# Investigation of Functional Data Analysis Techniques for Chemical Spectra

Joanne R. Wendelberger

Statistical Sciences Group, MS F600
Los Alamos National Laboratory
Los Alamos, NM  87545
joanne@lanl.gov

Joint Research Conference on
Statistics in Quality, Industry, and Technology
Knoxville, Tennessee
June 7-9, 2006

# Abstract

In many applications, the entity of interest is a function, rather than a single numerical value. Applications involving chemical spectra are introduced to illustrate statistical issues that arise in the analysis of functional data that are generated by chemical analysis instruments. Functional data analysis tools may be used to visualize and analyze the chemical data. Statistical issues such as quantification of variability, data visualization, dimension reduction, and data comparison are addressed.

# Introduction

- Look at 3 examples of chemical spectral data.
- Identify statistical issues in each application.
- Examine functional data analysis tools to address statistical issues.

# Functional Data

- In some experiments, the entity that is of interest may be a function rather than a single value.
- The underlying object may be a curve, a spectrum, a distribution, or some other functional response.
- The measured data may come as a curve or as a sample of measured values arising from the function of interest.

# Spectral Data

- Numerous instruments are available that generate spectral measurements of physical phenomena.
- A spectrum is often stored in digital form with values recorded at a large number of frequencies.
- The high dimensionality of spectral data requires appropriate techniques to visualize and understand data.

# LIBS Soil-Screening Example

- **L**aser **I**nduced **B**reakdown **S**pectroscopy
- Using a laser spark, sample material is vaporized, reduced to atomic constituents, and electronically excited to produce a spectral output.

# Multivariate Calibration Methods

1. Classical Least Squares
2. Inverse Regression
3. Partial Least Squares
4. Principal Components Regression

Martens and Naes (1989)
Haaland and Thomas (1988)

# Practical Considerations

- Baseline Corrections
- Frequency Range Selection
- Computational Requirements
- Data Reduction/Compression Techniques

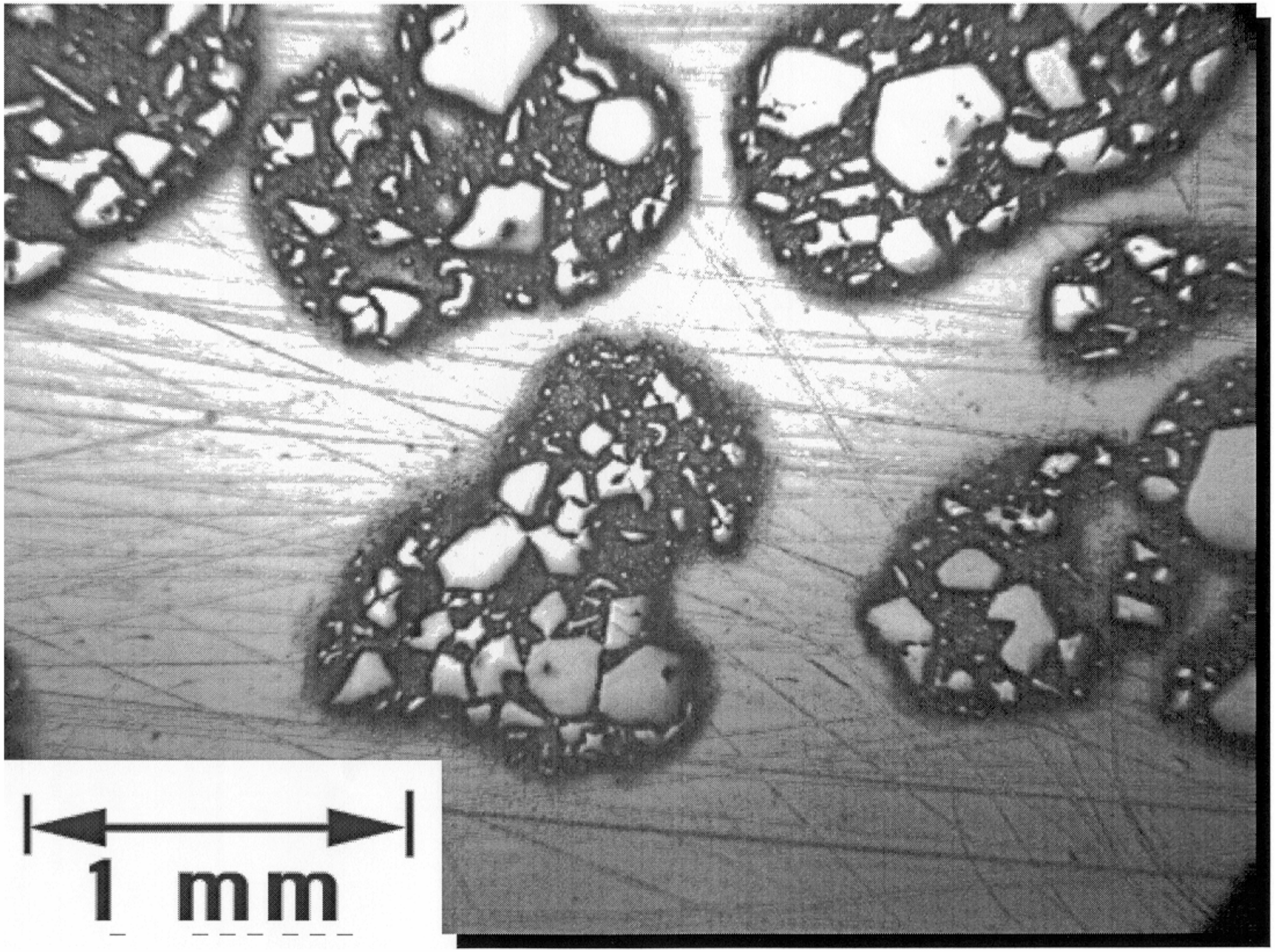
# Infrared Spectra of Aged and Unaged Heat-treated Materials

