SANDIA REPORT
SAND2001-3638
Unlimited Release
Printed January 2002

Reducing System Artifacts in Hyperspectral Image Data Analysis with the Use of Estimates of the Error Covariance in the Data

David Michael Haaland, Mark H Van Benthem, Christine M. Wehlburg,
Frederick W. Koehler IV

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

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

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/ordering.htm

SAND2001-3638
Unlimited Release
Printed January 2002

Reducing System Artifacts in Hyperspectral Image Data Analysis with the Use of Estimates of the Error Covariance in the Data

David M. Haaland, Mark H. Van Benthem, and Christine M. Wehlburg
Information Detection, Extraction and Analysis
Sandia National Laboratories
Albuquerque, New Mexico 87185-0886

Frederick W. Koehler IV
Spectral Dimensions, Inc.
Olney, MD 20832

Abstract

Hyperspectral Fourier transform infrared images have been obtained from a neoprene sample aged in air at elevated temperatures. The massive amount of spectra available from this heterogeneous sample provides the opportunity to perform quantitative analysis of the spectral data without the need for calibration standards. Multivariate curve resolution (MCR) methods with non-negativity constraints applied to the iterative alternating least squares analysis of the spectral data has been shown to achieve the goal of quantitative image analysis without the use of standards. However, the pure-component spectra and the relative concentration maps were heavily contaminated by the presence of system artifacts in the spectral data. We have demonstrated that the detrimental effects of these artifacts can be minimized by adding an estimate of the error covariance structure of the spectral image data to the MCR algorithm. The estimate is added by augmenting the concentration and pure-component spectra matrices with scores and eigenvectors obtained from the mean-centered repeat image differences of the sample. The implementation of augmentation is accomplished by employing efficient equality constraints on the MCR analysis. Augmentation with the scores from the repeat images is found to primarily improve the pure-component spectral estimates while augmentation with the corresponding eigenvectors primarily improves the concentration maps. Augmentation with both scores and eigenvectors yielded the best result by generating less noisy pure-component spectral estimates and relative concentration maps that were largely free from a striping artifact that is present due to system errors in the FT-IR images. The MCR methods presented are general and can also be applied productively to non-image spectral data.

Contents:

Figures

INTRODUCTION

The recent availability of commercial hyperspectral imaging Fourier transform infrared (FT-IR) spectrometers[1, 2] has increased the need for improved methods of analyses of the image data. Although most applications of this new technology have involved qualitative analysis of the image data, obtaining quantitative information from the hyperspectral image data is highly desirable. Any developments in the quantitative analysis of the FT-IR image data would also have application to the analysis of multispectral and hyperspectral image data obtained from remote sensors such as satellite or airborne hyperspectral imaging sensors. One opportunity available from the massive data obtained in a single hyperspectral image is the possibility of performing quantitative analysis of the image data without the use of calibration standards. This opportunity has recently been realized in the analysis of Raman hyperspectral images using multivariate curve resolution algorithms that employ constrained alternating least squares iterative methods to generate estimates of the pure-component spectra of species present in the image data and to obtain relative concentration maps of those chemical species.[3, 4]

We have found that the quantitative analysis of hyperspectral image data collected from FT-IR imaging systems that employ focal plane array (FPA) detectors can be greatly hampered by the presence of system artifacts in the spectral and image data. Methods are required to eliminate the detrimental effects of these systematic artifacts if they cannot be eliminated experimentally. We have found that the characteristics of the systematic artifacts can be captured in the error covariance structure of the data estimated from repeat spectral image differences. By coupling new generalized augmented classical least squares (ACLS) methods [5] with improved multivariate curve resolution techniques[6-11] and estimates of the error covariance structure of the data, we have been able to greatly minimize the detrimental effects of the systematic artifacts present in hyperspectral imaging FT-IR systems. Thus, improved pure-component spectral estimates and relative concentration maps from the image spectra have been realized without the requirement for calibration standards. These new methods will be described and demonstrated using an example of an FT-IR spectral image obtained from a sample of neoprene thermally aged in air at elevated temperatures.

EXPERIMENTAL

The sample investigated involved a sample of neoprene rubber. In order to obtain adequate signal to noise in the FT-IR image data, the neoprene sample did not contain the normal carbon filler found in many neoprene rubbers. The sample that was used to illustrate the methods described in this paper was obtained from a 2 mm x 4 mm x 4 mm (?) block of neoprene that was aged for 6 days in an air furnace held at a temperature of 140 °C. The sample was then potted in epoxy and the center of the sample was cut with a microtome to an approximate thickness of 15 μm. However, independent estimates of the sample thickness as a function of position on the sample indicated that the thickness varied by a factor of greater than three across the sample.

The spectrometer used in this study was a BioRad Stingray step-scan FT-IR imaging spectrometer equipped with a mercury-cadmium-telluride (MCT) FPA detector that had 4096 (64 x 64) detector elements. The sample was placed in the macro sampling compartment that provided a 4 mm x 4 mm transmission image to be obtained of the thin 2 mm x 4 mm aged neoprene sample. The sample was pressed flat onto a KBr window for support. Background single-beam images were obtained from a clear region of the KBr window next to the neoprene sample. Sample single-beam transmission image spectra were then obtained of the neoprene sample. The spectra were collected at a nominal spectral resolution of 16 cm^{-1} using a step rate of 2.5 steps/sec and co-adding 64 images at each step of the interferometer. Data collection time was approximately 5 minutes for each spectral single-beam image. Three image spectra were obtained in rapid succession of the KBr window background before moving the sample into the beam. Three image spectra were then obtained from the neoprene sample without moving the sample. Two of these repeat image spectra were used to obtain an estimate of the error covariance structure of the data.

Small numbers of spectra from each image were clear outliers and were removed from the analysis. These outliers generally exhibited spikes or steps in the raw interferograms. The sample images were trimmed to remove all pixels representing just the KBr window and mixed pixels at the extreme boundaries of the sample that contained sample and KBr window. Only pixels present in all single-beam images were retained for final analysis. The sample spectra were ratioed to backgrounds and converted to absorbance. Large, relatively complex baseline variations are present in these spectra. The presence of these baseline variations complicated attempts to extract pure-component spectra. Therefore, the spectra were preprocessed by isolating the spectral features of the sample and performing a separate linear baseline correction under each spectral band for each spectrum in the image.

Since the sample thickness was not uniform over the sample area, pathlength corrections to the data were required to obtain meaningful concentration maps of the various components across the sample. In order to perform pathlength corrections without direct measurements of the sample thickness, we identified a spectral band that was most representative of sample pathlength. This band was selected among the six baseline-corrected bands by performing a principal components analysis (PCA) [12] on each spectral band and examining the results carefully. As explained in another report, [13] necessary conditions for a spectral band to be representative of sample pathlength are that the band be present in all pixels of the image, that it has a consistent spectral shape in all pixels, and that only multiplicative changes are present in the band. Spectral bands that fulfill these conditions of only a multiplicative change will have the first PCA eigenvector represent nearly all the variance, and the first eigenvector will have the same shape as the average spectrum of the selected band. Of the six bands identified in the neoprene sample, the 1440 cm^{-1} band most nearly followed these criteria and was, therefore, selected as representing sample pathlength. Relative pathlength for each pixel was determined by assigning a relative pathlength of one to the average spectrum of the 1440 cm^{-1} band and using the average spectrum in a classical least squares (CLS) [14] prediction of all sample image spectra. The CLS analysis also included a simultaneous

fit of linear baseline components to the band. Relative pathlengths varied from 0.13 to 2.5. The thinnest pathlengths are at the edges of the sample. Because the spectra contained artifact from the system, the spectra were not scaled for pathlength since this process could serve to increase the variance of the artifact in the image. Rather, the MCR was performed on the uncorrected spectra, and the estimated concentrations were corrected for relative pathlength variations.

