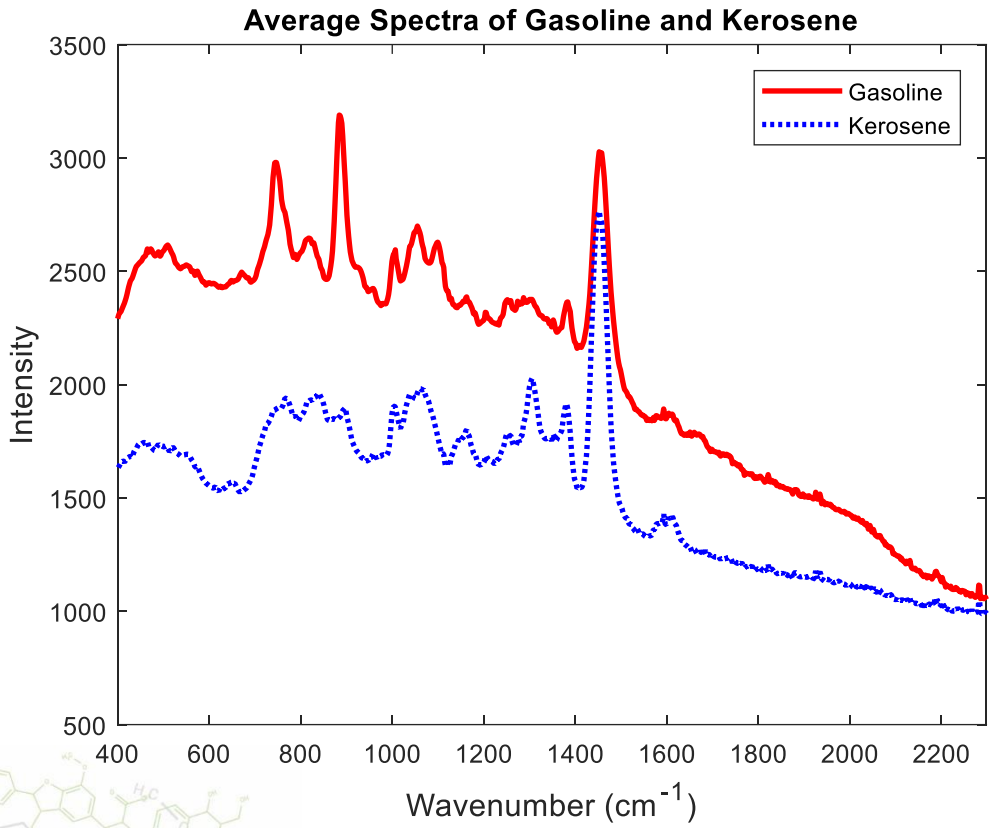


There is a need to differentiate highly similar compounds for forensic attribution and medical treatment

* Source of Structures Pubchem (pubchem.ncbi.nlm.nih.gov)

Real World Example

In reality most samples are complex mixture of an analyte of interest and adulterants



