*Anal. Chem.* **1997,** *69,* 2299−2311

# Maximum Likelihood Multivariate Calibration

Peter D. Wentzell* and Darren T. Andrews

*Trace Analysis Research Centre, Department of Chemistry, Dalhousie University, Halifax, NS, Canada B3H 4J3*

Bruce R. Kowalski

*Center for Process Analytical Chemistry, University of Washington, Box 351700, Seattle, Washington 98195-1700*

**Two new approaches to multivariate calibration are described that, for the first time, allow information on measurement uncertainties to be included in the calibration process in a statistically meaningful way. The new methods, referred to as maximum likelihood principal components regression (MLPCR) and maximum likelihood latent root regression (MLLRR), are based on principles of maximum likelihood parameter estimation. MLPCR and MLLRR are generalizations of principal components regression (PCR), which has been widely used in chemistry, and latent root regression (LRR), which has been virtually ignored in this field. Both of the new methods are based on decomposition of the calibration data matrix by maximum likelihood principal component analysis (MLPCA), which has been recently described (Wentzell, P. D.; et al. *J. Chemom.*, in press). By using estimates of the measurement error variance, MLPCR and MLLRR are able to extract the optimum amount of information from each measurement and, thereby, exhibit superior performance over conventional multivariate calibration methods such as PCR and partial least-squares regression (PLS) when there is a nonuniform error structure. The new techniques reduce to PCR and LRR when assumptions of uniform noise are valid. Comparisons of MLPCR, MLLRR, PCR, and PLS are carried out using simulated and experimental data sets consisting of three-component mixtures. In all cases of nonuniform errors examined, the predictive ability of the maximum likelihood methods is superior to that of PCR and PLS, with PLS performing somewhat better than PCR. MLLRR generally performed better than MLPCR, but in most cases the improvement was marginal. The differences between PCR and MLPCR are elucidated by examining the multivariate sensitivity of the two methods.**

Over the past several decades, advances in chemometrics have led to the development of a multitude of multivariate calibration methods for the analysis of chemical mixtures.[1−3] As a result, such methods are now routinely applied and are indispensable tools for solving many "real-world" problems. At times, the proliferation of multivariate calibration techniques seems unending and includes such methods as multiple linear regression (MLR), principal components regression (PCR), partial least-squares regression (PLS), continuum regression (CR), projection pursuit regression (PPR), locally weighted regression (LWR), and artificial neural networks (ANNs), among others. Each of these methods possesses its own strengths and weaknesses, and which works best for a given problem depends on the characteristics of the data and the objectives of the analysis. However, as research produces a clearer distillation of the similarities and differences among methods, a number of techniques, such as PLS and PCR, have established themselves as the practical workhorses of multivariate calibration. PCR is one of the oldest and most well-studied methods currently in use, and this paper describes two fundamental enhancements to the methodology involved which will extend its utility and reliability even further. Although the techniques described in this work are general in their applicability, the focus will be on spectroscopic data sets.

Traditional univariate calibration, which assumes no interference with the measured response variable, typically applies weighted or unweighted least-squares regression to a series of standards to develop the calibration model. Under the right conditions, the model developed in this manner will be optimal in a maximum likelihood sense. Maximum likelihood parameter estimation methods are widely used because of their desirable statistical characteristics.[4] In the present context, maximum likelihood estimation means that the parameters determined for the model are the ones most likely to give rise to the observed data, given the statistical characteristics of the measurement uncertainties. The conditions necessary for ordinary least-squares (weighted or unweighted, as appropriate) to provide maximum likelihood parameter estimates are (1) the form of the model needs to be correct (e.g., straight line, quadratic, intercept if necessary), (2) the measurement uncertainties in the response variable need to be uncorrelated and normally distributed and (in the case of weighted regression) have known variances, and (3) the measurement uncertainties in the concentrations ($x$ variable) need to be negligibly small relative to the uncertainties in the response variable ($y$). In practice, these ideal conditions are rarely met exactly, but maximum likelihood methods are often still regarded as the best alternative if the conditions are approximately valid.