The spectral data were collected with the use of the Bio-Rad Win-IR Pro software that came with the Stingray spectrometer. However, spectral analyses were performed using Matlab 6.1 code primarily written at Sandia National Laboratories. Matlab code for the SIMPLISMA algorithm was taken from Windig et al.[15]

THEORY

Since the primary goal of our analysis is to obtain the pure-component spectra and concentration maps of each component in the image, we must have a method of extracting this information directly from the spectral data since standards are not available for these samples. Andrew et al.[3, 4] have presented methods to accomplish the goal of quantitative analysis without standards using Raman hyperspectral image data. They employed an MCR algorithm that used constrained alternating least squares methods in an iterative analysis. The constraints applied involved non-negativity of the concentrations and the pure-component spectra since real concentrations and spectra will consist of all positive values. These constraints serve to limit the range of possible solutions in the analysis. In our analysis, we use not only non-negativity constraints but also equality constraints as outlined below to further limit the possible solutions and to minimize the detrimental effects of the system artifacts on the analyses.

The MCR methods require a starting point for the pure spectra and/or their corresponding concentrations. Although random numbers can be used as starting points, we chose to use the pure component spectra obtained from the SIMPISMA algorithm[15] as starting points. These pure spectra are generated by finding and extracting the purest spectral bands in the spectral data sets one at a time. From these purest pixels, often reasonable initial estimates of the pure-component spectra can be obtained.

Initially MCR was performed starting with the SIMPLISMA estimated pure-component spectra using only non-negativity constraints. The constrained alternating least squares algorithm used was a modification of the fast non-negative least squares algorithm from Bro.[6] Changes to Bro's algorithm were made to speed the convergence of the MCR analysis. Unfortunately, it became obvious that the spectral data were heavily contaminated by systematic errors that severely degraded the quality of the both the pure-component spectra and the concentration maps generated from the MCR analysis. The source of the systematic errors is thought to be related to systematic readout problems from the FPA detector. Independent of their source, the result is the presence of large correlated errors in the spectral domain and a banding artifact in the concentration maps. All the standard implementations of the MCR algorithms assume that the spectral errors are random, of constant variance and independent (i.e., uncorrelated). Unfortunately, all

these assumptions are violated in our FT-IR image data. In order to correct for the error structure in our MCR analysis, we could employ generalized least squares method that weights the analysis with an estimate of the error covariance structure in the data. However, the generalized least squares algorithms tend to be very slow and are not currently practical for the iterative MCR algorithms that are required for these large data sets. Therefore, we take a different approach that is currently practical. Our generalized augmented least squares (ACLS) methods create the opportunity to correct the MCR analysis for the presence of non-constant and correlated spectral errors in the data. However, an estimate of the error covariance structure of the spectral image data is still required. This estimate can be obtained from the repeat image spectral differences of the sample. If the sample does not move and does not change over time, then any changes in the image spectra at a given pixel are related to systematic changes in the system response as well as random noise.

In the following, vectors are represented as bold lower-case letters, matrices as bold upper-case letters, and scalars are in italics. Vectors are presented as columns with row vectors and transposed matrices represented by a superscript T. A superscript -1 indicates a matrix inversion and a superscript + represents the pseudoinverse of a matrix. The ACLS method is based on the standard linear additive model (Beer's law for absorbance spectra) given in Eq. 1.

$$\mathbf{A} = \mathbf{CK} + \mathbf{E_A} \qquad (1)$$

where $\mathbf{A}$ is the $n \times p$ spectral matrix of the absorbance spectra from the n pixels at p spectral frequencies, $\mathbf{C}$ is the $n \times m$ matrix of the m component concentrations, $\mathbf{K}$ is the $m \times p$ matrix of pure-component spectra, and $\mathbf{E_A}$ is the $n \times p$ matrix of spectral errors. Generally $\mathbf{E_A}$ is assumed to consist of identically distributed uncorrelated errors of zero mean. However, instrument drift, system artifacts, and a variety of possible errors that occur in collecting the interferogram will generate correlated error structure. The conversion to absorbance also has the effect of generating errors with non-uniform error variance. The error covariance structure of $\mathbf{E_A}$ has often been estimated in remote sensing applications with hyperspectral images with a shift difference generated from a single hyperspectral image. This shift difference assumes that the signal between pixels is slowly varying and therefore the spectral difference between adjacent pixels is representative of noise. The shift difference must be used if direct repeat image spectra cannot be obtained, for example in the case of a moving detector in remote sensing applications. However, in the laboratory, repeat image spectra can be obtained, and the difference between these images can be used to obtain an estimate of the error covariance structure of $\mathbf{E_A}$. Since errors in our image data are correlated, a PCA analysis of $\mathbf{E_A}$ can serve to separate the correlated errors from the random errors. Thus, Eq. 1 can be expanded as:

$$\mathbf{A} = \mathbf{CK} + \mathbf{TP} + \mathbf{E} \qquad (2)$$

where TP is the product of scores and eigenvectors from the PCA of $\mathbf{E_A}$ for the r scores and eigenvectors that represent the significant correlated errors contained in $\mathbf{E_A}$. The number of eigenvectors was selected based on a visual examination of the corresponding score images. Score images with nonrandom patterns were retained. $\mathbf{E}$ then contains primarily the random errors that are expected in any least squares analysis. We wish to estimate $\mathbf{C}$ and $\mathbf{K}$ without the detrimental influence of $\mathbf{T}$ and $\mathbf{P}$ that can contaminate the

concentration and pure-component spectral estimates. We can accomplish this goal by augmenting either **C** with **T** or **K** with **P** or both. By augmenting columns of **C** with selected columns from **T**, we force the least squares estimate of **K** to ignore any pattern in the image that has the same spatial distributions as any give set of column scores in the **T** vectors that are used to augment **C**. By augmenting the rows of **K** with selected rows from **P**, we require that the least squares estimate of **C** ignore any spectral shape contained in any given row eigenvector of **P** vectors that are used to augment **K**. The corresponding least squares calibration and prediction equations that are part of the alternating least squares MCR procedure are

$$\hat{\tilde{\mathbf{K}}} = (\tilde{\mathbf{C}}\tilde{\mathbf{C}}^{\mathrm{T}})^{-1}\tilde{\mathbf{C}}^{\mathrm{T}}\mathbf{A} = \tilde{\mathbf{C}}^{+}\mathbf{A} \quad (3)$$

$$\hat{\tilde{\mathbf{C}}} = \mathbf{A}\hat{\tilde{\mathbf{K}}}^{\mathrm{T}}(\hat{\tilde{\mathbf{K}}}\hat{\tilde{\mathbf{K}}}^{\mathrm{T}})^{-1} = \mathbf{A}\hat{\tilde{\mathbf{K}}}^{+} \quad (4)$$

respectively, where the ~ symbol represents augmented matrices and the ^ symbol indicates least squares estimated matrices. In practice, the MCR method implements the augmentation by using equality constraints to force the estimates of **K** and **C** to ignore sources of spectral and spatial variation that are contained in the augmented **C** and **K** matrices. A recent publication from our group describes several implementations of the equality constraints that are used in the constrained MCR analyses of the spectral images. We have the option of augmenting with either scores or eigenvectors or both. All three options are presented in this paper to identify the best procedure to use with these example data.