A polymer material was aged under varying aging conditions. Samples of unaged (control) materials and three different aged materials were heated and then allowed to cool down to ambient temperatures. Infrared spectra of the samples that were collected over time are being studied to try to understand chemical changes in the materials.

# Background

- Estane 5703 ® is a polymer material used in the binder formulation of a high explosive.
- With heating, the polymer "melts" and becomes a viscous fluid.
- With cooling, the polymer undergoes a self-organizing, phase separation resulting in solidification.

# PBX 9501 Molding Powder



# Polymer Data

- The data consist of time-dependent infrared (IR) spectra of samples of Estane 5703® after they are heated and cooled.
- The polymer used in the samples has been artificially aged by various means.
- Measurements were taken over a week timeframe.

Data was collected by Darla Thompson, Los Alamos National Laboratory.

# Tools for Spectral Aging Data

- Data visualization methods
- Multivariate analysis techniques including clustering and multivariate calibration
- Functional data analysis techniques including principal component analysis and other decomposition technqiues

Joint work with D. Scott, B. Ray, E. Kober, S. Marron

# Data Visualization

- The mean of the raw data is subtracted out to obtain better separation of the curves.
- Zooming is used to focus in on interesting sets of frequencies, showing smooth curves.
- Colorization is used to display time ordering.
- Principal components plots capture major behavior in a few directions.

# Principal Component Analysis

- Each spectrum may be thought of as a point in a high dimensional Euclidean space.
- Principal component analysis, obtained from an eigenanalysis of the covariance matrix, provides an orthogonal decomposition of the functional data.
- The first principal component corresponds to the direction of maximum variation.
- Examination of the first few principal components may reveal structure in the functional data.

# Estane Results

- Methods were developed for examining time-dependent spectra in a manner that requires minimal input (and possible bias) from the analyst.
- Color is used to enhance visualization of the spectra over time.
- Simple analysis on individual peaks indicated that one process dominated the spectral changes.
- One minor process was also observed.

