SURFACE AND INTERFACE ANALYSIS
Surf. Interface Anal. 2004; **36**: 203–212
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sia.1657

SIA

Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images†

Michael R. Keenan* and Paul G. Kotula

Sandia National Laboratories, Albuquerque, NM 87185-0886, USA

Received 1 July 2003; Revised 17 October 2003; Accepted 24 October 2003

Recent years have seen the introduction of many surface characterization instruments and other spectral imaging systems that are capable of generating data in truly prodigious quantities. The challenge faced by the analyst, then, is to extract the essential chemical information from this overwhelming volume of spectral data. Multivariate statistical techniques such as principal component analysis (PCA) and other forms of factor analysis promise to be among the most important and powerful tools for accomplishing this task. In order to benefit fully from multivariate methods, the nature of the noise specific to each measurement technique must be taken into account. For spectroscopic techniques that rely upon counting particles (photons, electrons, etc.), the observed noise is typically dominated by 'counting statistics' and is Poisson in nature. This implies that the absolute uncertainty in any given data point is not constant, rather, it increases with the number of counts represented by that point. Performing PCA, for instance, directly on the raw data leads to less than satisfactory results in such cases. This paper will present a simple method for weighting the data to account for Poisson noise. Using a simple time-of-flight secondary ion mass spectrometry spectrum image as an example, it will be demonstrated that PCA, when applied to the weighted data, leads to results that are more interpretable, provide greater noise rejection and are more robust than standard PCA. The weighting presented here is also shown to be an optimal approach to scaling data as a pretreatment prior to multivariate statistical analysis. Published in 2004 John Wiley & Sons, Ltd.

KEYWORDS: spectral imaging; principal component analysis; multivariate statistical analysis; Poisson statistics; data weighting

INTRODUCTION

Spectral imaging techniques are becoming the tools of choice for performing comprehensive surface and near-surface microcharacterization. Such instruments are capable of collecting a full spectrum at each point in a spatial array of points. In this paper, we will use the term 'spectrum image' expansively to describe data sets obtained as line scans, traditional two-dimensional images and data sets having three spatial dimensions. A time-of-flight secondary ion mass spectrometry (ToF-SIMS) depth profile would exemplify a typical three-dimensional image. Very large spectrum images can be acquired using current generation instruments. For instance, one commercially available energy-dispersive x-ray spectrometer[1] can collect a full 1024-channel spectrum at each pixel on a 1024×1024 grid. The spectrum image, in this case, comprises over 1 billion individual data elements. Once a third spatial dimension is allowed, the potential size of a spectrum image is seemingly unlimited. The remaining challenge is to extract the relevant chemical information from the mountain of spectral data. In this pursuit, multivariate statistical techniques have a prominent role to play.

Principal component analysis[2] (PCA) and other factor analysis methods[3] have a long history in analyzing very diverse types of multivariate data. Much of the large body of relevant literature is summarized in the works referenced above. Multivariate techniques also have been employed for some time to analyze surface spectroscopic data. Factor analysis was applied to SIMS images more than 15 years ago.[4] More recently, multivariate techniques have been used to advantage when analyzing electron energy-loss spectroscopy (EELS) elemental maps[5,6] and spectrum images obtained from x-ray fluorescence spectroscopy (XRF),[7,8] x-ray photoelectron spectroscopy (XPS),[9,10] energy-dispersive spectrometry (EDS),[11] ToF-SIMS[12–15] and a variety of other surface spectroscopies.[16–18] In these spectroscopic imaging applications, factor analysis may be used alone to extract the spectral characteristics of the pure chemical components that exist in the sample and to determine their spatial distributions. Alternatively, factor analysis can be used to denoise spectral data and reduce its dimension as a preprocessing step prior to cluster analysis or other classification procedure.[19] Because PCA is probably the most

*Correspondence to: Michael R. Keenan, Sandia National Laboratories, MS0886, PO Box 5800, Albuquerque, NM 87185-0886, USA. E-mail: mrkeena@sandia.gov
Contract/grant sponsor: US DoE; Contract/grant number: DE-ACO4-94AL85000.

ubiquitous form of factor analysis and is often the first step in many multivariate statistical methods, it will be the focus of this paper. It must be kept in mind, however, that the basic principles discussed here have a more general application.

Although the aforementioned imaging techniques exploit a varied set of physical phenomena, they share a common characteristic, namely that they count particles in order to form their respective spectra. The particles might be photons (XRF), electrons (XPS) or ions (ToF-SIMS), but the fact that they are counted means that the noise or variability in the measured spectra will likely be governed by Poisson statistics. An accessible review of the Poisson probability distribution has been given recently[20] but, for our purposes, the main result is that the estimated variance of any individual measurement is equal to the magnitude of the measurement itself. In other words, the absolute magnitude of the uncertainty is not constant over an entire data set but, rather, varies from datum to datum. The implication of this variation is that standard PCA, although providing an optimal model for the *total* variance in a data set, is not optimal in terms of describing the *chemically relevant* variance,[2,21] i.e. variance arising from small, chemically significant factors may be ignored while attempting to account for large, but chemically insignificant noise in major features. A number of approaches, including weighted PCA[22] and maximum likelihood PCA (ML-PCA),[21] have been taken to overcome the shortcomings of standard PCA in the presence of heteroscedastic noise. What these approaches share in common is the incorporation of error covariance information into the analysis. Similar approaches have achieved success in the analysis of remotely sensed multi- and hyperspectral images[23,24] but they do not appear to be applied commonly in the surface analysis community.

The remainder of this paper will be devoted to describing a weighted PCA method that is suitable for use with spectral data sets that exhibit Poisson statistics. Owing to the direct relationship between a Poisson variable and its uncertainty, the error covariance structure can be estimated directly from the data itself and has a particularly simple form in the absence of correlated errors. The method will be demonstrated by applying it to the analysis of a simple ToF-SIMS spectrum image whose interpretation is intuitively obvious. Using both the original data and reconstructions of the data at different noise levels, the advantages of weighting will become apparent. In particular, the process of selecting the number of significant factors is made more obvious, and these factors will be shown to represent the chemical information with higher fidelity. Improved noise rejection and robustness are also characteristic of the new method.

EXPERIMENTAL

The ToF-SIMS spectrum image that will be used to illustrate the Poisson-weighted PCA method is shown in Fig. 1. The sample itself consists of a copper grid that has been plated on aluminum and whose central region has been sputter-cleaned with a gallium ion beam. This data set, which is distributed as part of PHI's[25] WinCadence software for their TRIFT ToF-SIMS instrument, contains slightly more than 2.1 million ion counts in the mass range 7–149 amu. The ions detected were binned to a resolution of 1 amu, giving a 143-channel mass spectrum at each pixel. Because the spectrum image is 256×256 pixels in size, each spectrum contains approximately 33 total counts, on average.

All of the calculations presented here were performed using Matlab[26] Version 6.5. For the simulated spectrum images, Poisson noise was added to varying levels using the POISSRND function in the Matlab Statistics toolbox.

