Application of Maximum Likelihood Principal Components Regression to Fluorescence Emission Spectra

SUZANNE K. SCHREYER, MICHAEL BIDINOSTI, and PETER D. WENTZELL*

Trace Analysis Research Centre, Department of Chemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J3

The application of maximum likelihood multivariate calibration methods to the fluorescence emission spectra of mixtures of ace- 
naphthylene, naphthalene, and phenanthrene in acetonitrile is described. Maximum likelihood principal components regression (MLPCR) takes into account the measurement error structure in the spectral data in constructing the calibration model. Measure- 
ment errors for the fluorescence spectra are shown to exhibit both a heteroscedastic and correlated noise structure. MLPCR is com- 
pared with principal components regression (PCR) and partial least-squares regression (PLS). The application of MLPCR reduces the 
prediction errors by about a factor of two over PCR and PLS when a pooled estimate of the measurement error covariance ma- 
trix is employed. However, when only the heteroscedasticity is incor- 
thorated into MLPCR, no improvement in results is observed, in- 
dicating the importance of accounting for correlated measurement 
errors.

Index Headings: Multivariate calibration; Fluorescence; Principal components regression; Maximum likelihood; Measurement errors.

INTRODUCTION

Multivariate calibration techniques have been applied to a wide range of spectroscopic data for the determina- 
tion of individual analytes within mixtures. Typically, multi- 
channel measurements are obtained over a range of 
mixture compositions, with the level of each analyte in the 
calibration samples determined by a designed prepa- 
ration or an independent assay of real mixtures. Calibra- 
tion models are then built using any of a wide variety of 
methods. Some of the most common multivariate cali- 

drivation techniques employed in chemistry include mul- 
tilinear regression, classical least-squares regression, 
principal components regression (PCR), and partial least- 
squares regression (PLS). Each of these methods exhibits 
particular strengths and weaknesses, depending on the 
conditions under which it is used, but the most widely 
plied techniques for spectroscopic data are the latent 
variable methods, PCR and PLS. This is primarily be- 
cause of the capacity of these methods to deal with highly 
collinear data by reducing the dimensionality of the or- 
iginal data set.

In PCR, the latent variables are extracted in the form 
of scores by principal component analysis (PCA), which 
is closely related to singular value decomposition (SVD). 
This technique transforms the original data to a space of 
lower dimensionality, maximizing the amount of variance 
retained in the new space. PCA has been a valuable tech-
nique for dimensionality reduction, rank determina-
tion, and mixture analysis. A weakness of this model has been 
its inability to distinguish between measurement noise 

and systematic variance due to the chemical factors, es-

pecially if the measurement uncertainties at some chan-

nels are large. Furthermore, it is only when measurement 
uncertainties are independent and identically distributed 
(iid) with a normal distribution that a maximum likeli-

hood estimate of the subspace is obtained from PCA. In 
cases where measurement uncertainties are heteroscedas-
tic (nonuniform) and/or correlated with each other, the 
PCA solution will not be optimal in a maximum likeli-

hood sense. (In the context used here, “maximum like-

lihood” refers to the subspace model that gives rise to 
the highest probability density for the observed data 
based on prior assumptions of the dimensionality and 
measurement error distribution.)

In practice, spectroscopic measurements can exhibit 
both heteroscedastic and correlated measurement errors. 
For example, heteroscedastic noise can arise from count-

ing statistics (shot noise), nonlinear transformations (e.g., 
transmittance to absorbance), or variations in source in-
tensity with wavelength channel. Correlated measurement 
errors within a spectrum are also a common problem, 
lasting from conditions such as source intensity fluctua-
tions (flicker noise), spatial correlations of array detect-
ors, cell positioning errors, and electronic or digital 
smoothing of data. Correlated and heteroscedastic noise 
tends to be ignored in most multivariate calibration tech-
niques, leading to degradation of the quality of results 
obtained from the calibration model. Incorporating pre-
treatment techniques such as scaling and variable selec-
tion can substantially improve results, but these are gen-

erally suboptimal in the maximum likelihood sense.