RESULTS AND DISCUSSION

A visible video image of the approximate portion of the aged neoprene sample imaged by the FT-IR spectrometer is displayed in Fig. 1. A slight darkening from the center to the edge of the sample is the only indication in the visible image of the degradation of the sample due to the aging. The 1269 spectra from image pixels representing the sample are presented in Fig. 2. Note that there are significant complex baseline variations present in these data. The spectra that were subjected to linear baseline correction for each of the six significant spectral bands of neoprene are presented in Fig. 3. The preprocessed spectra in Fig. 3 were subjected to MCR analysis.

Figure 1. Video image of the thin neoprene sample aged for 6 days in air at 140° C.

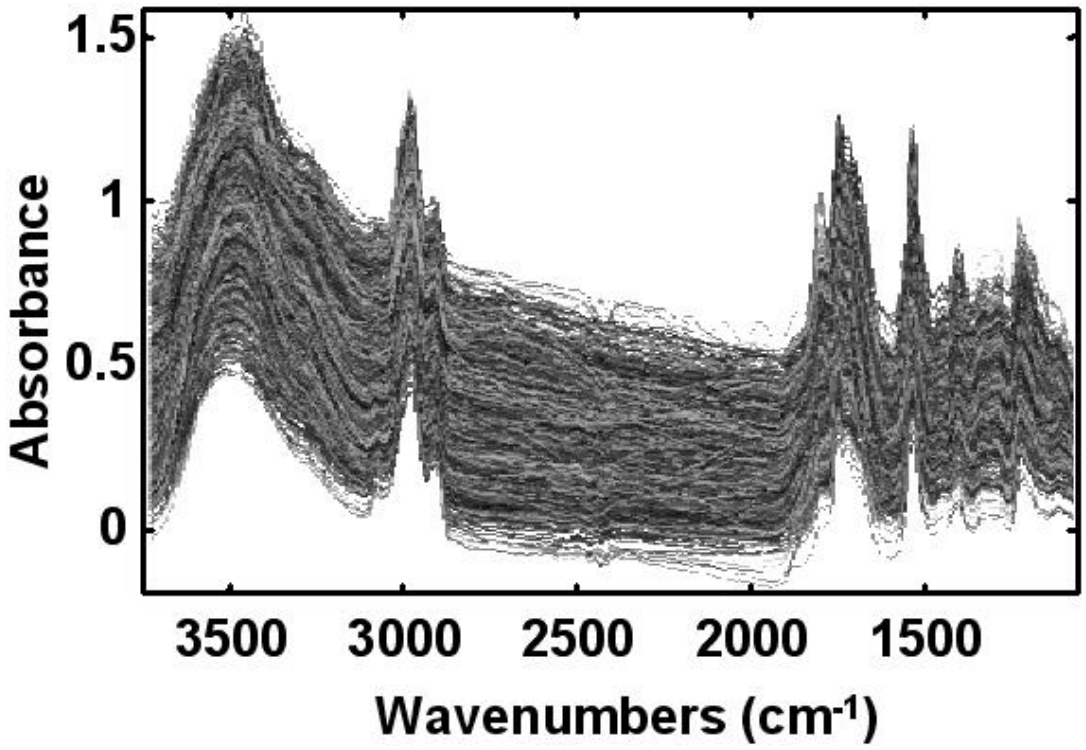

Figure 2. Spectra of neoprene pixels from spectral image of the neoprene sample.

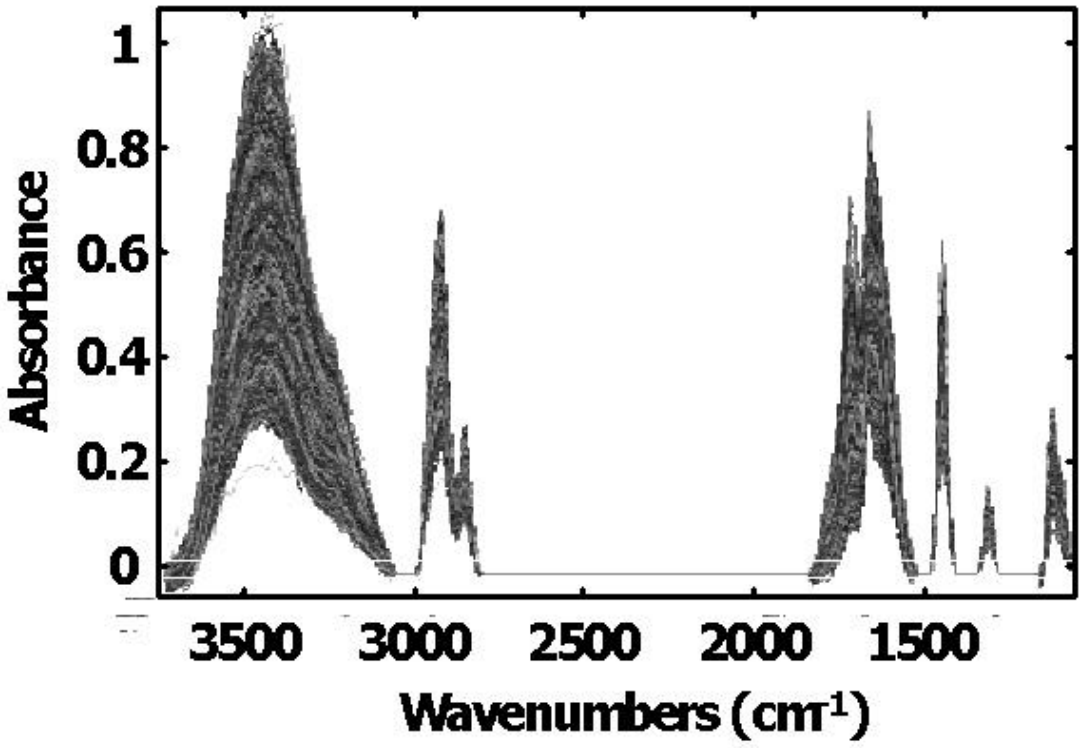

Figure 3. Spectra of neoprene bands from spectral image of the sample after removing a linear baseline from each of the 6 neoprene related spectral bands.

As mentioned in the Experimental Section, systematic artifacts were present in the FT-IR spectral image data. These artifacts were not always readily observable in the raw interferograms, single-beam spectra, or absorbance spectra. However, a PCA analysis of the spectral image data revealed the presence of the artifact as a systematic striped or banded pattern in the score images from the PCA analysis. As indicated in Fig. 4, this striped pattern was apparent in the PCA scores of interferograms, single-beams, and absorbance spectra obtained from the spectra of different sample or background images. The relative magnitude of the artifact varied considerably between various image data sets as evidenced by how many eigenvectors were generated before the artifact became noticeable. The data in Fig. 4 illustrates collected image data where the presence of the artifact varied from very small to significant as indicated by the fact that the striping artifact first becomes visible in factor 2, 5, or 10 depending on the image. In order to demonstrate the value of the ACLS methods applied to the MCR analysis, we choose as an example a spectral image where the artifact was a significant portion of the data.