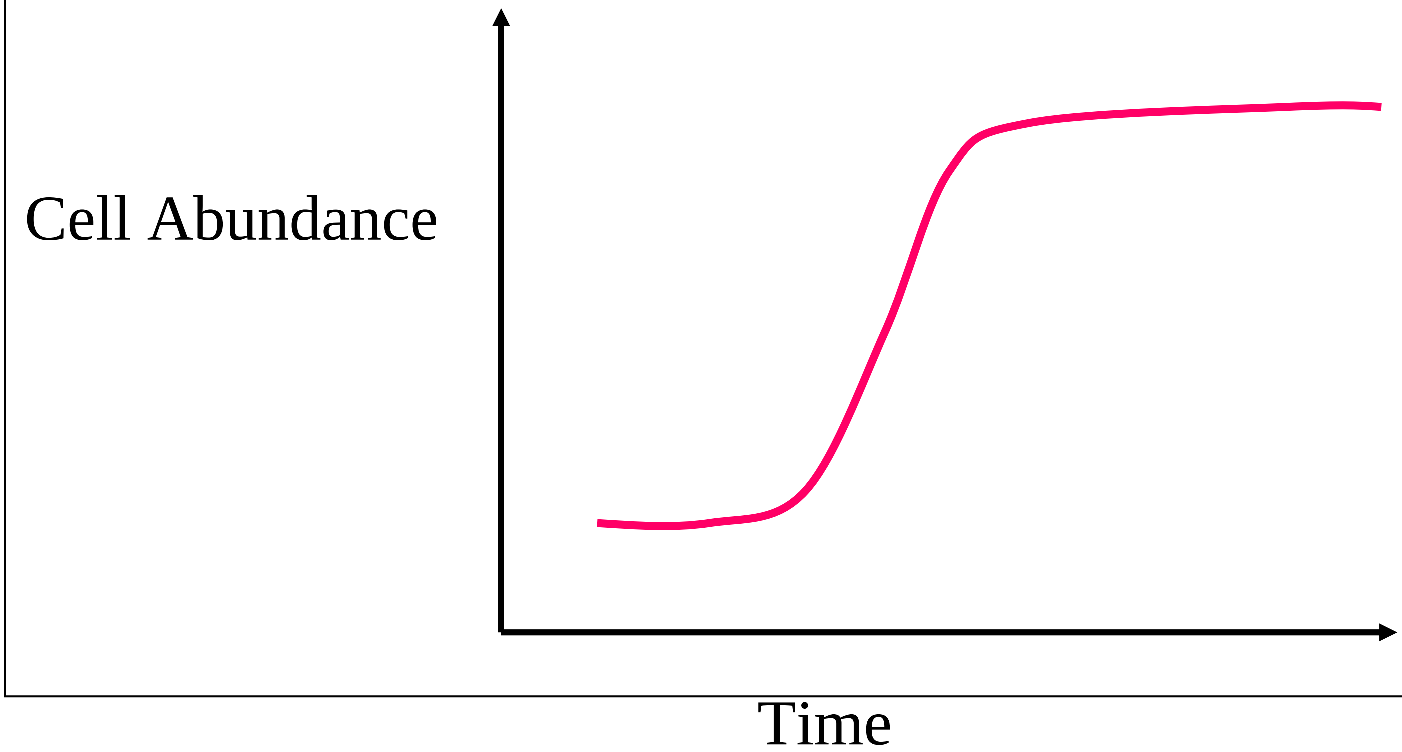
# Functional Variation Study

- A new project is underway to investigate the use of NMR/Mass Spectrometry to study metabolites in biological systems. As a first step, a baseline measurement study is needed.
- Need to quantify variability of spectral measurements as well as biological variability between samples.
- Baseline study will measure spectra for known compounds.
- Eventually want to compare measured unknown spectra to library catalogs of known spectra.

Jt. Work with L. Ticknor, P. Unkefer,LANL

# Orange Data

- Metabalomics experiments are being conducted that measure the abundance of cells as a function of time.



# Summary

- Chemical spectra provide measurements of underlying functions.
- Statistical methods can be used to visualize and analyze the spectral data.
- Statistical methods can be used to perform data summary, data visualization, quantification of variability, dimension reduction, and data comparison.

# Future Research

- Continue development of tools for visualization of chemical spectra.
- Examine alternative methods for decomposition of spectra.
- Explore methods for designing data collection strategies for chemical spectra.