In contrast to traditional univariate calibration, techniques such as PCR are known as *inverse* calibration methods because the concentrations are regressed on the responses (factor scores in PCR) rather than the other way around. Accordingly, PCR can only qualify as a maximum likelihood method if the uncertainties in the responses (scores) are negligible compared to those in the

(1) Thomas, E. V. *Anal. Chem.* **1994**, *66*, 795A−804A.
(2) Kowalski, B. R.; Seasholtz, M. B. *J. Chemom.* **1991**, *5*, 129−145.
(3) Beebe, K. R.; Kowalski, B. R. *Anal. Chem.* **1987**, *59*, 1007A−1017A.
(4) Larsen, R. J.; Marx, M. L. *An Introduction to Mathematical Statistics and Its Applications*, 2nd ed.; Prentice-Hall: Englewood Cliffs, NJ, 1986.

S0003-2700(96)01029-3 CCC: $14.00 

concentrations. While this is often true when the reference method for concentrations is relatively imprecise, there are many cases where the assumption is somewhat tenuous. Furthermore, PCR ignores the uncertainty in the spectroscopic data when it performs the initial decomposition by principal component analysis (PCA). As pointed out in earlier work,[5] PCA yields a maximum likelihood decomposition only when the measurement uncertainties are independent and identically distributed with an normal distribution ("*iid* normal"). It has long been known that spectroscopic measurements inherently possess nonuniform measurement standard deviations which can vary as a function of both signal amplitude and wavelength. Causes of such nonuniformity include variations in the source intensity profile, quantum effects (i.e., shot noise), nonlinear transformations (e.g., for absorbance measurements), and variations in detector noise characteristics. Furthermore, instrument characteristics, such as a finite spectral band-pass and source flicker, often lead to correlated noise characteristics.

Although the noise characteristics for most common spectroscopic methods have been well-studied,[6] this information is generally ignored in establishing multivariate calibration models. It should be apparent that, since each spectral measurement can possess a different uncertainty, each can also carry a different amount of information into the calibration procedure. In PCR, for example, PCA is first used to determine the subspace of the component spectra of a mixture. The spectrum of each calibration sample is then projected into this subspace to give a set of scores, or latent variables. These scores are used in the regression procedure to produce the PCR calibration model. This projection has the effect of combining the spectral measurements to reduce the overall error and also makes the regression step more mathematically tractable. Obviously, the quality of results obtained by PCR will depend on the quality of the estimation of the spectral subspace by PCA. Unfortunately, PCA tries to maximize the variance accounted for by the extracted latent variables, regardless of whether the variance is due to chemical effects (i.e., changes in chemical concentrations) or simply measurement uncertainty. Because of this, including measurements with a large uncertainty can degrade the quality of the calibration model developed by PCR. While this problem has been addressed informally through approaches such as scaling and wavelength selection, these pretreatments are generally suboptimal in a maximum likelihood sense.

In this paper, two new methods are described to account for measurement uncertainty in multivariate calibration. These methods are based on a recently described matrix decomposition procedure called maximum likelihood principal component analysis (MLPCA)[5] and will be referred to as maximum likelihood principal components regression (MLPCR) and maximum likelihood latent root regression (MLLRR). The new techniques are actually more general forms of PCR and latent root regression (LRR) and will produce solutions identical to those given by these methods under the right conditions. However, the new methods are better suited to providing optimal solutions in the maximum likelihood sense when there are nonuniform uncertainties in the data. It will be shown using both simulated and experimental data that MLPCR and MLLRR can provide significantly better predictive ability than conventional methods in realistic situations. Perhaps more importantly, the maximum likelihood methods provide a general unifying framework from which multivariate calibration methods can be examined.