Not only is the artifact clearly present in the PCA score images, it is also present in the spectral data. The systematic error (noise) can be observed in Fig. 5 that shows the mean-centered difference spectra between repeat images of the aged neoprene sample. Clearly, the noise in the difference spectra in Fig. 5 is not uniform. In fact, as expected by theory, the noise is higher in those regions where the absorbances are larger. Thus, the spectral noise is higher where the neoprene sample absorbs the IR beam. What is not clear from Fig 5 is that the noise in the sample is also correlated between spectral

frequencies. Thus, if the error is high at frequency p, it tends to be high at surrounding frequencies. The fact that the noise is high in broad spectral regions is not necessarily an indication of correlated error; it could simply be an indication that the noise is higher in broad spectral regions but uncorrelated between frequencies. The non-uniform and correlated nature of the noise in the spectral images can be best be observed by forming the error covariance matrix from the mean-centered repeat image spectral differences. This two-dimensional representation of the spectral error in the image is shown in Fig. 6A. Fig. 6B shows the ideal noise behavior of uniform variance and uncorrelated errors (i.e., only diagonal elements of constant magnitude). Clearly, the true error covariance structure of the spectral image data has non-uniform diagonal variance components and large off diagonal terms demonstrating that errors are highly correlated between frequencies. Thus, the errors observed from our measured spectral image data are a far departure from the ideal behavior presented in Fig. 6B.

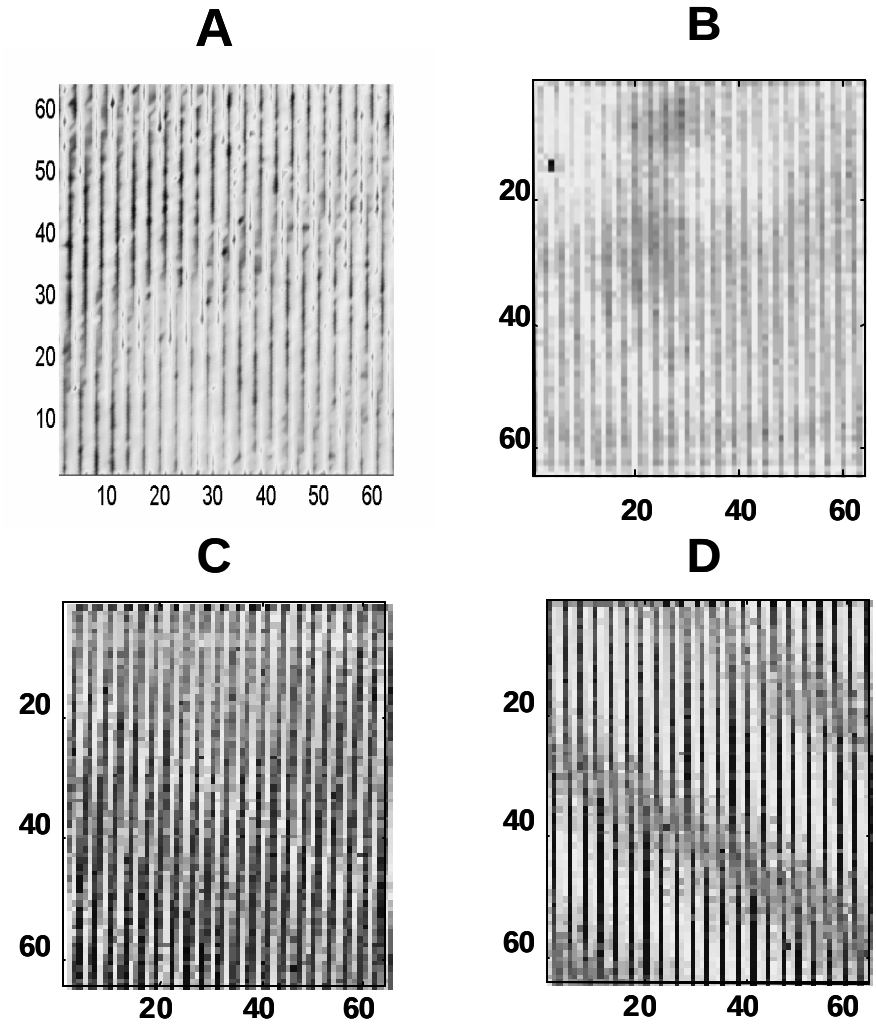

Figure 4. Examples of systematic artifacts in the spectral image scores, A) scores from factor 5 of an interferogram, B) scores from factor 10 of the negative log of a sample single beam, C) scores from factor 2 of two ratioed background single-beam spectra after conversion to absorbance, and D) scores from factor 3 from same two ratioed background single-beam spectra after conversion to absorbance.

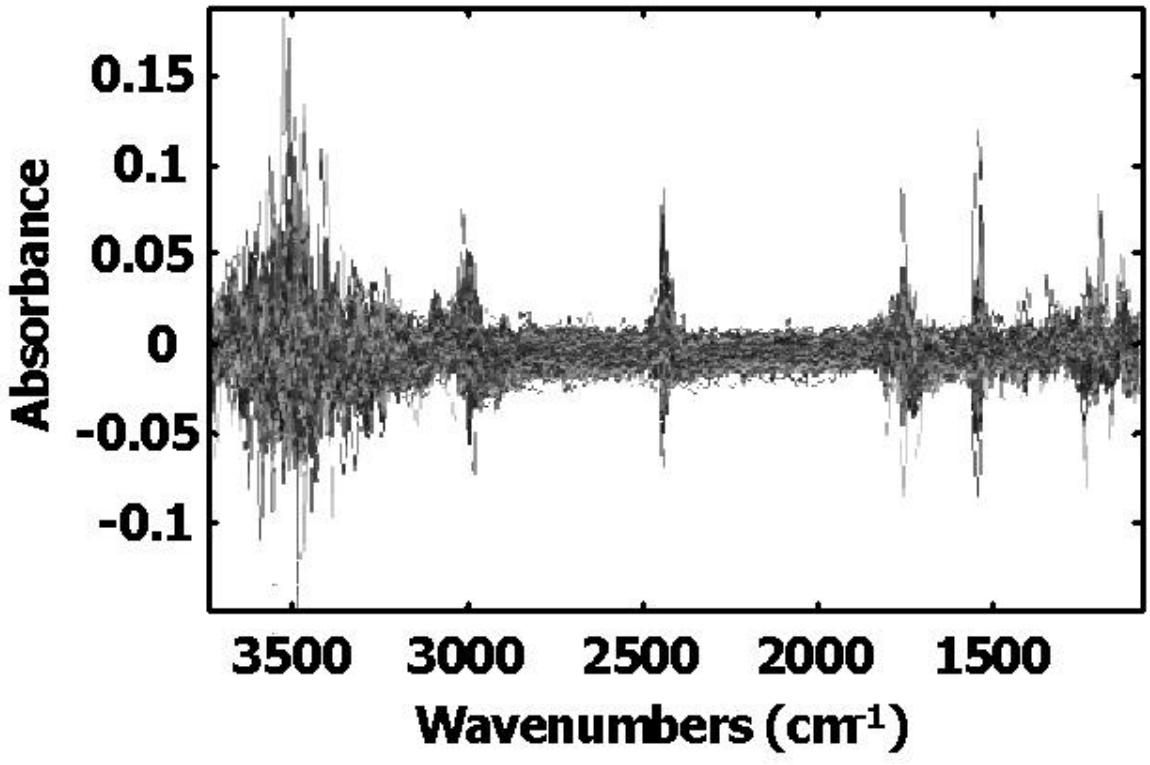

Figure 5. Mean-centered repeat spectra from the neoprene pixels in the spectral image.