Mixtures of Hydrocarbon Chains

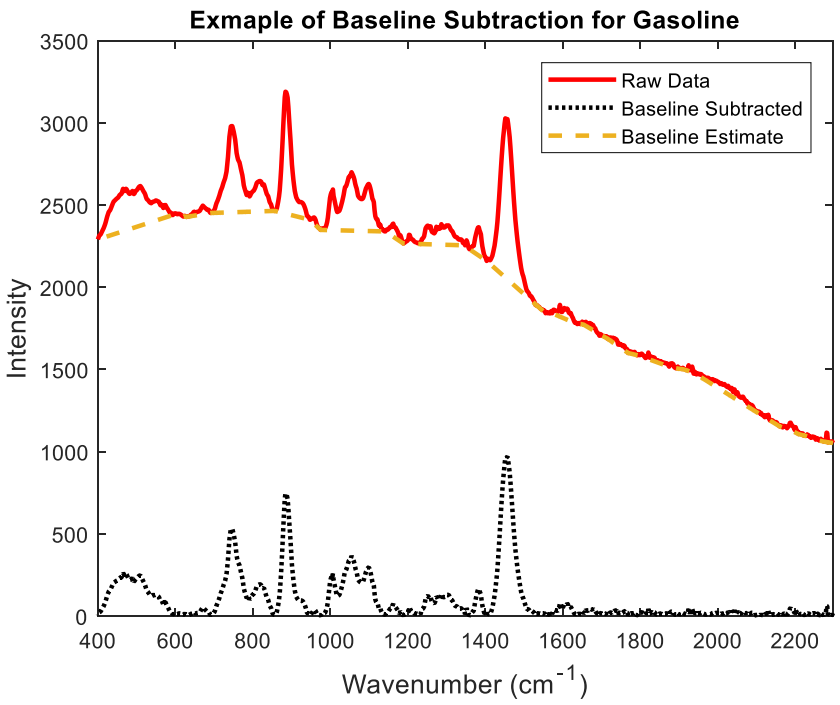
Gasoline C^7 – C^{11}

Kerosene C^{12} - C^{15}

Fire Hazard

Property	Kerosene	Gasoline
Flashpoint	>38°C	-43°C
Auto Ignition	220°C	280°C
Vapor Pressure	0.7 kPa	6.9 kPa

Baselining and Normalization



Baseline due to florescence

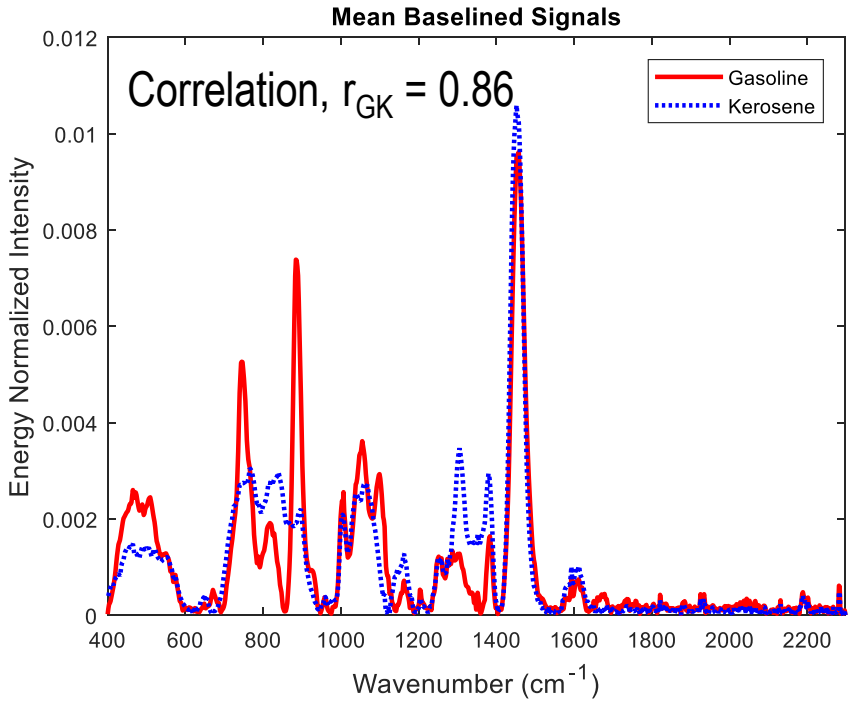
- Complicates analysis

Baseline Removal

- Iterative Median Filter *
  - Filter length 300 samples
  - 10 Iterations

* Guicheteau, et al. “Bacillus Spore Classification via Surface Enhanced Raman Spectroscopy and Principal Component Analysis”. 2008