In recent years, maximum likelihood principal com-

ponents analysis (MLPCA) and its regression counter-

part, MLPCR, have been shown to be effective multi-

variate data analysis techniques for dealing with cases of 
heteroscedastic and correlated measurement errors.1–5 
MLPCA and MLPCR are supersets of PCA and PCR that 
reduce to these more traditional methods if measurement 
errors are uncorrelated and uniform. Like PCA, MLPCA 
uses a least-squares approach to model data in a subspace 
of lower dimensionality, but incorporates measurement 
error information in the form of the error covariance ma-

trix. This results in a model that is optimal in a maximum 
likelihood sense, at least insofar as the measurement error 
structure is accurately known. It has been shown with 
both simulated and experimental data that MLPCA pro-
vides statistically better estimates of the data subspace 
except when the measurement error characteristics are 
poorly estimated. In terms of calibration, MLPCR has 
been shown to have significantly better predictive capa-

bility than PCR and PLS in spectroscopic applications 
where heteroscedastic and/or correlated measurement 

Received 17 July 2001; accepted 18 January 2002.

* Author to whom correspondence should be sent.
noise is present, including UV-visible absorption and near-infrared absorption and reflectance spectroscopy. By including the information in the measurement error covariance matrix, MLPCR can more easily separate chemical variance from other sources, resulting in a more reliable model.

The use of multivariate calibration with fluorescence emission spectroscopy has been extensively reported in the literature (see, for example, Refs. 6–12). A wide variety of other fluorescence techniques (excitation spectra, synchronous scanning, excitation-emission spectra) have also been employed in this context. In this paper, the application of MLPCR to fluorescence emission spectroscopy is investigated. Although this calibration method has proven effective for other spectroscopic techniques, fluorescence spectroscopy is substantially different in its signal and noise characteristics from the techniques studied so far. Depending on instrumental and chemical characteristics, measurement precision may be limited by shot noise, source flicker noise, and/or blank noise. The first two of these can lead to heteroscedastic noise characteristics, while the last two can contribute to correlated noise. In contrast, other techniques are typically dominated by other factors such as detector noise (IR), scatter (NIR), or the effects of transformation (UV-visible). Therefore, it would be useful to examine to what extent improvements can be realized by the application of MLPCR to fluorescence data.

**THEORY**

**Principal Components Regression.** Conventional PCR begins with the application of PCA using the singular value decomposition (SVD) algorithm on the spectra obtained for calibration samples. This leads to the decomposition:

\[ X = USV^T \]  

(1)

where \( X \) is the matrix of spectra with \( m \) samples and \( n \) wavelength channels. Assuming \( m < n \), the SVD gives the matrices \( U \) \((m \times m)\), \( S \) \((m \times m)\), and \( V \) \((n \times m)\). These matrices are then truncated to rank \( p \), where \( p \) is typically determined by cross-validation. The resulting truncation gives:

\[ \tilde{X} = \tilde{U}\tilde{S}\tilde{V}^T \]  

(2)

where \( \tilde{U} \) \((m \times p)\), \( \tilde{S} \) \((p \times p)\), and \( \tilde{V} \) \((n \times p)\) are the truncated matrices. Assuming that \( p \) is the true (error free) rank of \( X \), this will represent the optimal estimation of the \( p \)-dimensional model in a maximum likelihood sense if the measurement errors in \( X \) are iid normal. The scores matrix, \( \tilde{T} \), is given by:

\[ \tilde{T} = \tilde{U}\tilde{S} \]  

(3)

Given an \( m \times 1 \) vector of concentrations or properties, \( y \), independently measured for the calibration samples, the regression model on the scores is given by:

\[ y = \tilde{T}\hat{q} + e \]  

(4)

where \( \hat{q} \) is the \( p \times 1 \) estimate of the regression vector for the scores and \( e \) is a vector of residuals. The regression vector is estimated by:

\[ \hat{q} = (\tilde{T}^T\tilde{T})^{-1}\tilde{T}^Ty = \tilde{S}^{-1}\tilde{U}^Ty \]  