PRINCIPAL COMPONENT ANALYSIS

The application of PCA in the specific case of multivariate image analysis has been described at length in a book of the same title.[27] Consider an $m \times n$ matrix $\boldsymbol{D}$ comprising the spectral data where m is the number of spectra (i.e. pixels) and n is the number of spectral channels. Assuming, without loss of generality, that $m > n$, the goal of PCA is to extract the useful information in the n-dimensional data set into a lower dimensional subspace. From a geometric point of view, PCA begins by finding that single direction in n-dimensional space that best describes the location of the data. That vector gives the direction of the first principal component. Once found, a second direction orthogonal to the first is determined that best accounts for the variation in the data that is not described by the first. This represents the second principal component. The process continues, with each new principal component maximally accounting for the variation in the data that is orthogonal to all preceding components. Typically, the first few principal components will contain the chemical information of interest. If there are p such components, the remaining $n - p$ components are assumed to describe experimental noise. Limiting further analysis to the p-dimensional subspace defined by the first p principal components provides the desired noise and dimensional reduction.

In matrix terms, PCA seeks to factor the data matrix $\boldsymbol{D}$

$$\boldsymbol{D} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} \tag{1}$$

where the superscript T represents a matrix transpose. As PCA is commonly implemented, $\boldsymbol{T}$ is an $m \times p$ matrix having orthogonal columns and is traditionally called the score matrix, whereas $\boldsymbol{P}$ is an $n \times p$ matrix having orthonormal columns and is called the loading matrix. Often, the data are 'centered' by subtracting the mean of all of the spectra from each individual spectrum prior to performing the matrix factorization. Recently, Bro and Smilde[28] have shown that the fit obtained by performing PCA on mean-centered data is equivalent to fitting the original data to Eqn. (1) plus a common offset. Because there is no common offset in ToF-SIMS data (i.e. there is a true zero), mean centering has no effect on the quality of the fit. Consequently, we will use the original, non-centered data in all of the calculations presented here.

There is a strong connection between PCA and the singular value decomposition (SVD)[29] of $\boldsymbol{D}$, which is a common technique for computing the principal components. The SVD of the data matrix can be written

$$\boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} \tag{2}$$

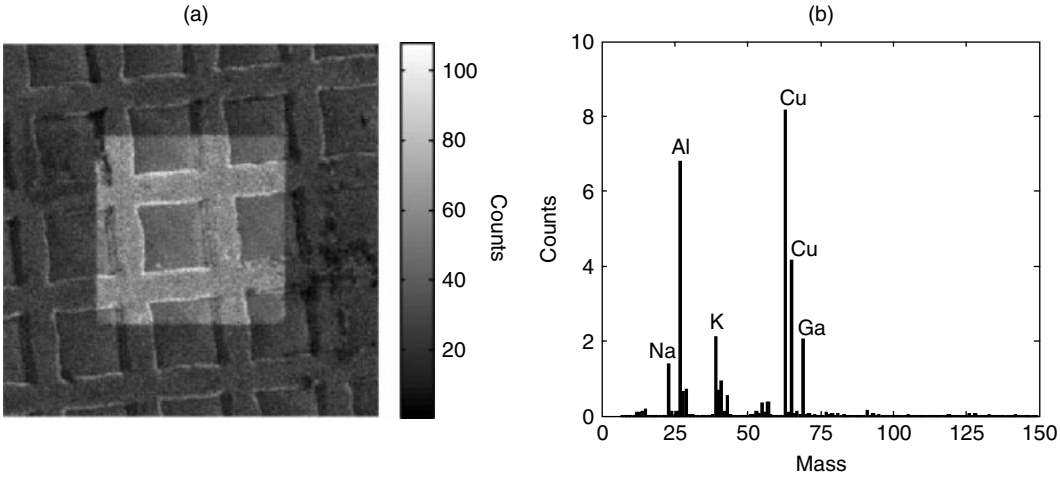

Figure 1. Total ion image (a) and mean spectrum (b) for the ToF-SIMS example spectrum image. The area of the sample imaged is 100 × 100 μm in size.

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices and $\boldsymbol{\Sigma}$ is a diagonal matrix with the singular values of $\boldsymbol{D}$ being placed along the diagonal. Comparison with Eqn. (1) shows that the scores and loadings can be obtained simply as

$$\boldsymbol{T} = (\boldsymbol{U}\Sigma)_p \quad \text{and} \quad \boldsymbol{P} = \boldsymbol{V}_p \tag{3}$$

The subscript p indicates that only the first p columns of the respective matrices are being retained. A principal component reconstruction of the data $\hat{\boldsymbol{D}}$ then can be computed as

$$\hat{\boldsymbol{D}} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} \tag{4}$$

The singular values themselves have important significance. The squares of the singular values are equal to the eigenvalues of the cross-product of the data matrix $\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}$. Each individual eigenvalue, in turn, quantifies the amount of variance in the data set that is accounted for by the corresponding principal component. Malinowski[30,31] has shown that for errors that are uncorrelated and uniform (i.e. a measurement's variance is independent of the measurement's magnitude) all of the eigenvalues that describe error are statistically equal once they have been normalized by the degrees of freedom inherent in their calculation. Assuming uniform error, a plot of the sorted eigenvalues should be a virtually horizontal line over the range of components that describes only noise and should exhibit positive deviation from that line when the principal components begin to describe systematic spectroscopic variation in addition to the error. This result is of fundamental importance. It says that if the assumption of uniform, uncorrelated error is satisfied, all of the information arising from the p spectroscopically active pure components in a sample will be localized in the first p principal components. Principal component analysis, in this case, will provide a model that is optimal in the sense that the chemical information is captured in the smallest possible number of components.

Of course, for the spectroscopic techniques considered here, where the observed spectra derive from Poisson processes, the noise is far from uniform and the standard PCA model will not be optimal in the sense outlined above. Principal components describing real, but small, spectroscopic features may be interspersed among, and confounded with, principal components that describe the noise. Figure 2 shows the eigenvalues of the copper grid data set as computed using standard PCA. Given the simplicity of the sample, one might intuitively expect that four principal components would be sufficient to describe the chemical variation in the spectrum image. These would include copper and aluminum, sputtered and not. The eigenvalue plot, on the other hand, suggests that there are perhaps 12 or 13 significant components. Figures 3 and 4 show the first six score images and loading vectors, respectively, as computed with standard PCA. Because the data have not

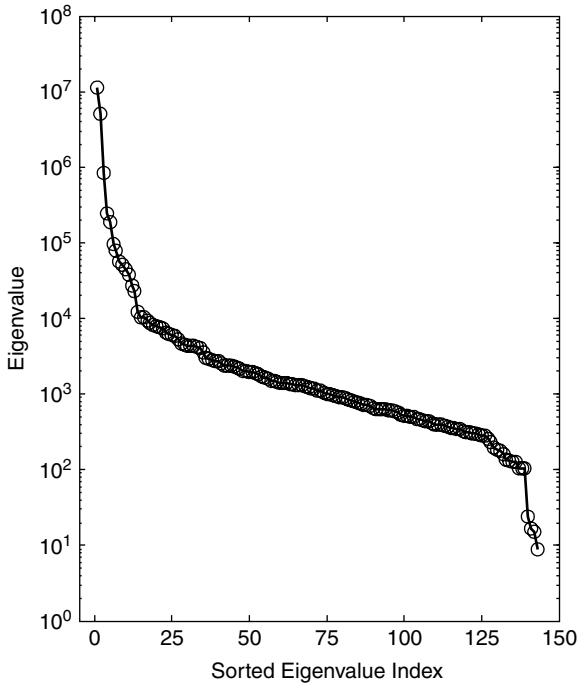

Figure 2. Sorted eigenvalues obtained using standard PCA for the copper grid example.