Throughout this work, a number of assumptions and simplifications have been made that should be clarified from the outset. First, it has been assumed that measurement errors are normally distributed. While the principles of maximum likelihood estimation are general in nature, mathematical tractability in the development of MLPCA demanded that this restriction be imposed. Although this assumption may not be strictly valid in all cases, it is viewed as reasonable and, unless the violation is severe, should not greatly diminish the general utility of the methods, just as simple regression is often used without strict adherence of the underlying assumptions. A second assumption made by MLPCA for maximum likelihood estimation (and by weighted regression, for that matter) is that measurement error variances are exactly known. In practice, however, this is rarely the case, so true maximum likelihood estimates are technically unattainable for real experimental data. Nevertheless, it will be shown that variance estimates are sufficient to obtain significant improvement in results; i.e., that some knowledge of measurement uncertainty is often better than an implicit assumption of uniform variance. Finally, throughout this work, it has been assumed that measurement errors are uncorrelated; i.e., the error covariance matrix is diagonal. While such a condition can be controlled in simulations, it is almost certainly invalid for experimental measurements. It is demonstrated, however, that significant improvement in predictive ability can be achieved even when the assumption of uncorrelated errors is tenuous. There are two main reasons for excluding error covariance in this work. First, while estimates of measurement error variance are often available, knowledge of the covariance matrix in practice is still quite rare, so we wished to demonstrate the utility of these methods when the covariance matrix is unavailable. Second, although the theory of MLPCA is capable of dealing with correlated errors,[5] there are several practical problems that need to be addressed. These include rank deficiency in the estimated error covariance matrix and the computational burden of large matrices. These subjects are beyond the scope of the present work and will be dealt with in a subsequent paper.

## THEORY

**Principal Components Regression.** For the purposes of this discussion, it will be assumed that we are trying to develop a calibration model for a single analyte in the presence of multiple unknown interferences, and that the measurements consist of spectroscopic data (although other analytical techniques could also be employed). Conventional PCR begins with a set of calibration samples for which the concentration of the analyte has been obtained by some independent means. The first step in the procedure is to apply PCA to the spectra of the calibration samples. This is usually done by way of singular value decomposition (SVD) to give

$$\mathbf{X} = \mathbf{USV}^{\mathrm{T}} \tag{1}$$

Here, $\mathbf{X}$ is the matrix of spectra in the calibration set ($m$ samples by $n$ wavelengths). The component concentrations in the calibration set should reflect the distribution of those concentrations for

(5) Wentzell, P. D.; Andrews, D. T.; Hamilton, D. C.; Faber, K.; Kowalski, B. R. *J. Chemom.*, in press.
(6) Ingle, J. D; Crouch, S. R. *Spectrochemical Analysis*; Prentice-Hall: Englewood Cliffs, NJ, 1988.

future samples (i.e., the calibration set should span the space of samples to be predicted), and the number of samples and wavelength channels should be greater than the number of independently observable components in the mixtures. Assuming that $m < n$, the SVD gives the matrices $\mathbf{U}$ ($m \times m$), $\mathbf{S}$ ($m \times m$) and $\mathbf{V}$ ($n \times m$). These matrices are truncated by removing the right-hand columns and bottom rows to give $\tilde{\mathbf{U}}$ ($m \times p$), $\tilde{\mathbf{S}}$ ($p \times p$), and $\tilde{\mathbf{V}}$ ($n \times p$), where $p$ is the "pseudorank" of $\mathbf{X}$, or the number of independently observable components. In practice, $p$ is usually unknown, but can be estimated by statistical means or cross-validation. In this work, the tilde ("~") will be used to distinguish the truncated matrices and the quantities derived from them. The truncation gives $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^{\mathrm{T}} = \tilde{\mathbf{T}}\tilde{\mathbf{V}}^{\mathrm{T}}$, where $\tilde{\mathbf{X}}$ is the estimated data matrix and $\tilde{\mathbf{T}} = \tilde{\mathbf{U}}\tilde{\mathbf{S}}$ is called the scores matrix for the truncated solution. Alternatively, in a model and parameters framework, we have

$$\mathbf{X} = \tilde{\mathbf{T}}\tilde{\mathbf{V}}^{\mathrm{T}} + \mathbf{E} \tag{2}$$

where $\mathbf{E}$ is the $m \times n$ matrix of residuals. PCA obtains $\tilde{\mathbf{T}}$ and $\tilde{\mathbf{V}}$ by minimizing the sum of the squares of the elements in $\mathbf{E}$. This estimation is optimal in a maximum likelihood sense as long as $p$ represents the true pseudorank and the measurement errors for the elements of $\mathbf{X}$ are *iid* normal.

