Identifying Molecules in Mass Spectra Using Principal Component Analysis (PCA)

Bradley Powers, Bernice Mills

Sandia National Laboratories, Livermore, CA 94550

Abstract

Many materials outgas as they are heated in a vacuum, compromising a sealed vacuum if they cannot be gettered. Our goal is to be able to identify which molecules are emitted and at what temperature. Determining specific species from the fragments produced in the mass spectrometer is extremely difficult since there is always more than one species outgassed at any time. We have chosen to use PCA because it has been used previously with mass spectral data. The commercial PCA software we used was unable to separate the spectra into pure, or almost pure, species. Other strategies for interpreting the data are suggested.

Introduction

- Materials heated in a vacuum.
- Mass spectra of the vacuum are recorded periodically resulting in a **very large** matrix of masses versus time.
- PCA software factors data into two much smaller matrices.
- Analyze the data after factorization.

What does PCA do?

$$\begin{array}{cc} & D \\ & \text{Masses} \rightarrow \\ \text{Time} \downarrow & \begin{pmatrix} d_{11} & \cdots & \cdots & \cdots & \cdots & d_{1n} \\ \vdots & \ddots & & & & \vdots \\ \vdots & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ d_{m1} & \cdots & \cdots & \cdots & \cdots & d_{mn} \end{pmatrix} \end{array}$$

Very Large

m time intervals

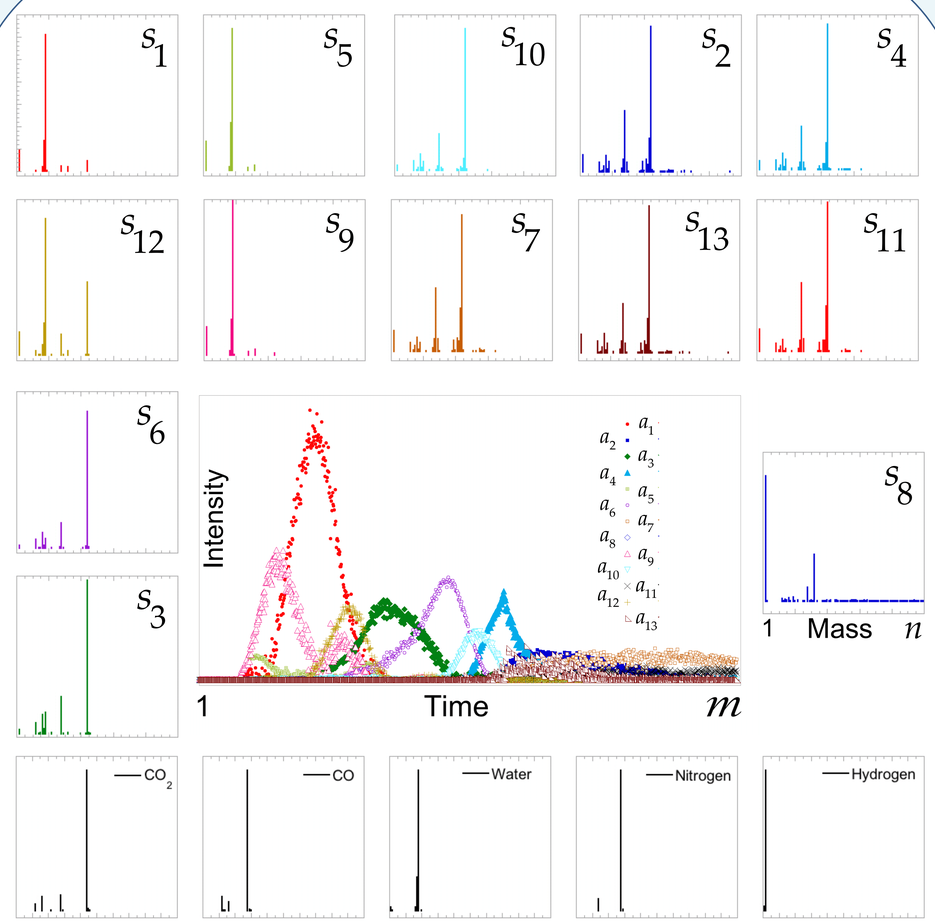
n masses

PCA finds j components

$j \ll m, n$

$$\begin{array}{cc} & A \\ \text{Time} \downarrow & \begin{pmatrix} a_{11} & \cdots & a_{1j} \\ \vdots & & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj} \end{pmatrix} \end{array} \bullet \begin{array}{c} S \\ \text{Masses} \rightarrow \\ \begin{pmatrix} s_{11} & \cdots & \cdots & \cdots & \cdots & s_{1n} \\ \vdots & & \ddots & & & \vdots \\ s_{j1} & \cdots & \cdots & \cdots & \cdots & s_{jn} \end{pmatrix} \end{array}$$

Example of Principal Components Found Outgassed From a Material, (j=13)



This material is based upon work supported by the S.D. Bechtel, Jr. Foundation and by the National Science Foundation under Grant No. 0952013 and Grant No. 0934967. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the S.D. Bechtel, Jr. Foundation or the National Science Foundation."

Goals

- To break up data into more easily identifiable parts so as to…
- Determine which species are being emitted inside the vacuum with some ease and ultimately…
- Create an easy to use procedure to determine individual species from large amounts of mass spectral data

Results

- Several component spectra are similar, e.g. s_1, s_5, s_{12}, and s_9
- s_{12} seems to be a combination of s_1 and s_6
- None of the components represent a pure species

Conclusions

- Spectra did not separate clearly because there was never one species present at any time
- We need to use a different approach if we cannot collect data for individual species

What Now?

- Try using correlation matrices
- Find a way to eliminate known species

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

SAND2011-5386 P