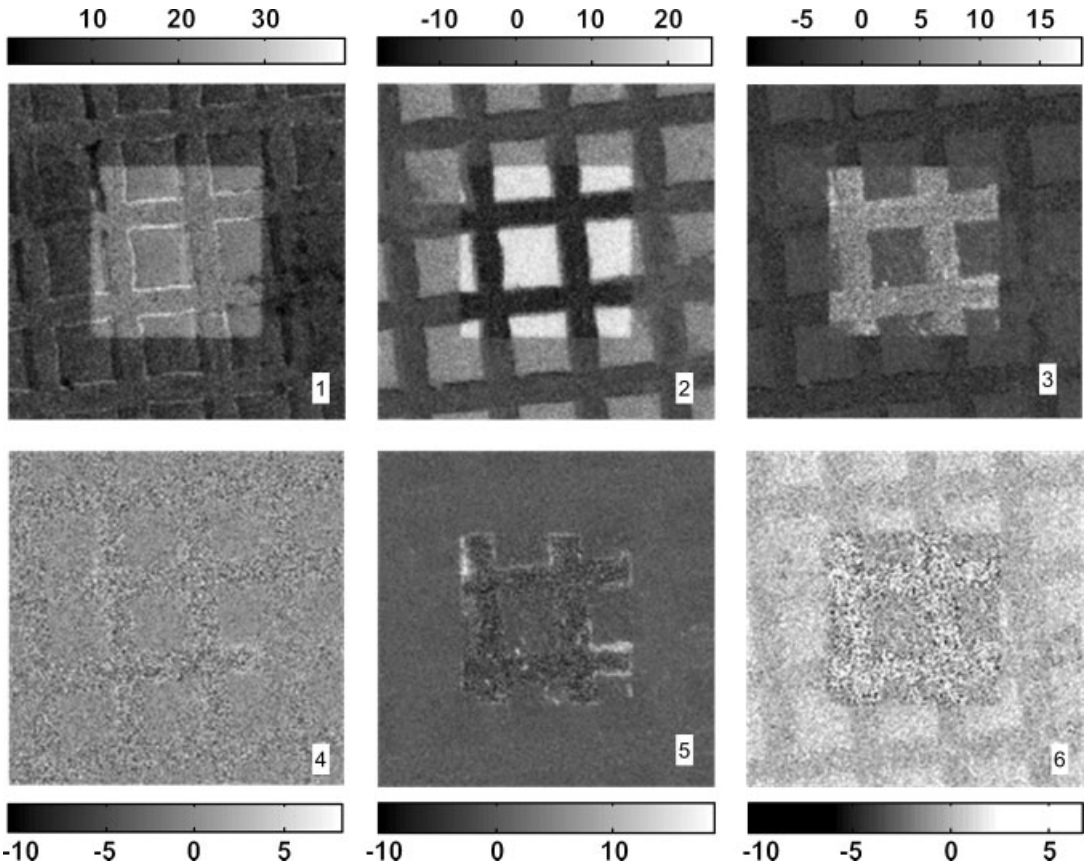

Figure 3. First six score images for the copper grid data set computed by standard PCA.

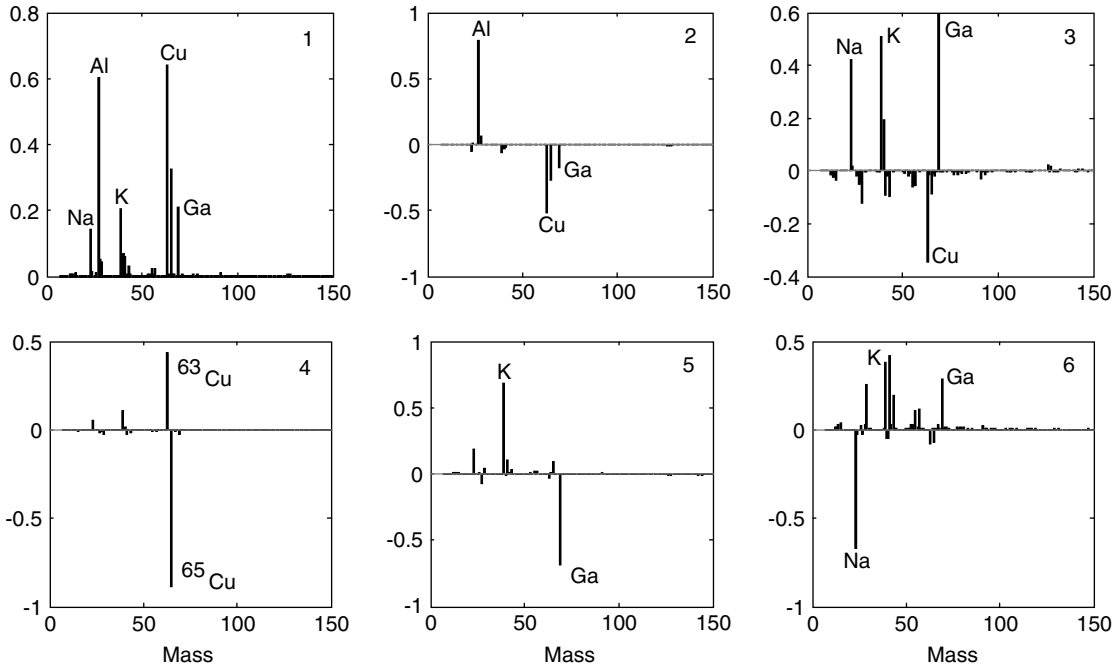

Figure 4. First six loading vectors for the copper grid data set computed with standard PCA.

been mean-centered, the first loading vector resembles the mean spectrum and is the single vector that best describes the data in a least-squares sense. The second component represents the anticorrelation of aluminum and copper, and the third component shows gallium from the sputtering process together with sodium and potassium. The latter species are likely to be residual from the plating process and were uncovered by the cleaning operation.