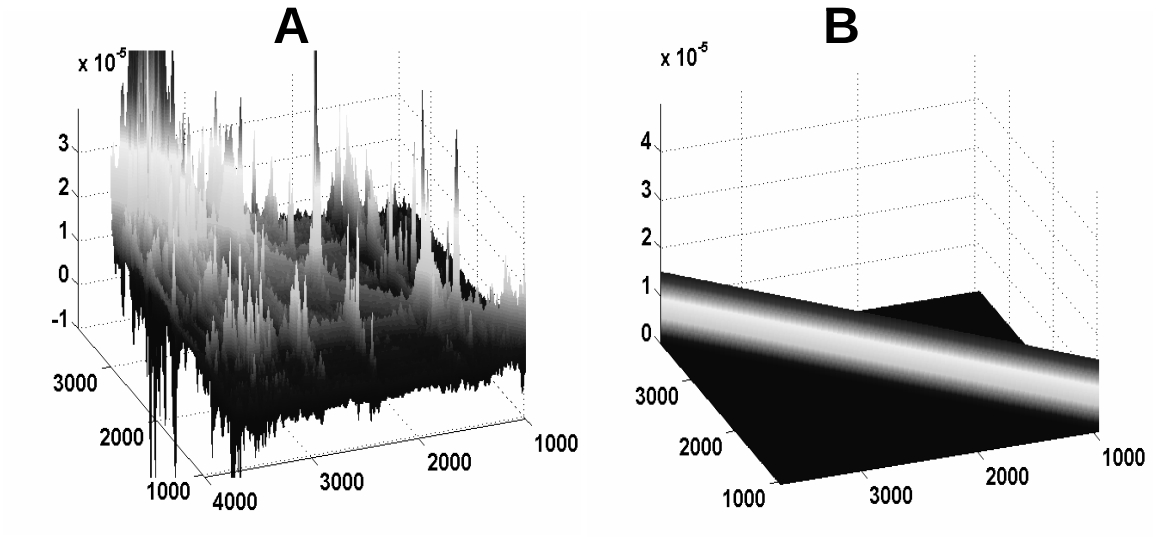

Figure 6. A) Covariance error structure of the mean-centered repeat spectra, B) Ideal error covariance structure assumed for standard least-squares analyses.

The non-constant noise variance for infrared absorbance spectra is expected since the log transformation required to obtain absorbance spectra takes noise variance that is expected to be constant at all frequencies for MCT detectors and makes the noise nonuniform over the spectral frequencies. The expected form of the noise variance of the difference of two repeat absorbance spectra at a single frequency can be estimated as outlined below.

$$A = \log \frac{(I_s + s_{s1})}{(I_s + s_{s2})} = \log(I_s + s_{s1}) - \log(I_s + s_{s2}) \qquad (5)$$

where I_s is the noise-free single-beam intensity of the two repeat sample spectra and σ_{s1} and σ_{s2} are the standard deviations of the noise from repeat spectra 1 and 2, respectively. From Eq. 5, we can obtain estimates of the error variances in absorbance (s_A^2) by taking a Taylor series expansion of the error variance in the single-beam intensities in Eq. 5 and retaining only the first non-zero term for both terms on the right-hand side of Eq. 5. The result is:

$$s_A^2 \approx \left(\frac{dA}{dI_s}\right) s_{s1}^2 + \left(\frac{dA}{dI_s}\right) s_{s2}^2 = \frac{s_{s1}^2}{I_s^2} + \frac{s_{s2}^2}{I_s^2} \approx \frac{2s_s^2}{I_s^2} \qquad (6)$$

where the last term on the right-hand side of Eq. 6 assumes that the error variance is comparable for both repeat single-beam spectra. From Eq. 6, we see that the error variance for the absorbance spectra is inversely proportional to the square of the single-beam intensity for the infrared spectrum. In a simple weighted least squares fit of the data, the weights would be proportional to the square root of the inverse of the spectral absorbance variance. If experimental estimates follow the expected theory, then the approximate proper weights would be proportional to the average single beam spectrum obtained from the image spectra. Fig. 7 compares the estimated weighting function for absorbance error variance obtained from the repeat sample image differences and the theoretical weighting function based on the average single-beam intensity obtained from the sample image data. Clearly, the experimental and theoretical estimates for the weighting function are very similar demonstrating that the repeat spectral differences can lead to reasonable estimates for the error variance of the absorbance spectral data. Confirmation that the full covariance structure of the error is captured by the repeat image spectral differences rests in our ability to use this information to reduce or eliminate systematic errors during the image analysis.

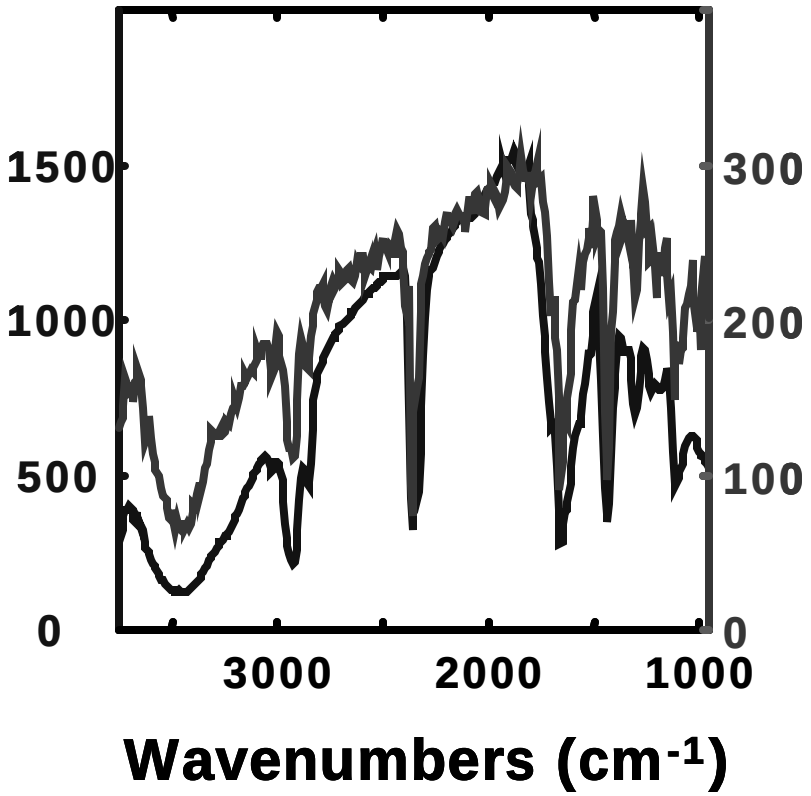

Figure 7. The upper trace is the inverse of the square root of the variance component from the mean-centered repeat spectra (left axis). The lower trace is the average single-beam spectrum of the neoprene pixels (right axis).

Conventional MCR methods were applied to the spectral image data presented in Fig. 3 using non-negativity constraints on both concentrations and pure-component spectra. Three pure components appeared to be adequate to represent the original thermally modified neoprene and the decomposition products. Using more than three pure components yielded results that were not as physically meaningful as obtained with three components. The normalized MCR pure-component spectra obtained with three components are presented in Fig. 8. The trace corresponding to component 1 is characteristic of the thermally aged but unoxidized neoprene sample. The component 2 trace indicates a decomposition product that contains a significant carbonyl decomposition product. The component 3 trace indicates that a second decomposition product is present in the aged sample that is characterized by a hydroxyl dominated decomposition product with very little formation of a carbonyl species. This second decomposition product has a different spatial profile in the sample than the first. Note that the high-frequency region of the pure-component spectra is quite noisy.