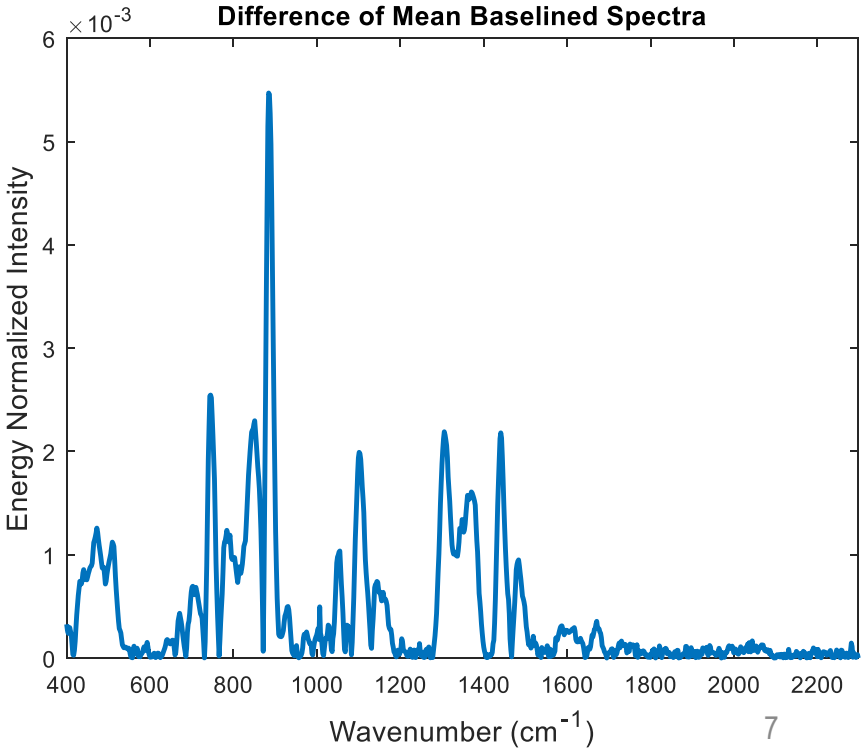
Difference of Spectral Signals



High correlation leads to mis-identification

→ Need to find regions that provide the most information !

Simple magnitude difference shows many areas of possible significant difference but which are the most important ?



Maximum Discrimination Approach

- Itakura-Saito Spectral Distance Measure
  - Approximation of distance between the PSD of two Gaussian Distributions

$$d(P_1, P_2) = \frac{1}{2}\int_0^B \left[\frac{P_1(f)}{P_2(f)} - \ln\left(\frac{P_1(f)}{P_2(f)}\right) - 1\right] df$$

$$+\frac{1}{2}\int_0^B \left[\frac{P_2(f)}{P_1(f)} - \ln\left(\frac{P_2(f)}{P_1(f)}\right) - 1\right] df$$

where B is the highest frequency

Shape of the Metric

Let $g(x) = \frac{P_1}{P_2}$

we can rewrite MDA equation as

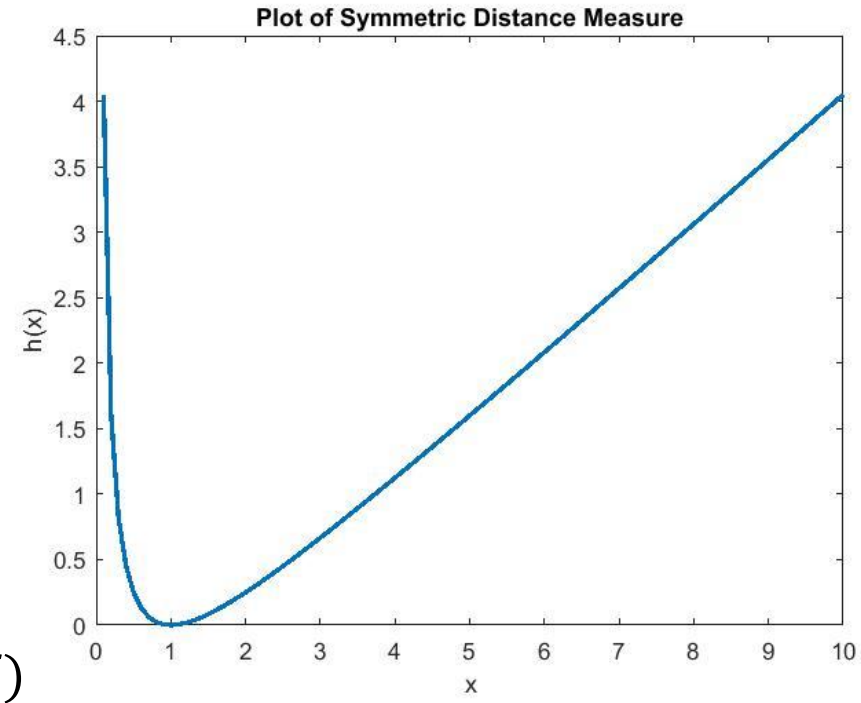
$$\mathrm{h}(x) = \frac{1}{2}\left(g(x) + g\left(\frac{1}{x}\right)\right)$$