Component 3 together with components 5 and 6 describe, primarily, correlations and anticorrelations among gallium, sodium and potassium. Perhaps component 4 is the most interesting of the principal components. This component

describes, primarily, the anticorrelation of masses 63 and 65, which correspond to the two naturally occurring isotopes of copper. Because there is no obvious physical reason why the copper isotope ratio should vary on the scale of the measurement, the observed anticorrelation is curious. To gain some insight into what this component represents, consider that in the absence of noise and non-copper-related fragments the difference between the mass 63 signal and the mass 65 signal should be zero once each has been normalized by its respective natural abundance in copper. Deviations from zero simply reflect noise in the copper channels. Figure 5 compares an image of the copper noise computed in this fashion with the magnitude of the component 4 score. Clearly, the copper noise and component 4 are highly correlated (correlation coefficient = 0.97). Thus, we are left to conclude that component 4, although undeniably important in accounting for the overall variation of the data, is describing noise rather than meaningful chemical information. In other words, by assuming that noise is uniform, standard PCA finds that it can most effectively account for spectral variation by fitting large magnitude noise associated with high intensity signals rather than trying to fit small magnitude, yet chemically important, features.

WEIGHTED PRINCIPAL COMPONENT ANALYSIS

The shortcomings of standard PCA for describing the chemically interesting content of spectral data become apparent in the case when noise is not uniform. Several approaches have been taken to adapt PCA to accommodate heteroscedastic noise. These approaches include maximum likelihood PCA[21] and various forms of weighted PCA.[2,22] The maximum likelihood approach is perhaps the most general in that it allows each individual data element to have its own associated uncertainty. For the Poisson data considered here, such estimates are readily obtained because the estimated variance of a Poisson variable is simply equal to the value of the variable itself. In other words, if d_{ij} is the number of counts in the ijth data element, one estimate of its variance is

$$\mathrm{var}(d_{ij}) = d_{ij} \tag{5}$$

Unfortunately, the maximum likelihood PCA method makes extreme computational demands. In addition, although we have individual estimates of uncertainty, they may not be good estimates. Typical ToF-SIMS spectrum image data sets, for instance, reflect low counting rates that, in turn, yield large relative uncertainties and highly sparse data matrices containing many zeros (>92% of the copper grid data elements are zero).

A different approach was presented by Cochran and Horne.[22] Their basic idea is to transform the data into an alternate space so that the uncertainty in the data is more uniform in that space. Standard PCA then can be performed in the alternate space to obtain a model that is optimal for describing the chemical information. Finally, the model can be back-transformed to give scores and loadings that can be interpreted in real, physical terms. Cochran and Horne showed that the optimal transformation is easily obtained by scaling the data when the estimated variance can be separated into the product of two factors: a row factor and a column factor

$$\mathrm{var}(d_{ij}) = g_i h_j \tag{6}$$

In this equation, g_i is a function assumed to be constant for all elements across the ith row and, likewise, h_j is a function that is assumed to be constant for all elements down the jth column. This expression can be written as the outer product

$$\mathrm{var}(\boldsymbol{D}) = \boldsymbol{g}\boldsymbol{h}^{\mathrm{T}} \tag{7}$$

Here, the vectors $\boldsymbol{g}$ and $\boldsymbol{h}$ are assumed to be column vectors. The optimally weighted data $\tilde{\boldsymbol{D}}$ are then given by

$$\tilde{\boldsymbol{D}} = (a\boldsymbol{G})^{-1/2}\boldsymbol{D}(b\boldsymbol{H})^{-1/2} \tag{8}$$

where a and b are arbitrary constants (they simply scale the eigenvalues) and $\boldsymbol{G}$ and $\boldsymbol{H}$ are diagonal matrices with

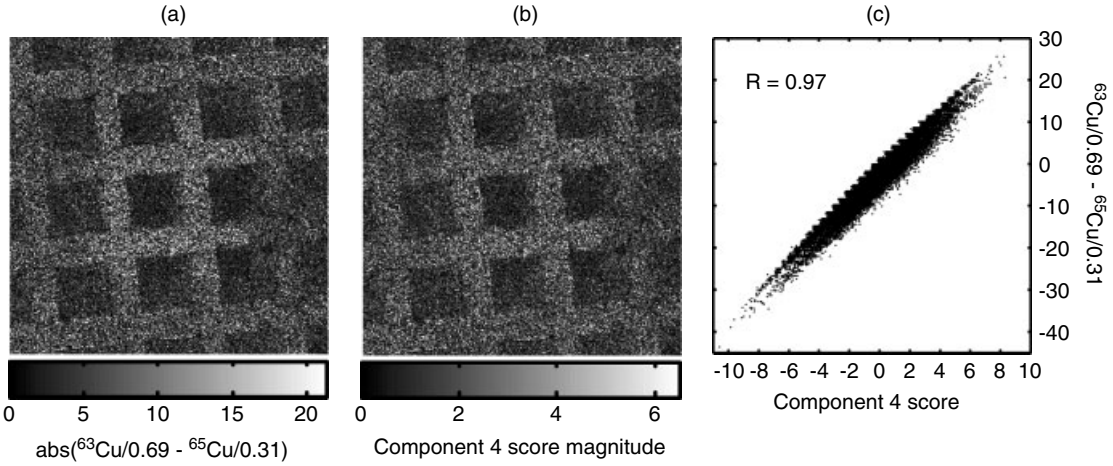

Figure 5. (a) Noise in the copper isotope ratio measured as the absolute difference between the abundance-normalized counts in mass channels 63 and 65. (b) Magnitude of the component 4 scores. (c) The pixel-by-pixel correlation between copper noise and the component 4 score. Some extreme values (<1% of the total number of pixels) were clipped in the images to improve contrast. All of the data is included in (c).

the elements of $\boldsymbol{g}$ and $\boldsymbol{h}$, respectively, along the diagonals. Performing standard PCA in the weighted space gives

$$\tilde{\boldsymbol{D}} = \tilde{\boldsymbol{T}}\tilde{\boldsymbol{P}}^{\mathrm{T}} \tag{9}$$

Equations (1), (8) and (9) can be combined to provide the transformations needed between the principal components in the weighted and physical spaces, namely

$$\begin{aligned} \boldsymbol{T} &= (a\boldsymbol{G})^{1/2}\tilde{\boldsymbol{T}} \\ \boldsymbol{P} &= (b\boldsymbol{H})^{1/2}\tilde{\boldsymbol{P}} \end{aligned} \tag{10}$$

Note that the term 'principal component' is being used somewhat loosely here. The columns of $\boldsymbol{T}$ and $\boldsymbol{P}$ will no longer be orthogonal after the back-transformation. However, $\boldsymbol{T}$ and $\boldsymbol{P}$ can be considered principal components in the sense that they provide a basis that maximally accounts for the chemically related variation in the data.

The foregoing arguments were cast in terms of an optimal data transformation. In fact, the weighted PCA solution itself can be shown to be optimal in a maximum likelihood sense. Assuming that individual uncertainties are described identically by the rank-one model in Eqn. (7), weighted PCA and maximum likelihood PCA yield identical fits. The equivalence of scaling (i.e. weighting with diagonal matrices) and maximum likelihood estimation given rank-one uncertainty estimates was pointed out recently[28] and we have verified the result using a published maximum likelihood PCA algorithm[21] on the copper grid data.