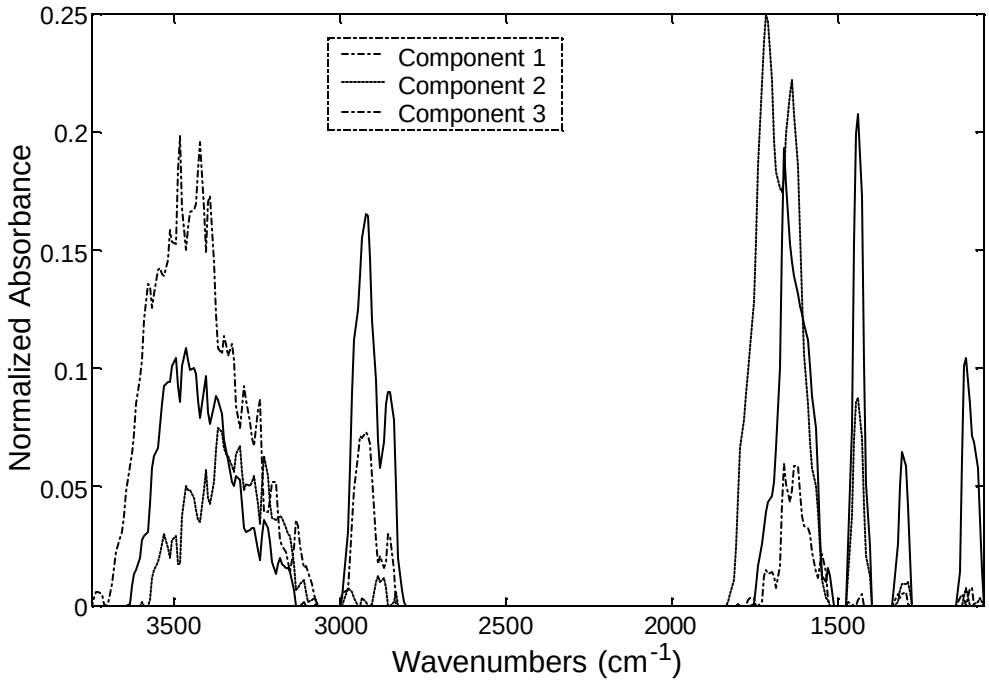

Figure 8. Estimated pure-component spectra using standard non-negativity constrained MCR.

Pathlength-corrected relative concentration maps representing the three pure are presented in Fig. 9. Some of the striped artifact presented in Fig. 4 is clearly visible in at least two of these concentration maps indicating that the MCR analyses are corrupted by this artifact. The carbonyl decomposition product (middle image) demonstrates a large concentration gradient that is high at the outer edges of the sample. The hydroxyl component decomposition product shown on the right side of Fig. 9 is more variable and extends somewhat beyond the region containing the carbonyl decomposition product. The thermally aged but unoxidized neoprene on the left of Fig. 9 is present somewhat uniformly throughout the sample although it may have been reduced near the edge of the sample by the oxidation. Full evaluation and interpretation of the results and interpretation of the MCR results will be the subject of another manuscript that includes images and evaluations from many more sample-aging conditions. The primary purpose of this paper is to demonstrate the power of ACLS methods coupled with MCR in reducing the influence of system artifacts.

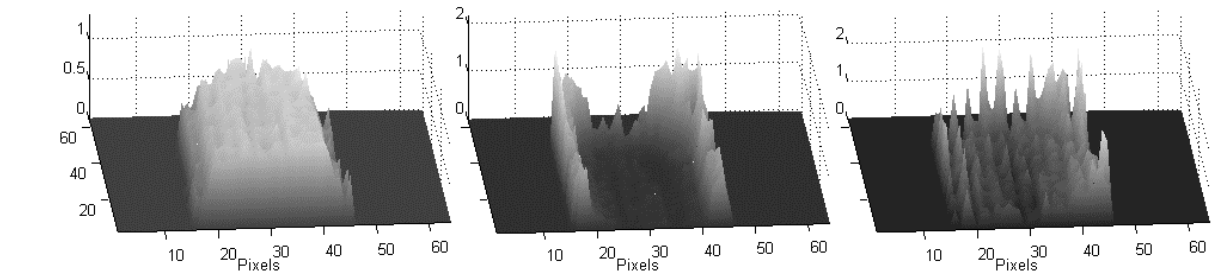

Figure 9. Concentration maps for components 1, 2, and 3, respectively obtained from the standard non-negativity constrained MCR.

MCR was then applied to the same spectral image data with the concentration matrix augmented by the first five sets of scores from the mean-centered repeat image spectral differences. The MCR was implemented with equality constraints on these five sets of image scores. In applying these equality constraints on the scores, we correct the alternating least-squares generation of the pure-component spectra for the presence of score-related concentration errors in the image data. Fig. 10 presents the three pure-component spectra generated from the MCR augmented with the scores. Note that the noise in the high-energy region of the pure-component spectra is greatly reduced in these pure-component estimates. More detailed structure becomes available in the OH stretching region of the spectra (3000 – 3700 cm^{-1} spectral region) with score augmentation. The derivative nature of the hydroxyl band for the carbonyl dominated decomposition product (component 2) may indicate some hydrogen bonding with the hydroxyl dominated decomposition product. Since we have not corrected the concentrations when augmenting with scores, the concentration maps indicate essentially no improvement in the striping artifact relative to the standard non-augmented MCR analysis of the same data.

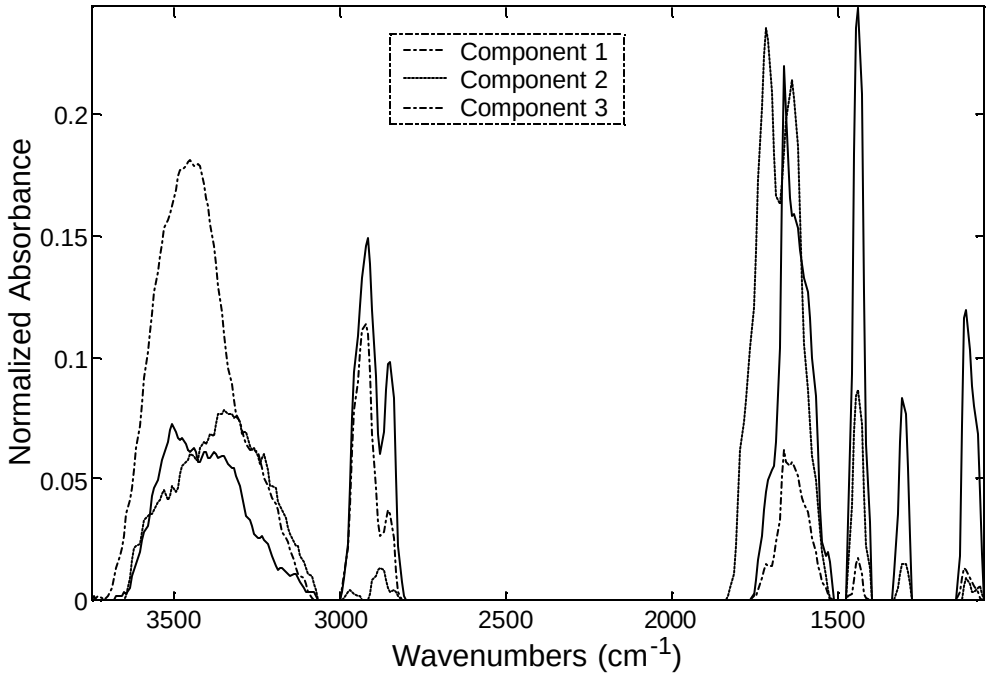

Figure 10. Estimated pure-component spectra augmenting with the first five scores from the mean-centered repeat spectra and using non-negativity and equality constrained MCR.