since

$$g(x) = x - \ln(x) - 1 \geq 0, \qquad x > 0$$

Then $h(x)$ is minimum at $x = 1$

That is, $d(P_1, P_2) = 0$ when $P_1(f) = P_2(f)$

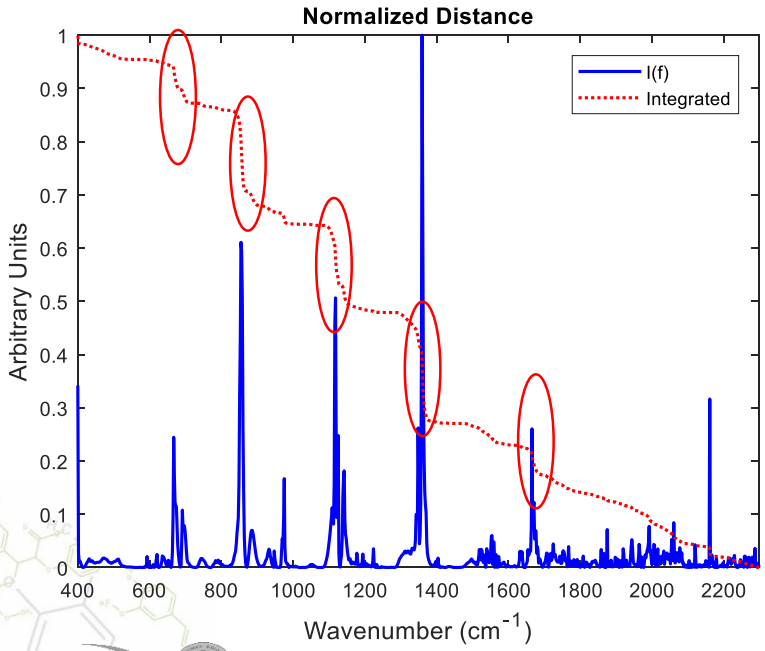


Application of MDA to Data

Applying the distance measure to each spectral band

$$\mathrm{I}(f) = \frac{1}{2}\int_0^B \left[\frac{P_1(f')}{P_2(f')} - \ln\left(\frac{P_1(f')}{P_2(f')}\right) - 1\right] df'$$

