

Contributed Talk, for presentation at the 9th International Conference on X-ray Absorption Fine Structure (XAFS), to be held in Grenoble, France on 8/25-30/96.

CONF-9608104--3

ANL/CHM/CP--89747

**THE ANALYSIS OF MIXTURES: APPLICATION OF PRINCIPAL COMPONENT ANALYSIS TO XAS SPECTRA**

Stephen R. Wasserman\*, Chemistry Division, Argonne National Laboratory, 9700 S. Cass Ave., Argonne, IL, USA

708-252-3527

708-252-9288 (FAX)

SRW@ANL.GOV

\* Author to whom correspondence should be addressed

RECEIVED

SEP 03 1996

OSTI

The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. W-31-109-ENG-38. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED *ph*

MASTER

**DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Many samples which are subjected to XAS analysis contain the element of interest in more than one chemical form. The interpretation of the spectra from such samples is often not straightforward, particularly if appropriate model systems are not available. We have applied principal component analysis (PCA) to real and simulated systems which contain mixtures of several species for a given element. PCA has been extensively used for the analysis of other types of spectra, including MS, IR, and UV-VIS. The application of PCA to XAS is illustrated by examining the speciation of iron within coals. PCA can determine how many different species that contain a particular element are present in a series of spectra. In tandem with model compounds, principal component analysis can suggest which of the models may contribute to the observed XAS spectra.

## 1. INTRODUCTION

X-ray absorption spectroscopy has proved extremely useful for characterizing samples in which the absorbing atom is found within only a single species. However, in many cases, the absorbing atom is located in several chemical environments. If the identities of the individual species in such samples are known, then analysis of the XAS spectra can provide information on the relative amounts of each compound within the material. Unfortunately, particularly in environmental applications, the constituent species within the sample are not known *a priori*. It is nevertheless possible to analyze a series of such spectra and obtain information on the number of distinct species that are present. In this paper we describe the application of principal component analysis (PCA) to X-ray absorption spectroscopy. The iron near edge (XANES) spectra from a series of coals serves as an illustrative example.

## 2. THEORY OF PRINCIPAL COMPONENT ANALYSIS (PCA)

The history of principal component analysis extends over 60 years for the behavioral sciences.[1] It has been used for the analysis of chemical spectra for more than 30 years. In this section we summarize the technique following the terminology of Malinowski and Howery.[2]

The starting point for PCA is the recognition that every spectrum can be represented as a vector. In the following discussion we will use the term spectrum to refer to the experimental data and vector to denote the mathematical representation of that spectrum. For XANES spectra, each part of the vector indicates the intensity of absorption at a particular energy. If the energy scales for the original spectra differ, interpolation can be used to place the spectra on a common scale. We construct a data matrix, [D], in which the columns contain the individual vectors. The covariance matrix for the data, [Z], is then calculated by multiplying the transpose of the data matrix, [D]<sup>T</sup>, and the data matrix together (eq. 1). Each member of [Z] is

$$[Z] = [D]^T [D] \quad (1) \quad [R] = [D][C]^T \quad (2) \quad [D]_{\text{repro}} = [R][C] \quad (3)$$

the inner product between two vectors and represents how similar the individual spectra are to one another. The eigenvectors of [Z] are abstract components for the spectra. In themselves, these components do not have any direct physical meaning. The corresponding eigenvalues,  $\lambda$ , represent the relative importance of each component.

The eigenvectors are then used to reproduce the original data matrix. A column matrix, [C], is formed, containing 1 or more of the most significant eigenvectors, as determined from the eigenvalues, in its rows. From the column matrix, a row matrix, [R], is created by eq. 2. Lastly, the reproduced data, [D]<sub>repro</sub>, which is the product from multiplication of the row and column matrices, is compared to the original data, [D]. If these data matrices agree within experimental error, then these components are sufficient to explain the data. If this procedure does not reproduce the data well, the process is repeated after adding the next most significant eigenvector to the column matrix, until the data are reproduced. The *n* eigenvectors necessary to reconstruct the data, the primary eigenvectors, represent the presence of *n* distinct species in the original spectra. The remaining, secondary, eigenvectors contain the experimental error and spectral noise. This procedure is equivalent to the methods in linear algebra for determining the dimensionality of the vector space for the spectra.

One advantage of PCA is that it provides a method for testing whether a particular pure material is present in the original mixtures. The power of this procedure, called target transformation, is that one can test a particular spectrum without requiring any knowledge regarding the other species which may be present in the samples. In target transformation, a test vector, Test,

$$\text{Test}' = [R][\lambda]^{-1}[R]^T \text{Test} \quad (4)$$

representing the spectrum of interest, is transformed into Test', according to eq. (4).  $[\lambda]^{-1}$  is a square matrix that contains the reciprocals of the primary eigenvalues as its diagonal elements. If Test and Test' agree within experimental error, the spectrum represented by Test corresponds to a species present in the samples. One complication is that each of the original vectors will transform successfully, even though some or all of them may represent the spectra of mixtures, rather than of pure species.