(5)

For a matrix of spectra of unknown samples, \( X_{unk} \), the concentration can be estimated by first making an orthogonal projection of the spectra into the PCA subspace to obtain the scores and then applying the regression model:

\[ \tilde{T}_{unk} = X_{unk}\tilde{V} \]  

(6)

\[ \hat{y}_{unk} = \tilde{T}_{unk}\hat{q} \]  

(7)

Alternatively, this can be represented by the more familiar form:

\[ \hat{y}_{unk} = X_{unk}\tilde{V}\hat{q} = X_{unk}\tilde{b} \]  

(8)

where \( \tilde{b} = \tilde{V}\tilde{S}^{-1}\tilde{U}^Ty \) is the \( n \times 1 \) regression vector in the original space. Because of the greater convenience of the direct application of Eq. 8, it is the commonly represented form. However, it has been implicitly assumed here that all of the measurement errors are iid normal. If this assumption is not valid, then the estimation of the subspace will not be optimal in a maximum likelihood sense, and an alternative model, which incorporates information about the spectral measurement errors in the estimation of the subspace, is required. Under these circumstances, it is not always possible to obtain a universal regression vector in the original wavelength space because the calculation of scores requires a non-orthogonal projection.

**Maximum Likelihood PCR.** The motivation for the application of MLPCR in calibration problems is the need to more effectively handle the wide variety of error structures that can exist in experimental data. In essence, the application of this technique is the same as for PCR except for two important differences: (1) MLPCA, rather than PCA, is employed in the decomposition step, and (2) a maximum likelihood projection, rather than an orthogonal projection into the PCA subspace, is used in the prediction step. The theoretical aspects of MLPCA and MLPCR have been addressed in the literature, so only a brief description will be given here.

A core element of the MLPCA algorithm is an accurate description of the heteroscedasticity and covariance of the measurement noise. This is usually expressed in terms of a measurement error covariance matrix. Various implementations of the MLPCA algorithm are available, depending on the complexity of the error structure. In the most general form, a matrix \( X \) \((m \times n)\) may have measurement error correlations amongst all of its elements. Under these circumstances, it would be necessary to describe \( mn(mn-1)/2 \) covariance terms and \( mn \) variance terms. Although the solution to this case has been theoretically demonstrated, the computational difficulties in this situation make it impractical to implement in most circumstances. For calibration problems, it is generally safe to take the next step towards simplification, which is to assume that there are no measurement error correlations among samples (rows). Under these circumstances, the error structure for each row of \( X \) can be described by its own error covariance matrix, which is defined as:

\[ \Sigma_i = E(e_i^T e_i) = E[(x_r - \bar{x}_i)^T(x_r - \bar{x}_i)] \]  

(9)

In this equation, \( \Sigma_i \) \((n \times n)\) is the measurement error covariance matrix for row \( i \) of the matrix \( X \), \( E \) denotes the expectation operator, and \( e_i \) is the residual error vector
for row i of X, which is defined as the difference between the measured row vector, \(x_i\), and the true (unobservable) row vector, \(\hat{x}_i\). Normally the error covariance matrix is estimated from the residuals of replicate samples around the mean or from an assumed error structure in the data. Once this estimate is available, MLPCA is carried out by minimizing the objective function, \(S^2\), given by,

\[
S^2 = \sum_{i=1}^{n} (x_i - \hat{x}_i \Sigma^{-1}(x_i - \hat{x}_i))^T
\tag{10}
\]

where,

\[
\hat{x}_i = x_i \Sigma^{-1} \tilde{V}(\tilde{V}^T \Sigma^{-1} \tilde{V})^{-1} \tilde{V}^T
\tag{11}
\]

The minimization is carried out on the truncated loadings matrix, \(\tilde{V} (n \times p)\), subject to the usual constraints of orthonormality. The minimization uses an alternating least-squares procedure. Following convergence, the final MLPCA estimates of \(\hat{U}, \hat{S}, \text{and} \tilde{V}\) are obtained by SVD on \(\tilde{X}\) and truncation of the solution to rank \(p\). Assuming that the estimates of the error covariance matrices are accurate, this procedure produces a maximum likelihood estimate of the PCA subspace. Even in cases where the error covariance estimates are only approximate, MLPCA has generally been shown to provide superior results when heteroscedastic/correlated noise is prevalent.