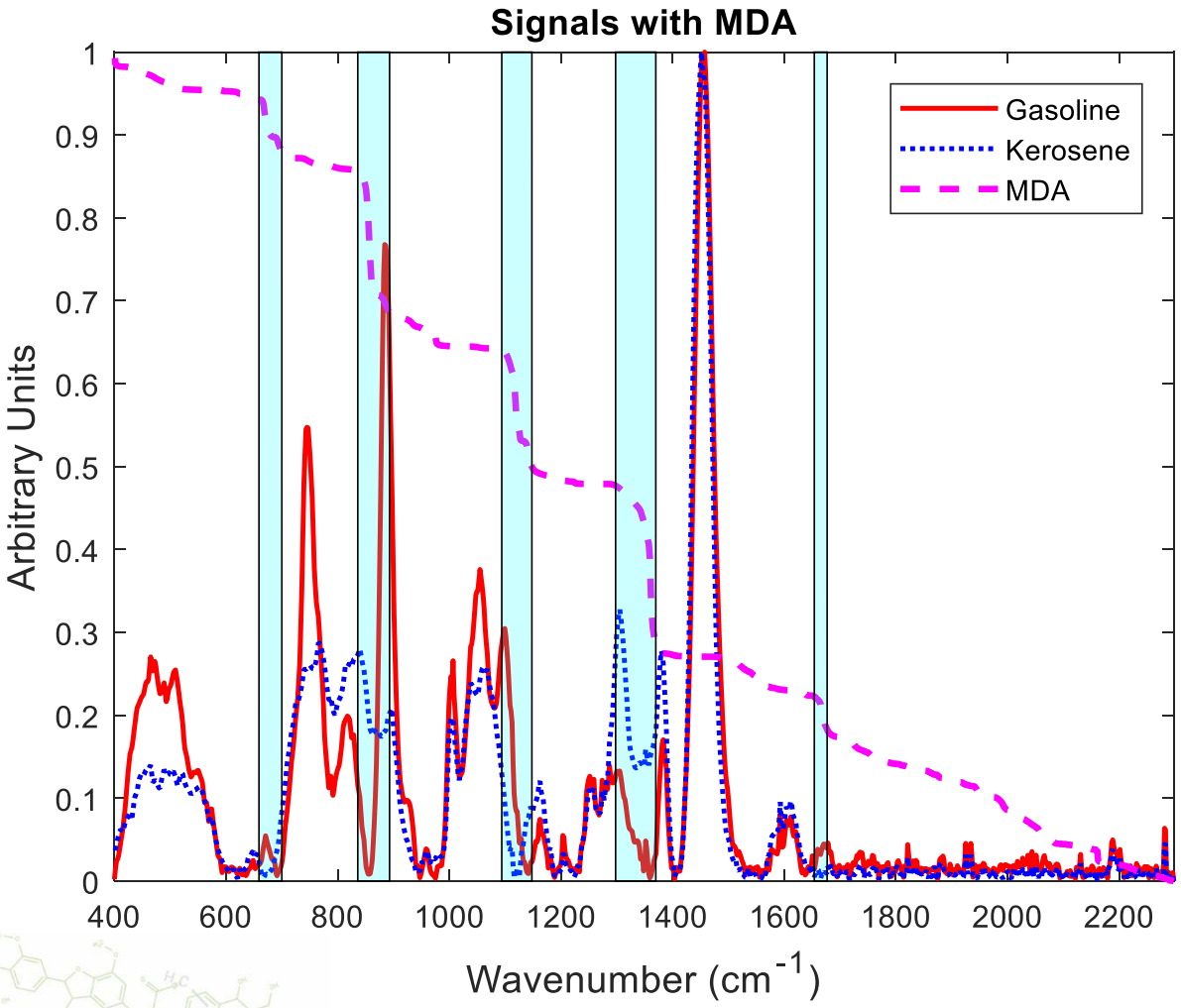
$$+\frac{1}{2}\int_0^B \left[\frac{P_2(f')}{P_1(f')} - \ln\left(\frac{P_2(f')}{P_1(f')}\right) - 1\right] df'$$



Maxima "information" occurs when $I(f)$ is maximum

However, we are looking for regions of maximum information so if we integrate $I(f)$ and look at regions of maximum change, i.e., maximum slope

MDA Results



Maximum slope occurs at 659 – 700, 836 – 893, 1094 – 1148, 1298 – 1370, and 1654 – 1677 cm^{-1}

Examine the correlation in and out of those bands:

$r_{in\ band} = 0.24$

$r_{out\ of\ band} = 0.93$

Current Effort

- Examination of highly similar compounds for attribution
- Autonomous determination of regions of maximum discrimination
  - Spline based approach
  - Statistical decision methods
- End to end system demonstration

Acknowledgements

- Dr. Jason Guicheteau – Spectroscopist

  Date Collection and Interpretation

  “That does not look like a spectra”

- Dr. Steven Kay – Electrical Engineer

  Methods and Theoretical Development