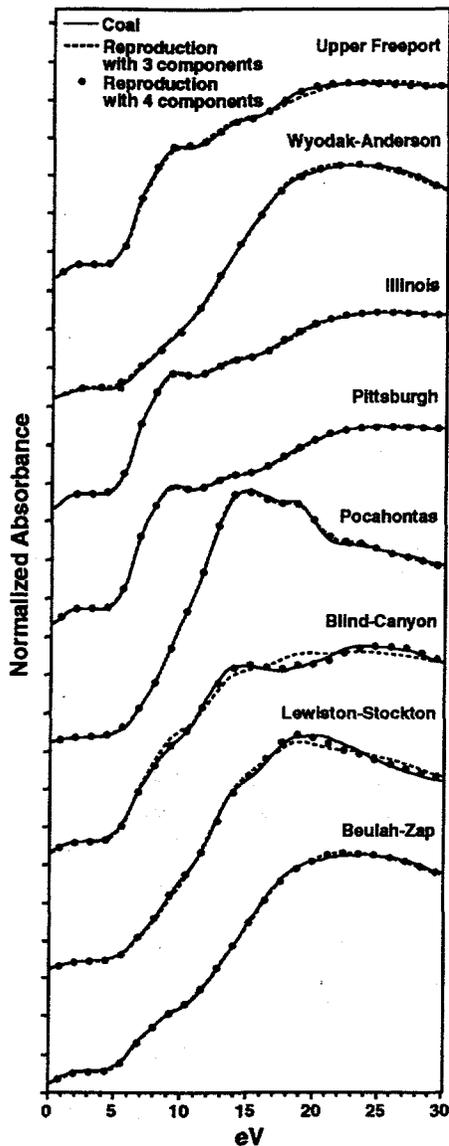


Figure 1: Iron K edge X-ray Absorption Near Edge Spectra (XANES) from the Argonne Premium Coal Samples (—) and data reproductions using 3 (---) and 4 (●) principal components.

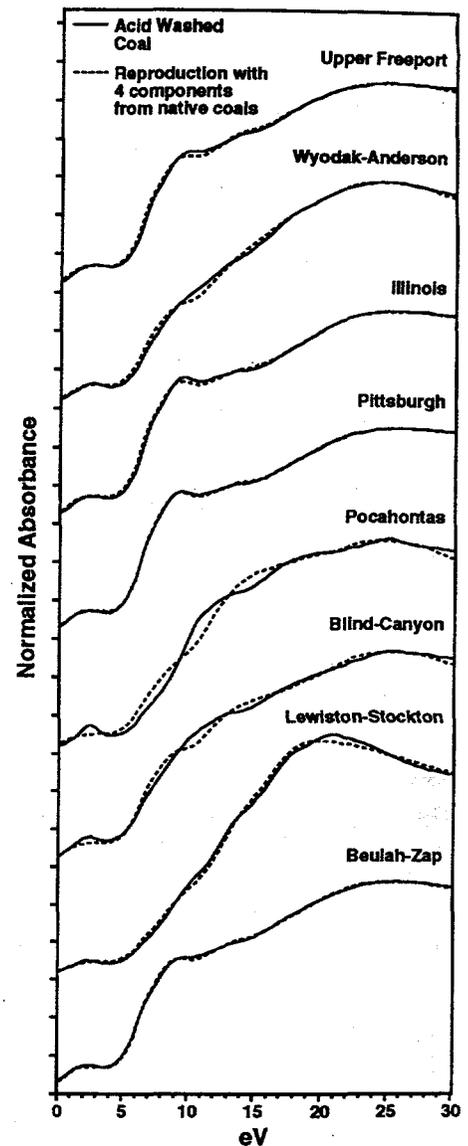


Figure 2: Iron K edge XANES spectra from the Argonne Premium Coal Samples after extraction with hydrochloric acid (—) and data reproductions using the 4 principal components from the native coals (---)

### 3. APPLICATIONS

Figure 1 shows the XANES spectra for iron in the eight coals of the Argonne Premium Coal Samples (APCS): Upper Freeport, Wyodak - Anderson, Illinois #6, Pittsburgh #4, Pocahontas, Blind Canyon, Lewiston-Stockton, and Beulah-Zap.[3] Figure 1 also contains the results of fitting these spectra to the first three and first four principal components. The former set of components is fairly successful in reproducing seven of the eight spectra, although for two spectra, that from Wyodak-Anderson and Beulah-Zap, the quality of the reconstruction is marginal. The 3 components, however, fail to reproduce the spectrum from the Blind Canyon coal. Inclusion of the fourth component results in a successful modeling of each of the spectra. This result indicates that there are (at least) four different species contributing to this series of spectra. In deriving these components, the XANES spectra were normalized to the edge jump,  $\mu_0$ . Other normalizations are possible, including the standard normalization of linear algebra, in which the absolute value of each vector is 1. Changing the way the spectra are normalized will not usually affect the number of primary components. However, altering the normalization will affect the individual eigenvectors and their relative importance. For the spectra analyzed here, normalization to  $\mu_0$  results in relative weightings for the spectra which differ by up to 15 percent from the equal weightings that occur with the use of vectors that are normalized to an absolute value of 1.