Once the MLPCA decomposition has been carried out, regression is carried out in the scores space as before with the use of Eq. 5. The prediction step is slightly modified, since the unknown spectrum must first be projected into the scores space using a maximum likelihood projection rather than an orthogonal projection.

\[
\tilde{i}_{\text{unk}} = x_{\text{unk}} \Sigma_{\text{unk}}^{-1} \tilde{V}(\tilde{V}^T \Sigma_{\text{unk}}^{-1} \tilde{V})^{-1} \tilde{V}^T
\tag{12}
\]

Prediction can then be carried out with Eq. 7. Note that in this fairly generalized case it is not possible to derive a regression vector, \(\tilde{b}\), in the original wavelength domain because each unknown sample may have a different error covariance matrix, leading to a different projection angle.

Further simplifications of this approach are possible given appropriate error structures. For example, if the measurements exhibit heteroscedasticity but no correlations, the error covariance matrices will consist only of diagonal elements containing the variances of each measurement. While this does not change the implementation strategy, it does simplify the inversion of the error covariance matrix, an operation which is slightly problematic when the matrix is singular.\(^4^5\)

Another simplification is to assume that all of the samples (rows) have the same error covariance matrix. While this may not be strictly true in a particular application, it is often a reasonable approximation where spectra are similar, especially since each row covariance is typically calculated with a small number of replicates and therefore prone to substantial error. Adopting this assumption has three significant advantages. First, error covariance estimates from each sample can be pooled to obtain a more reliable overall estimate, which in turn improves the results from MLPCA. Second, there is an algorithmic advantage in that the MLPCA solution does not have to be obtained iteratively. Brown et al.\(^5\) demonstrated that the MLPCA solution in this case could be obtained by a rotation and scaling of the original space followed by PCA, with the characteristics of the rotation matrix determined by the error covariance matrix. Finally, if it is assumed that unknown samples express the same error covariance matrix, a regression vector in the wavelength domain can be obtained by:

\[
\tilde{b} = \Sigma^{-1} \tilde{V}(\tilde{V}^T \Sigma^{-1} \tilde{V})^{-1} \tilde{S}^{-1} \hat{U}^T y
\tag{13}
\]

Continuing to build on this simplification, one could further assume that there was no error correlation, but only heteroscedasticity, which is only a function of wavelength (i.e., not a function of the sample). In this case, the MLPCA estimates are obtained by a simple scaling of the data. This can be extended to the trivial case of uniform errors, at which point the MLPCA reduces to PCA.

In this work, the calibration was carried out assuming equal error covariance matrices for all of the rows. In addition, the assumption of uncorrelated errors (i.e., heteroscedasticity only) was evaluated by performing the calibration using only the diagonal of the pooled error covariance matrix. This was done to assess the relative roles of heteroscedasticity and correlated noise in contributing to a suboptimal calibration.

**EXPERIMENTAL**

**Procedure.** Fluorescence emission spectra were obtained from mixtures of three polycyclic aromatic hydrocarbons (PAHs): acenaphthylene (ace), naphthalene (nap), and phenanthrene (phe). Five replicate sets of spectra were obtained from each of 27 mixtures. A three-level, three-factor factorial design was used to prepare the mixtures and a randomized order was employed for each of the five replicate blocks. A blank containing only the solvent (acetonitrile) was run before and after each block.

**Reagents and Samples.** Naphthalene (Fisher) was used as received. Acenaphthylene (Aldrich) and phenanthrene (BDH) were recrystallized prior to use. Stock solutions of the individual samples were prepared by mass in acetonitrile (Anachemia, spectrophotometric grade, 99.9%). The final concentration ranges were approximately 0.10 to 0.34 µg/g (ace), 0.018 to 0.063 µg/g (nap), and 0.0072 to 0.027 µg/g (phe).