The weighted PCA procedure outlined above is straightforward. The primary difficulty resides in making the rank-one estimate of the data variance that is required in Eqn. (7). Fortunately, this estimation is considerably simplified in the case of Poisson statistics where Eqn. (5) holds. Assuming that counts in the spectral domain are statistically independent of counts in the spatial domain, and that row and column sums will be preserved, the best estimate of var($\boldsymbol{D}$) in a maximum likelihood sense can be achieved readily. This is accomplished by relating g_i and h_j to the probabilities of finding a count in the ith row and jth column, respectively. If p_i is the probability of finding a count in the ith row and q_j is the probability of finding a count in the jth column, then the expected value of var(d_{ij}) is simply

$$E[\mathrm{var}(d_{ij})] = E(d_{ij}) = p_i q_j d_{\cdot\cdot} \tag{11}$$

where E represents an expectation and a subscript dot indicates summation over the corresponding index (thus, $d_{\cdot\cdot}$ represents the total number of counts in $\boldsymbol{D}$). The requisite probabilities are readily computed from the row and column sums

$$\begin{aligned} p_i &= d_{i\cdot}/d_{\cdot\cdot} \\ q_j &= d_{\cdot j}/d_{\cdot\cdot} \end{aligned} \tag{12}$$

Then, letting $\boldsymbol{d}_m = [\, d_{1\cdot} \quad d_{2\cdot} \quad \cdots \quad d_{m\cdot}\,]^T$ be a vector whose elements are the m row sums and $\boldsymbol{d}_n = [\, d_{\cdot 1} \quad d_{\cdot 2} \quad \cdots \quad d_{\cdot n}\,]^T$ be the corresponding vector of the n column sums, the expectation of var($\boldsymbol{D}$) can be written

$$E[\mathrm{var}(\boldsymbol{D})] = \frac{1}{d_{\cdot\cdot}}\boldsymbol{d}_m\boldsymbol{d}_n^{\mathrm{T}} \tag{13}$$

Comparison with Eqn. (7) shows that $\boldsymbol{g}$ and $\boldsymbol{h}$ can be expressed as

$$\begin{aligned} \boldsymbol{g} &= \boldsymbol{d}_m/\sqrt{d_{\cdot\cdot}} \\ \boldsymbol{h} &= \boldsymbol{d}_n/\sqrt{d_{\cdot\cdot}} \end{aligned} \tag{14}$$

The weighting factors can be made more physically intuitive by setting $a = \sqrt{d_{\cdot\cdot}}/n$ and $b = \sqrt{d_{\cdot\cdot}}/m$ in Eqn. (8). With this choice, the weighting matrix $a\boldsymbol{G}$ is simply a diagonal matrix with the properly unfolded mean image along its diagonal, and the diagonal of the matrix $b\boldsymbol{H}$ consists of the mean spectrum.

It is worthwhile at this point noting the fundamental difference between the estimates of var($\boldsymbol{D}$) represented by Eqn. (13) and those computed directly via Eqn. (5). In the latter case, the uncertainties associated with each individual data element are assumed to be independent of one another. No use is made of the fact that correlations exist within the data. The rank-one estimate of variance, on the other hand, assumes that the variance, just like the data itself, can be described by a linear model. Through the statistical aggregation of the uncertainty estimates, as embodied in the mean spectrum and mean image, improved estimates of individual uncertainties are realized.

The superiority of the weighted PCA method for capturing chemical information and producing more interpretable results is readily demonstrated with the copper grid example. Figure 6 shows the eigenvalues obtained for the copper grid data set after weighting.

In contrast to the results from the unweighted analysis, the eigenvalue plot obtained here unambiguously shows the presence of four significant factors. The first five score images obtained by weighted PCA are shown in Fig. 7, and Fig. 8 displays the corresponding loading vectors. The first three components appear quite similar to the corresponding

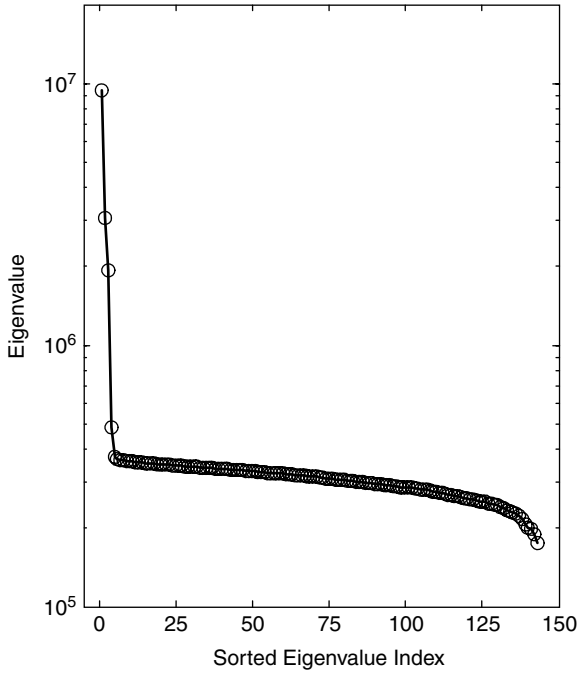

Figure 6. Sorted eigenvalues obtained by PCA of the weighted data matrix for the copper grid example.