Additional insights into the constituents of these coals can be gained from the use of PCA in conjunction with experiment. Each of the coals was extracted with hydrochloric acid in order to remove some of the iron from the organic core of the coals. While many iron compounds are susceptible to extraction, pyrite,  $\text{FeS}_2$ , a material found in many coals, is resistant to such treatment. Figure 2 contains the spectra for the eight coals after acidic processing. Also included are the reproductions of these spectra using the four primary components from the native coals. Several of the reproductions are poor, in particular those for Wyodak-Anderson, Pocahontas, and Blind-Canyon. There are two possible explanations for the fact that the components from the fresh coals do not successfully model the acid washed samples. The first is that, during the extraction, new iron compounds were created. Alternatively, minor species that were originally in the coals now represent a greater percentage of the iron content because of removal of other iron compounds. We find that when the principal components are derived for the acid washed coals, four components are necessary to reconstruct the spectra. When all sixteen spectra from the fresh and acid-treated coals are analyzed in this manner, there are six principal components. These latter two findings are consistent with the fact that the components for the fresh coals do not work well with the acid-treated samples. All of these conclusions have been reached without making any assumptions as to which iron species exist in the coals.

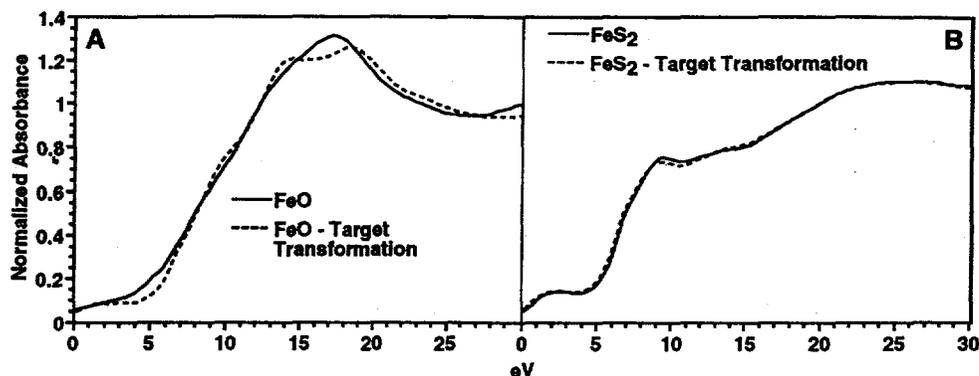


Figure 3: Target transformations for the iron K edge XANES spectra of (A) ferrous oxide,  $\text{FeO}$ , and (B) pyrite,  $\text{FeS}_2$ .

We have used target transformation to test whether certain compounds are present in these coals. The result of two such tests are shown in Figure 3. The transformed spectrum of ferrous oxide,  $\text{FeO}$ , does not agree well with the original one. This result indicates that  $\text{FeO}$  is not present in the Argonne Premium Coals. Nevertheless, the reconstruction does crudely mirror the position and shape of the  $\text{FeO}$  XANES spectrum. This observation suggests that some sort of ferrous,  $\text{Fe(II)}$ , species is a constituent of the samples. A successful transformation, that for pyrite, also appears in Figure 3. We therefore conclude that pyrite is found in the APCS.

#### 4. LIMITATIONS

An implicit assumption in principal component analysis is that the observed spectrum for a mixture is a linear combination of the spectra from the separate constituents. This assumption is usually valid for XAS spectra, although there may be some instances in which it is not true. In addition, anything which changes the shape and position of the XANES spectrum will affect the PCA analysis. PCA will be most successful for a series of spectra taken at a single beamline under identical experimental conditions. Errors in calibration and changes in monochromator crystals and/or mode of detection will all result in apparent increased experimental error. Thickness effects represent another possible complicating factor for PCA analysis.

#### 5. CONCLUSIONS

Principal component analysis is a useful method for examining the XAS spectra from samples that contain mixtures of compounds. This paper has presented the basic considerations that go into a PCA analysis. Other enhancements to the technique examine the effects of experimental error. Additionally, special types of target transformations can be used to determine which spectra reflect the presence of the same species. For samples about which there is little initial information regarding the compounds contained within, PCA can provide a significant first step in gaining insight into the nature of the samples.

#### Acknowledgments

This research was supported by the Division of Chemical Sciences, Office of Basic Energy Sciences, U. S. Department of Energy under contract W-31-109-ENG-38.

#### References

- [1] Aries R. E., Lidiard D. P., and Spragg R. A., *Chemistry in Britain* (September, 1991) 821-824.
- [2] Malinowski E. R. and Howery D. G., *Factor Analysis in Chemistry* (John Wiley and Sons, New York, 1980).
- [3] Wasserman S. R., Winans R. E., and McBeth R., *Energy & Fuels* 10 (1996) 392-400.