MCR was then applied to the same spectral image data the first five sets of eigenvectors from the mean-centered repeat image spectral differences added to the pure-component spectral matrix. The MCR was implemented with equality constraints on these five eigenvectors. There is some improvement in noise of the high-energy side of the hydroxyl-dominated pure component, but there is little, if any, improvement in the noise on the other pure-component spectra. However, the concentration maps exhibit dramatic improvement in the quality. The striping artifact is nearly completely removed when the eigenvectors augment the MCR as shown in Fig. 11. These improvements in concentration maps were expected since the augmentation of the pure-component spectra with the eigenvectors corrects the concentration estimates. This correction of concentrations with spectral augmentation was described in detail in our earlier paper describing the prediction-augmented CLS (PACLS) method applied to non-image spectral data.

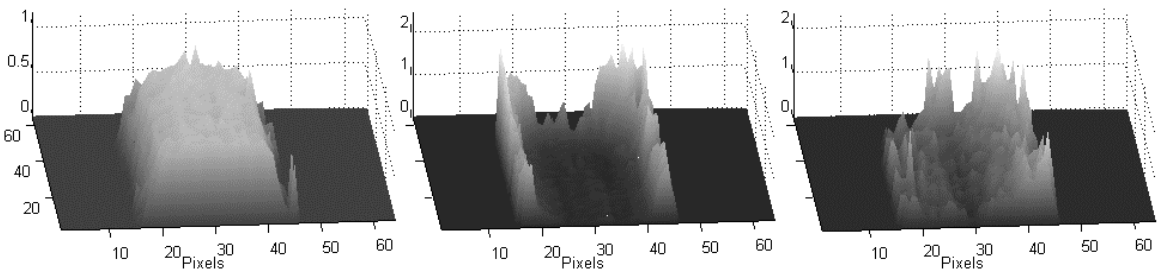

Figure 11. Concentration maps for components 1, 2, and 3, respectively obtained augmenting with the first five eigenvectors from the mean-centered repeat spectra and using non-negativity and equality constrained MCR.

Finally, we tested augmenting the MCR method with both scores and eigenvectors from the repeat image spectral differences. In this implementation, one simply augments with scores during the calibration phase of the MCR and augments with the eigenvectors during the prediction phase. Thus, at each iteration in the MCR procedure, any estimates related to the augmented vectors are replaced with the original eigenvectors from the PCA decomposition of the repeat image differences. The combined augmentation with scores and eigenvectors generates improved pure-component spectra and corrected concentration maps. The pure-component spectra with scores and eigenvector augmentation are presented in Fig. 12. These spectra have the low noise of the scores-augmented MCR and are also slightly modified, especially in the high-energy hydroxyl spectral region. The concentration maps with scores and eigenvector augmented MCR are presented in Fig. 13. Like the eigenvector- augmented MCR concentration maps, these images exhibit greatly reduced striping artifact. The reduction in the artifact is emphasized in Fig. 14, which shows the difference between the standard MCR and the fully augmented MCR composition maps for all three components. Clearly, a significant reduction in the system artifact has been achieved when augmenting with estimates of the error covariance information derived from the difference between two repeat images.

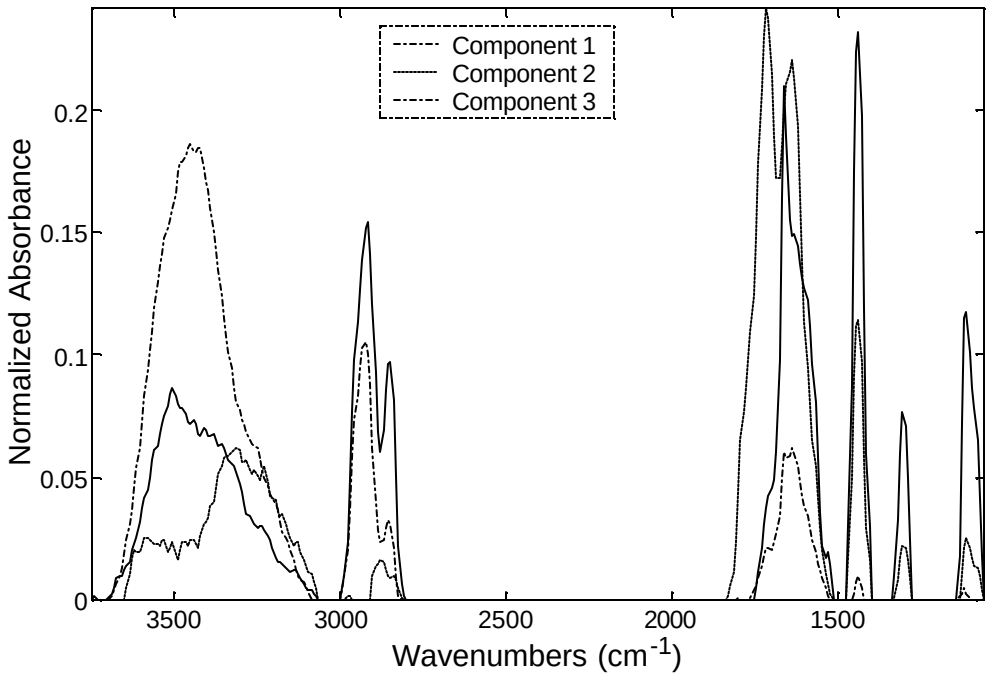

Figure 12. Estimated pure-component spectra augmenting with the first five sets of scores and eigenvectors from the mean-centered repeat spectra and using non-negativity and equality constrained MCR.

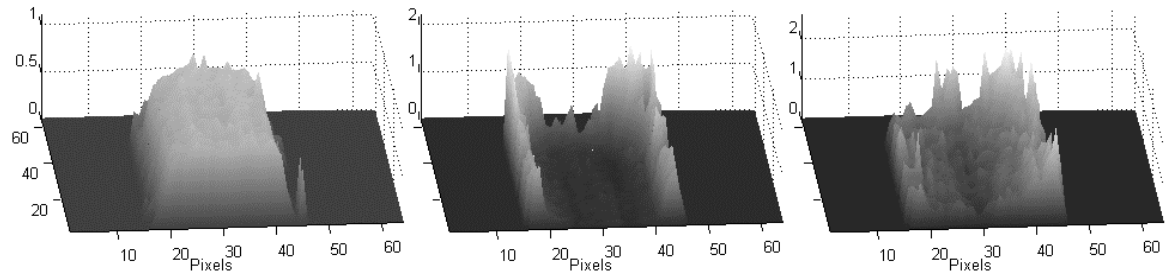

Figure 13. Concentration maps for components 1, 2, and 3, respectively obtained augmenting with the first five sets of scores and eigenvectors from the mean-centered repeat spectra and using non-negativity and equality constrained MCR.