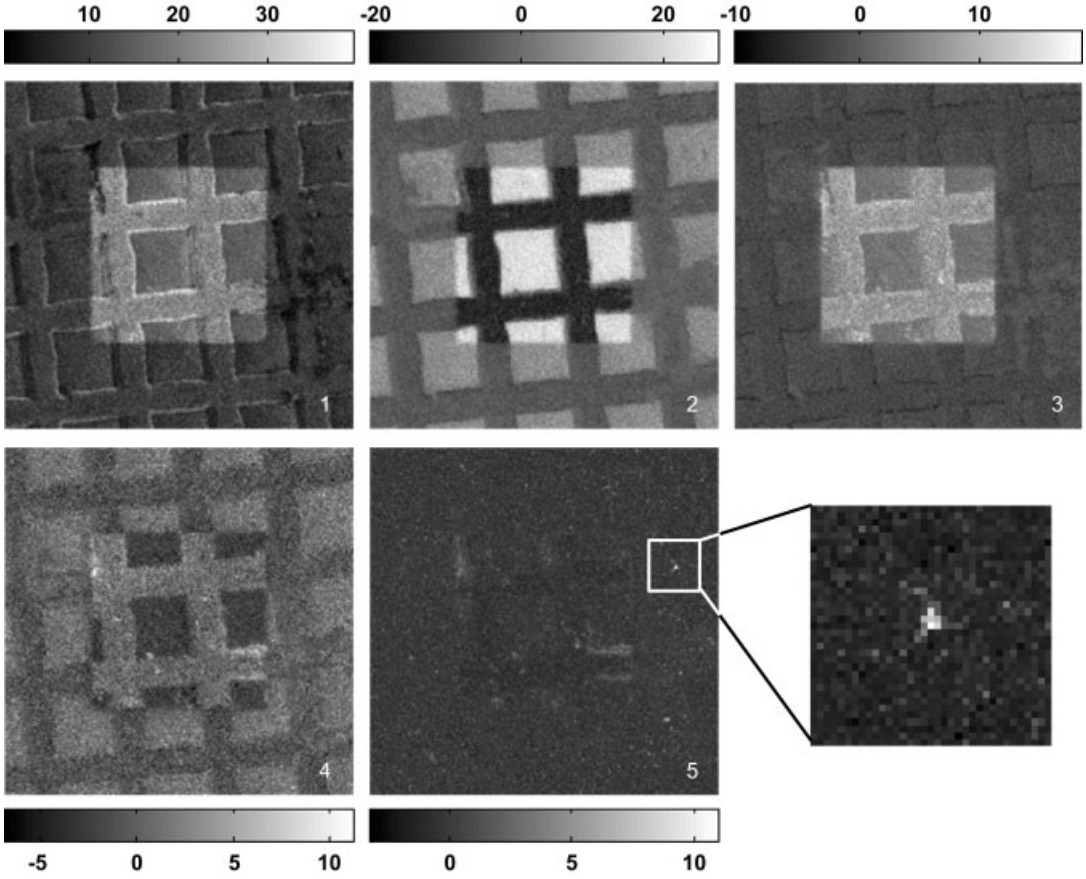

Figure 7. The first five score images obtained by weighted PCA of the copper grid data set.

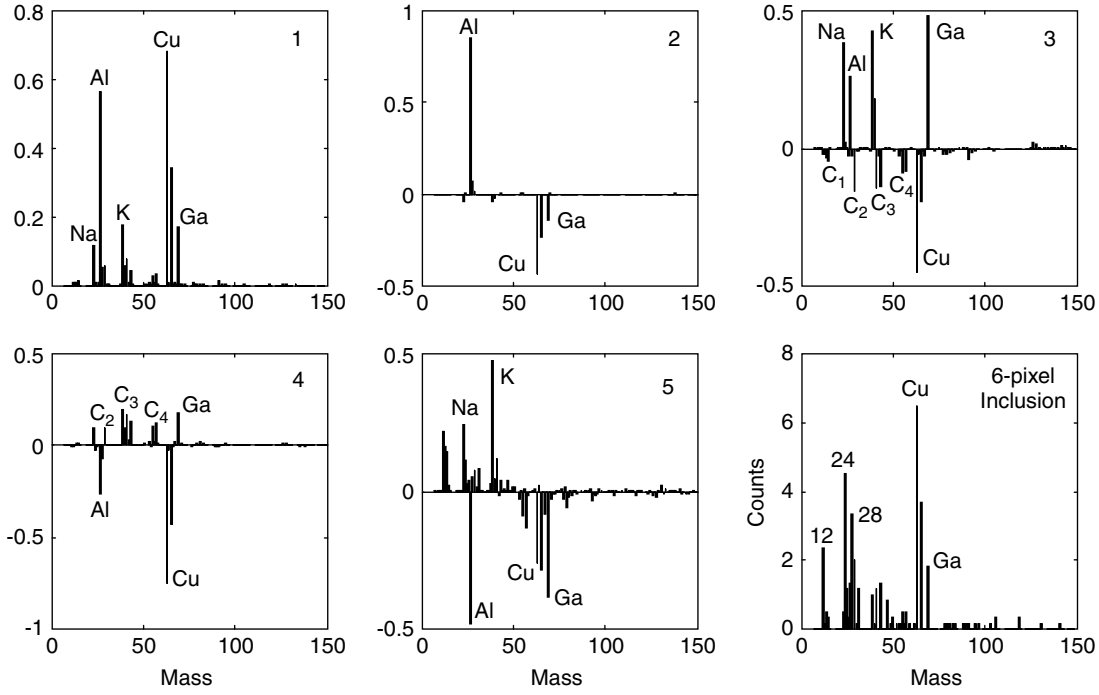

Figure 8. The first five loading vectors from the weighted PCA analysis of the copper grid example. The mean mass spectrum of the six-pixel inclusion identified by component 5 is also shown; C_x indicates a family of hydrocarbon fragments.

components found by the unweighted PCA analysis. The fourth component, however, rather than describing noise in the copper isotope ratio, is helping to describe the real hydrocarbon layer present in the non-sputtered region of the sample. A fifth component is also shown that appears to detect a small inclusion in the sample. A comparison of the actual spectra in the inclusion with the spectra of the surrounding uncleaned copper suggests that this is in fact a real feature containing several characteristic masses not found elsewhere in the image. This demonstrates the critical need for proper weighting of the data if one is seeking to solve the needle-in-a-haystack problem.

SIMULATION RESULTS

Several assertions regarding the relative performances of the weighted and unweighted PCA methods were made in the previous sections in the context of the copper grid example. A quantitative comparison of the two techniques, however, requires that 'truth' be known. For instance, it can be shown that unweighted PCA fits noisy experimental data more closely than does the weighted PCA method. To demonstrate that weighted PCA provides a better explanation of the real chemical information, on the other hand, presupposes that we know the precise chemistry of the sample. Although such precise knowledge is difficult to come by in a real sample, it is easy to incorporate into realistic simulations. In the present case, the copper grid data set was semi-simulated by constructing a noise-free version of the data from a weighted PCA model comprising the first five principal components. Noise was then added using a Poisson random number generator to simulate counting rates ranging from 2.5 to 250 counts per pixel, on average. The eigenvalues obtained by the weighted PCA method for the simulated data sets are shown in Fig. 9. At the higher counting rates, the presence of all five of the chemical components is clearly indicated. In these cases, the principal components derived are virtually indistinguishable from the corresponding components computed using the original data. At the lower counting rates, the signals arising from minor constituents become insignificant with respect to the noise and fewer non-noise principal components are detected. In the following comparisons, data reconstructions were made using the number of significant components appropriate to each noise level.