**Instrumentation.** Fluorescence emission spectra were obtained from samples in a 1-cm quartz cuvette on a Shimadzu RF-301PC spectrofluorometer with a xenon lamp excitation source. The excitation wavelength was 278 nm and the emission wavelength was scanned between 310 and 460 nm in steps of 2 nm. A medium scan speed was used and the slit width for both excitation and emission was set at 3.0 nm.

**Computational Aspects.** Data analysis was carried out using in-house algorithms written in MatLab v. 6.0 (The MathWorks, Inc., Natick, MA) on a Pentium-based personal computer. PLS analysis was performed using the PLS Toolbox v. 2.1 (Eigenvector Research, Inc., Manson, WA).

**RESULTS AND DISCUSSION**

Representative spectra of three of the mixture samples (extracted from the 100, 010, and 001 corners of the factorial design) are shown in Fig. 1. The pure component
spectra in the mixtures are sufficiently overlapped that differentiation of the individual components is not possible. Thus, multivariate calibration analysis would be advantageous for the simultaneous determination of the individual components.

Component concentrations were selected to give a noise level of ~1% of the maximum signal, which was observable but not excessive. Error covariance matrices were estimated for each set of sample replicates by Eq. 14:

\[
\Sigma_i = \sum_{k=1}^{N} (x_i(k) - \bar{x}_i)(x_i(k) - \bar{x}_i)/(N-1)
\]  

(14)

In this equation \(x_i(k)\) \((1 \times n)\) is the measured spectrum for replicate \(k\) of sample \(i\) and \(\bar{x}_i\) is the mean spectrum of the \(N\) replicates for that sample \((N = 5\) in this case). After examination to verify sufficient similarity, these individual sample error covariance matrices were pooled using Eq. 15 to give an overall estimate that was applied to all spectra.

\[
\Sigma_{avg} = \frac{\sum_{i=1}^{m} \Sigma_i}{m}
\]  

(15)

Figure 2 shows the average covariance matrix in the form of contour and mesh plots. One of the features which is immediately apparent is that the magnitudes of the variance and covariance elements are directly related to the corresponding spectral intensities. This is more apparent in Fig. 3, which shows the pooled standard deviations at each channel (i.e., the square root of the diagonal of \(\Sigma_{avg}\)) plotted with the mean spectrum. With the exception of the offset in the standard deviations, the match between the two curves is very close. This type of behavior in the variance (and hence the covariance) is not unexpected since shot noise exhibits a square-root dependence on the signal intensity and flicker noise is directly proportional to the signal intensity. However, another feature that is prominent in the mesh plot of Fig. 2 is the positive offset everywhere. This indicates that all channels have a significant positive correlation. Such behavior is also anticipated, since this is the characteristic expected with blank noise or related effects, such as cell positioning errors. The fact that heteroscedasticity and correlated noise are both substantial contributors to the error structure in this case is a fortunate outcome of this experiment, since it allows the impact of both features to be examined.

Four calibration methods were evaluated for these data: PCR, PLS, MLPCR using only the diagonal of the error covariance matrix (MLPCR - diag), and MLPCR using the full error covariance matrix (MLPCR - cov). As noted before, the purpose of comparing the last two methods was to determine whether improvements brought about by MLPCR could be attributed to heteroscedasticity alone. To compare these methods in a statistically meaningful way, a resampling experiment was carried out. The fluorescence data were divided into a calibration set, consisting of three of the five replicates for each sample, and a prediction set, which was made up of the remaining two replicates. For each sample, the replicates used for calibration and prediction were randomly selected to ensure no block correlation effects. The regression model was then built using the calibration data with each of the four regression methods already described, and the quality of the model was evaluated using the usual metrics, namely the root mean squared error of calibration (RMSEC) and the root mean squared error of prediction (RMSEP). This process was repeated ten times using a new randomization for each trial. (While this type of partitioning is useful here, it is not recommended for routine model building.)