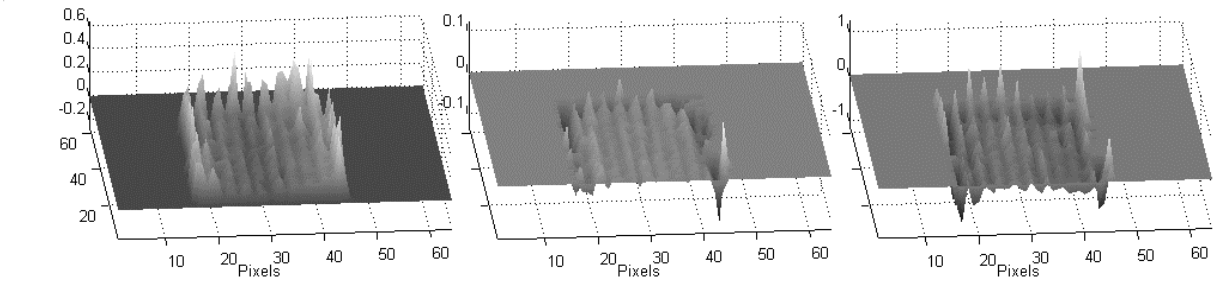

Figure 14. Differences in the concentration maps of components 1, 2, and 3, respectively from the standard MCR and MCR augmented with the first five sets of scores and eigenvectors from the mean-centered repeat image spectra of the neoprene pixels.

CONCLUSIONS

Significant improvements have been made in the quantitative analysis of FT-IR hyperspectral image data. We have demonstrated the quantitative analysis of spectral image data without the need for calibration standards. The standard method of applying MCR with non-negativity constraints to spectral data has been improved with the augmentation of the CLS calibration and prediction solutions with estimates of the error covariance structure in the spectra. Mean-centered repeat image spectral differences provide an estimate of the error covariance structure in the image spectra and give the analyst an opportunity to minimize or eliminate the influence of system artifacts on the analysis of the image data. The MCR with ACLS is implemented not only with non-negativity constraints, but also with the addition of equality constraints applied to the augmented eigenvectors and/or scores generated from mean-centered repeat image difference spectra. We have demonstrated with an aged neoprene sample containing system artifacts in the image data that augmentation with scores improves the pure-component spectral estimates while augmentation with the corresponding eigenvectors improves primarily the relative concentration maps derived from the image data. Augmentation with both scores and eigenvectors improves both the pure-component spectra and the concentration maps.

Of course the best solution for eliminating the detrimental effects of system artifacts is to identify the experimental sources of the artifacts and eliminate the artifact at its source. Software could then be used to minimize the detrimental effects of any small residual artifact. However, augmenting the steps of the MCR analysis corrects for non-uniform variance that is inherent in infrared spectral data, and therefore, ACLS methods should always be used in the MCR analysis of infrared data. ACLS methods will generally be useful in the analysis of any type of spectral data where estimates of the error covariance structure of the data can be obtained. We are now applying MCR methods to a wide variety of spectral data beyond hyperspectral image data. We expect that the MCR methods with a variety of constraints and coupled with augmentation will have applicability to the quantitative and qualitative analysis of multivariate spectral data.

ACKNOWLEDGEMENTS

We would like to acknowledge Mathias Celina for preparing the neoprene samples. Michael R. Keenan provided useful discussions and improvements in the Matlab code for the constrained MCR algorithms. David K. Melgaard is acknowledged for software that facilitated the preprocessing of the spectral image data.

REFERENCES

1. Lewis, E.N., et al., *Fourier-Transform Spectroscopic Imaging Using an Infrared Focal-Plane Array Detector.* Analytical Chemistry, 1995. **67**(19): p. 3377-3381.
2. Koenig, J.L. and C.M. Snively, *Fast FT-IR imaging: Theory and applications.* Spectroscopy, 1998. **13**(11): p. 22-28.
3. Andrew, J.J., et al., *Raman imaging of emulsion systems.* Applied Spectroscopy, 1998. **52**(6): p. 790-796.
4. Andrew, J.J. and T.M. Hancewicz, *Rapid analysis of Raman image data using two-way multivariate curve resolution.* Applied Spectroscopy, 1998. **52**(6): p. 797-807.
5. Haaland, D.M. and D.K. Melgaard, *New Augmented Classical Least Squares Methods for Improved Quantitative Spectral Analyses.* Vibrational Spectroscopy, 2001: p. In Press.
6. Bro, R. and S. DeJong, *A fast non-negativity-constrained least squares algorithm.* Journal of Chemometrics, 1997. **11**(5): p. 393-401.
7. Tauler, R., B. Kowalski, and S. Fleming, *Multivariate Curve Resolution Applied to Spectral Data from Multiple Runs of an Industrial-Process.* Analytical Chemistry, 1993. **65**(15): p. 2040-2047.
8. Saurina, J., S. Hernandezcassou, and R. Tauler, *Multivariate Curve Resolution Applied to Continuous-Flow Spectrophotometric Titrations : Reaction between Amino-Acids and 1 ;2-Naphthoquinone-4-Sulfonic Acid.* Analytical Chemistry, 1995. **67**(20): p. 3722-3726.
9. Casassas, E., I. Marques, and R. Tauler, *Study of Acid-Base Properties of Fulvic-Acids Using Fluorescence Spectrometry and Multivariate Curve Resolution Methods.* Analytica Chimica Acta, 1995. **310**(3): p. 473-484.
10. Tauler, R., et al., *Application of a New Multivariate Curve Resolution Procedure to the Simultaneous Analysis of Several Spectroscopic Titrations of the Copper(Ii)-Polyinosinic Acid System.* Chemometrics and Intelligent Laboratory Systems, 1995. **27**(2): p. 163-174.
11. Esteban, M., et al., *Multivariate curve resolution with alternating least squares optimisation: a soft-modelling approach to metal complexation studies by voltammetric techniques.* Trac-Trends in Analytical Chemistry, 2000. **19**(1): p. 49-61.
12. Martens, H. and T. Naes, *Multivariate Calibration*. 1989, Chichester: John Wiley & Sons.
13. Koehler IV, F.W. and D.M. Haaland, *Quantitative Analysis of Ink-jet Inks with an FT-IR Imaging Spectrometer.* Applied Spectroscopy, 2001: p. In Preparation.
14. Haaland, D.M. and R.G. Easterling, *Improved Sensitivity of Infrared-Spectroscopy by the Application of Least-Squares Methods.* Applied Spectroscopy, 1980. **34**(5): p. 539-548.
15. Windig, W., *Spectral data files for self-modeling curve resolution with examples using the Simplisma approach.* Chemometrics and Intelligent Laboratory Systems, 1997. **36**(1): p. 3-16.

DISTRIBUTION:

1 Dr. Mathias Celina
Queensland University of Technology
Faculty of Science
School of Physical Science
Brisbane, Australia

1 Fredrick W. Koehler IV
Spectral Dimensions, Inc.
3416 Olandwood Court
Suite 210
Olney, MD 20832

1	MS-0188	Donna L. Chavez	1030
1	MS-0885	Duane B. Dimos	1802
1	MS-0886	Nancy B. Jackson	1812
1	MS-0886	Christine M. Wehlburg	1812
1	MS-0886	Mark H. Van Benthem	1812
1	MS-0886	Michael R. Keenan	1812
5	MS-0886	David M. Haaland	1812
1	MS-0889	David K. Melgaard	1835
1	MS-9018	Central Technical Files	8945-1
2	MS-0899	Technical Library	9616
1	MS-0612	Review & Approval Desk For DOE/OSTI	9612
1	MS-0161	Patent and Licensing Office	11500