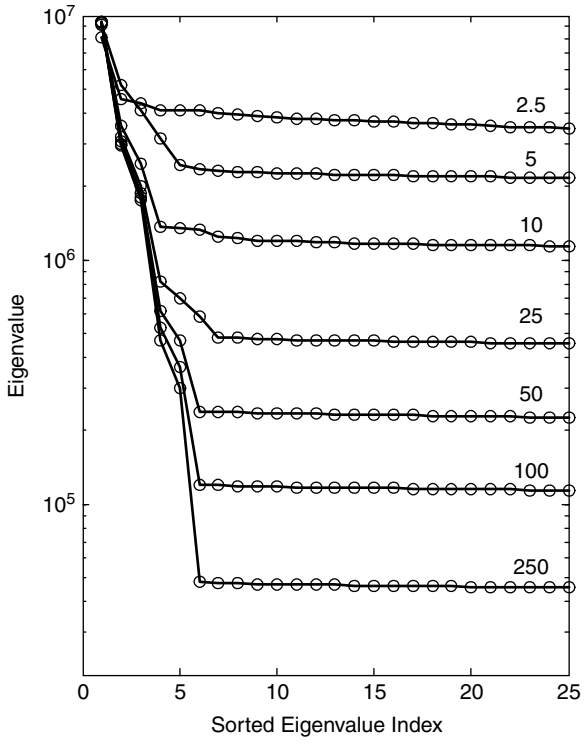

Figure 9. Eigenvalues obtained by the weighted PCA method for the simulated copper grid data at several counting rates. The numbers on the curves represent the average counts per pixel in the data set.

Two metrics were used to assess the performance of the weighted and unweighted PCA methods. The first, the relative sum of squared residuals RSSR, is defined by

$$\mathrm{RSSR} = \left\|\hat{\boldsymbol{D}} - \boldsymbol{D}\right\|_{\mathrm{F}}^2 \Big/ \left\|\boldsymbol{D}\right\|_{\mathrm{F}}^2 \tag{15}$$

and provides a measure of how closely the model fits the raw data. In this expression, $\left\| \bullet \right\|_{\mathrm{F}}^2$ represents the squared Frobenius norm of the matrix, which is computed simply as the sum of the squared elements of the matrix. The second metric is the relative sum of squared errors RSSE, which measures how the model deviates from the true values

$$\mathrm{RSSE} = \left\|\hat{\boldsymbol{D}} - \boldsymbol{D}_{\mathrm{True}}\right\|_{\mathrm{F}}^2 \Big/ \left\|\boldsymbol{D}_{\mathrm{True}}\right\|_{\mathrm{F}}^2 \tag{16}$$

The relative performances of the two PCA methods are compared in Fig. 10. Standard PCA provides a better fit to the raw data than does weighted PCA. This is in keeping with the well-known[2] characteristic of PCA that it produces, for a given number of components, the best possible representation of the data in the least-squares sense. Owing to the non-uniform noise, however, unweighted PCA does not provide the best fit to the underlying chemical model. Clearly, weighted PCA outperforms standard PCA when it comes to describing the real chemical content of the data.

As a final comparison, the ability of the weighted and standard PCA methods to detect the minor inclusion can be evaluated. Whether or not this inclusion is 'real' in the original data, it is truly real by construction in the simulations. Figure 11 shows the score images and portions of the loading vectors for the single principal component that best describes the inclusion. For this particular comparison, the 100-counts-per-pixel simulation was employed. The inclusion is clearly identified by both PCA methods. It is interesting to note, however, that while the weighted PCA method finds the inclusion to be the fifth most important component of the five known sources of chemical variation, the standard PCA method detects the inclusion in the 14th principal component. In other words, standard PCA finds at least nine components, which primarily represent noise, that are deemed more important than the real chemical inclusion for describing the overall variance in the data. It would seem highly unlikely that this component would be discovered in a typical standard PCA analysis.

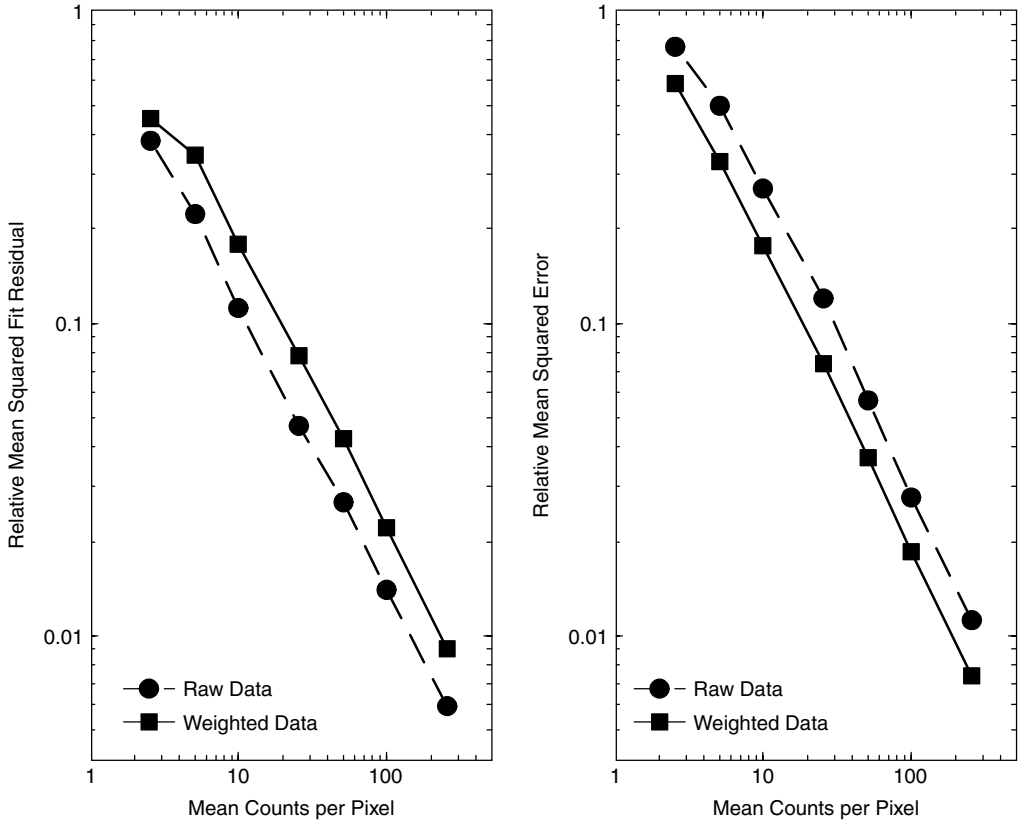

Figure 10. Comparison of the PCA model fit to the raw data (left) and model error (right) for the weighted and unweighted PCA methods.