To determine the number of latent variables to be used in each of the models, leave-one-out cross-validation was used. In this approach, the calibration model is construct-
Fig. 2. Contour (top) and mesh plot (bottom) representations of the pooled measurement error covariance matrix for the fluorescence spectra.
ed using all but one sample in the calibration set, which is then predicted with the model. This is repeated for all 27 samples and the root mean squared error of cross-validation (RMSECV) is calculated and plotted against the number of latent variables (LVs) used in the model. Typical cross-validation results for PCR and MLPCR (covariance included) are shown in Fig. 4. For phenanthrene and naphthalene, four LVs appear optimal, while for acenaphthylene good prediction seems to require only three. Although one might expect an optimum of three latent variables in all cases given that this is a three component mixture, other factors such as a baseline offset may increase this number. The results for all calibration methods and all trials were essentially the same, so for simplicity, four latent variables were used for all components in all cases in the results reported here.

The results of this study are summarized in Fig. 5, which provides both the RMSEC and RMSEP for each calibration method. Error bars represent one standard deviation of the ten resampling trials. In all cases, the RMSEC and RMSEP are comparable, as expected for a valid model. The figure also shows that PCR and PLS give virtually identical results for all three analytes. This is typical for systems with a well-defined rank and indicates that PLS, despite its perceived advantages over PCR, does not have any special ability to handle heteroscedastic and correlated measurement errors. It is clear from the figure that MLPCR using the pooled error covariance matrix gives an improvement of about a factor of two in the prediction error over PCR and PLS. This corresponds to errors of 1.6, 0.97, and 0.56% relative to the maximum concentrations for acenaphthylene, naphthalene, and phenanthrene, respectively. This factor of improvement is similar to that observed for visible absorbance spectroscopy in another application, but not as good as that observed for a near-infrared reflectance application, where improvements as much as a factor of four were observed. Interestingly, the use of MLPCR with only the diagonal of the covariance matrix to account for heteroscedascity without correlation does not produce any improvement over PCR or PLS. This suggests that covariance is a more important factor than var-

**Fig. 3.** Comparison of measurement standard deviation (A) and mean fluorescence spectrum (B).

**Fig. 4.** Typical results for cross-validation of calibration data by PCR (top) and MLPCR (bottom).
Fig. 5. Calibration and prediction results for acenaphthylene (A), naphthalene (B), and phenanthrene (C) by various calibration methods.

Conclusions

In spite of the fact that analytical instruments in use today rarely exhibit noise characteristics that are iid normal, most calibration methods employed implicitly assume such a structure. Through the use of MLPCR in the fluorescence application presented here, it has once again been demonstrated that superior performance can be achieved by taking the measurement noise characteristics into account through the error covariance matrix. Furthermore, it was shown that, in this case, incorporation of heteroscedasticity alone did not improve on the results obtained by PCR and PLS. This suggests that the simple scaling used in some applications to account for heteroscedasticity may not be optimal when correlated noise is present.

Maximum likelihood principal components regression provides a general way to incorporate a variety of error structures into the calibration problem. Computationally, the use of MLPCR with a pooled error covariance matrix increases the computational burden in the calibration step by less than a factor of two, so this aspect should not detract from its use. Perhaps the biggest deterrent is the need for replicate measurements to estimate the error covariance matrix. Although this requires additional samples to be run for a given design, it is obvious from this work that extracting the optimal amount of information from a measurement also requires a knowledge of its noise characteristics, so these additional experiments may be warranted. Furthermore, it may be possible to develop models for these noise characteristics on particular instruments, removing the need to obtain replicate measurements for every new application.
The improvement in predictive ability with MLPCR in this case was about a factor of two, a result which is consistent with other applications that have been reported.\textsuperscript{2,4,5} For applications such as this, where the prediction errors are around 1%, this improvement may not appear to be that significant, but given that many multivariate calibration applications, such as those employed for biological or medical samples, push the limits of detectability, such improvements can have a profound impact and justify a closer examination of the role of measurement uncertainty in multivariate techniques.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council (NSERC) and The Dow Chemical Company (Midland, MI).