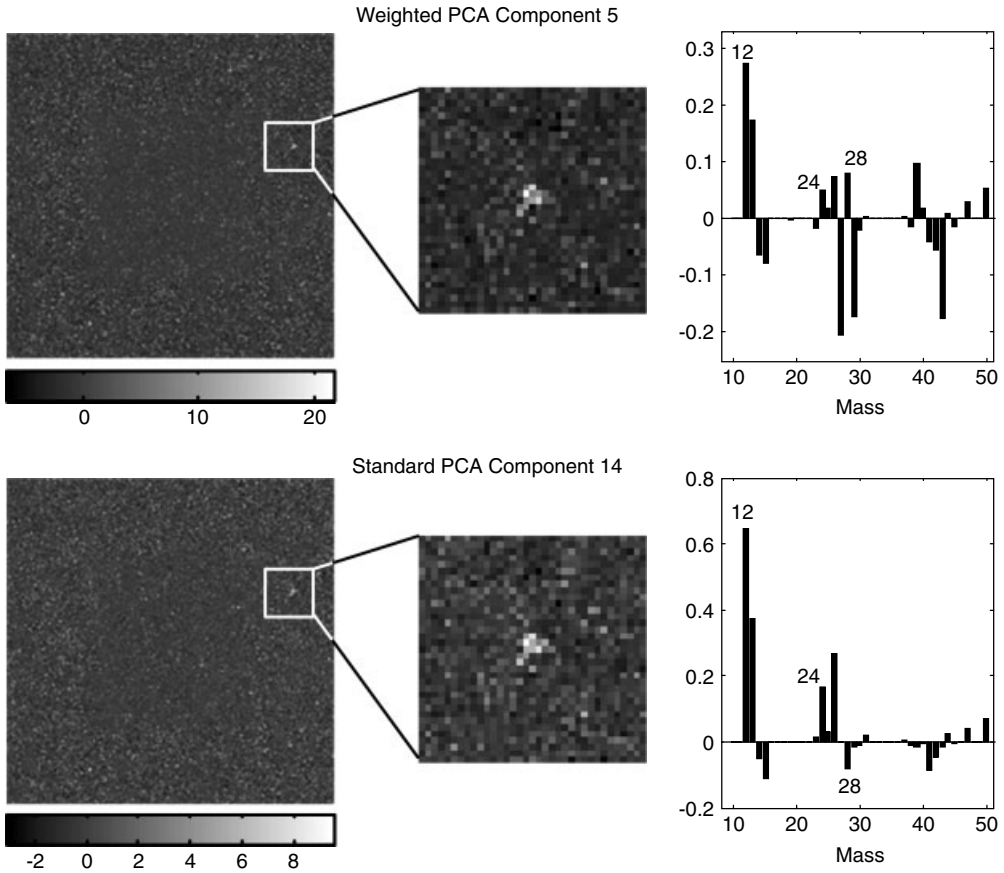

Figure 11. Score images and loading vectors for the single principal component that best describes the six-pixel inclusion in the copper grid data set simulated at 100 counts per pixel on average.

SUMMARY AND CONCLUSION

The statistical properties of spectroscopic data that arise from techniques that rely upon particle counting have been explored. The Poisson nature of such data ensures that its uncertainty is inherently non-uniform throughout the data set. This violates a fundamental assumption of standard multivariate techniques, such as PCA, which assume that the absolute magnitudes of the uncertainties are uniform and independent of the size of the respective individual data elements. A simple approach to weighting the data to make the uncertainties more uniform was presented. The optimal weighting under the constraints that the weighting matrices are diagonal and row and column sums are preserved during variance modeling was shown to involve simply the mean observation intensity (or mean image in the case of a spectrum image) and the mean spectrum. Using a simple and intuitive example, the clear superiority of the weighted PCA approach as compared to standard PCA was demonstrated. The weighted analysis provides eigenvalues that are more reliable for estimating the number of chemically important components. In addition, weighted PCA is more effective at segregating the chemically important information into the most significant principal components. Although ToF-SIMS data and PCA were used as exemplars of Poisson data and the multivariate statistical technique, respectively, the same ideas are directly applicable to other Poisson spectroscopies and multivariate analyses as well.

Acknowledgements

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-ACO4-94AL85000.

REFERENCES

1. NORAN System SIX, Thermo NORAN, 2551 West Beltline Highway, Middleton, WI, USA.
2. Jollife IT. *Principal Component Analysis* (2nd edn). Springer-Verlag: New York, 2002.
3. Malinowski ER. *Factor Analysis in Chemistry* (3rd edn). Wiley: New York, 2002.
4. Kargacin M, Kowalski B. *Anal. Chem.* 1986; **58**: 2300.
5. Trebbia P, Bonnet N. *Ultramicroscopy* 1990; **34**: 165.
6. Bonnet N, Brun N, Colliex C. *Ultramicroscopy* 1999; **77**: 97.
7. Cross B, *50th Annual Meeting of the Electron Microscopy Society of America*. Bailey GW, Bentley J, Small JA (eds). San Francisco Press: San Francisco, 1992; 1752.
8. Kessler T, Hoffmann P, Greve T, Ortner H. *X-Ray Spectrom.* 2002; **31**: 383.
9. Artyushkova K, Fulghum JE. *J. Electron Spectrosc. Relat. Phenom.* 2001; **121**: 33.
10. Artyushkova K, Fulghum J. *Surf. Interface Anal.* 2002; **33**: 185.
11. Kotula P, Keenan M, Michael J. *Microsc. Microanal.* 2003; **9**: 1.
12. Latkoczy C, Hutter H, Grasserbauer H, Wilhartitz P. *Mikrochim. Acta* 1995; **119**: 1.
13. Vanden Eynde X, Bertrand P. *Surf. Interface Anal.* 1997; **25**: 878.
14. Biesinger M, Paepegaey P, McIntyre N, Harbottle R, Petersent N. *Anal. Chem.* 2002; **74**: 5711.
15. Tyler B. *Appl. Surf. Sci.* 2003; **203**: 825.
16. Prutton M, Barkshire I, Kenny P, Roberts R, Wenham M. *Philos. Trans. R. Soc. London, Ser. A* 1996; **354**: 2683.
17. Bonnet N. *J. Microsc. (Oxford)* 1998; **190**: 2.
18. Prutton M, Wilkinson DK, Kenny PG, Mountain DL. *Appl. Surf. Sci.* 1999; **141/142**: 1.
19. Bonnet N. In *Advances in Imaging and Electron Physics*, vol. 114, Kazan B, Mulvey T, Hawkes P (eds). Academic Press: San Diego, 2000; 1.
20. Thompson W. *Comput. Sci. Eng.* 2001; **3**: 78.
21. Wentzell P, Andrews D, Hamilton D, Faber K, Kowalski B. *J. Chemomet.* 1997; **11**: 339.
22. Cochran RN, Horne FH. *Anal. Chem.* 1977; **49**: 846.
23. Green A, Berman M, Switzer P, Craig M. *IEEE Trans. Geosci. Remote Sens.* 1988; **26**: 65.
24. Lee J, Woodyatt S, Berman M. *IEEE Trans. Geosci. Remote Sens.* 1990; **28**: 295.
25. Physical Electronics Inc., 6509 Flying Cloud Drive, Eden Prarie, MN, USA.
26. The Mathworks, Inc., 3 Apple Hill Drive, Natick, MA, USA.
27. Geladi P, Grahn H. *Multivariate Image Analysis.* Wiley: Chichester, 1996.
28. Bro R, Smilde A. *J. Chemomet.* 2003; **17**: 16.
29. Golub GH, Van Loan CF. *Matrix Computations* (3rd edn). Johns Hopkins University Press: Baltimore, 1996.
30. Malinowski ER. *J. Chemomet.* 1987; **1**: 33.
31. Malinowski ER. *J. Chemomet.* 1988; **3**: 49.