The reduction in the dimensionality of the problem by PCA is the key to PCR, since it improves the reliability of the solution. The actual regression is carried out using orthogonal projections of the spectra onto the subspace determined by PCA, i.e., the scores. The regression assumes a model of the form

$$\mathbf{y} = \tilde{\mathbf{T}}\mathbf{q} + \mathbf{f} \tag{3}$$

where $\mathbf{y}$ is the $m \times 1$ vector of analyte concentrations for the calibration set, $\mathbf{q}$ is a $p \times 1$ regression vector, and $\mathbf{f}$ is an $m \times 1$ vector of errors. The least-squares solution to this problem is

$$\hat{\mathbf{q}} = (\tilde{\mathbf{T}}^{\mathrm{T}}\tilde{\mathbf{T}})^{-1}\tilde{\mathbf{T}}^{\mathrm{T}}\mathbf{y} = \tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^{\mathrm{T}}\mathbf{y} \tag{4}$$

In this equation and elsewhere, the caret ("^") is used to indicate an estimated quantity. In the prediction step, the scores for the unknown spectrum are given by

$$\tilde{\mathbf{t}}_{\mathrm{unk}} = \mathbf{x}_{\mathrm{unk}}\tilde{\mathbf{V}} \tag{5}$$

where $\tilde{\mathbf{t}}_{\mathrm{unk}}$ and $\mathbf{x}_{\mathrm{unk}}$ are row vectors of length $p$ and $n$, respectively. The unknown concentration is then estimated by

$$\tilde{y}_{\mathrm{unk}} = \tilde{\mathbf{t}}_{\mathrm{unk}}\hat{\mathbf{q}} \tag{6}$$

More typically, the intermediate step of calculating the scores is incorporated into an $n \times 1$ regression vector, $\hat{\mathbf{b}}$, that is multiplied directly by the spectrum to obtain the concentration estimate

$$\tilde{y}_{\mathrm{unk}} = \tilde{\mathbf{x}}_{\mathrm{unk}}\hat{\mathbf{b}} \tag{7}$$

$$\hat{\mathbf{b}} = \tilde{\mathbf{V}}\hat{\mathbf{q}} = \tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^{\mathrm{T}}\mathbf{y} \tag{8}$$

In conventional PCR, the representations in eqs 6 and 7 are equivalent, but this is not the case for MLPCR, as discussed in the next section.

**Maximum Likelihood PCR.** When applied in the proper context, conventional PCR is a powerful tool for the quantitative analysis of multicomponent mixtures. However, it suffers from a number of weaknesses. One of these is that it relies on SVD to obtain a reliable estimation of the $p$-dimensional subspace that contains the component spectra. In essence, the eigenvectors produced by SVD (the columns of $\tilde{\mathbf{V}}$) describe a $p$-dimensional hyperplane in the $n$-dimensional wavelength space and should contain all of the pure spectral vectors. As long as the measurement errors in all of the calibration spectra are all *iid* normal, the $p$-dimensional hyperplane determined by SVD will be an optimal model for the data in a maximum likelihood sense (assuming the system is linear and the pseudorank, $p$, has been correctly specified). However, if the measurement errors are not independent with uniform variance, this will no longer be true, and the estimation of the subspace will be suboptimal.

There are several potential solutions to this problem. First, it may be possible to scale the data in such a way that all of the measurement standard deviations become equal. It has been shown, however, that in order for this to work in a manner which preserves the structure of the data, the matrix of measurement standard deviations must have a rank of unity[7] (e.g., when the uncertainty at each wavelength is independent of signal amplitude). This restriction is frequently violated for experimental data sets, making it impossible to obtain an optimum solution through simple scaling. A second approach to the problem is to perform wavelength selection, removing those channels that significantly violate the assumption of *iid* errors. This assumes, however, that noise is a function only of wavelength and not signal amplitude. Furthermore, although a portion of the spectrum may appear noisy, it may also be the region which is richest in information about the analyte of interest. Wavelength selection has also been performed by using leave-one-out cross-validation. In addition to being very time consuming, this approach only mitigates the problem of finding the optimal subspace and does not address the source of the problem.

What is needed is a modeling method which accounts for spectral measurement errors in the estimation of the spectral subspace. Such a method, called maximum likelihood principal component analysis (MLPCA), was recently introduced[5] and is the basis of MLPCR. In MLPCA, the eigenvectors are chosen to provide the optimal estimation of the $p$-dimensional hyperplane in a maximum likelihood sense. The optimality of the estimation is, strictly speaking, contingent on the assumption of normally distributed measurement errors with known variances and covariances, but relaxation of these conditions (i.e., near normality and/or estimated variances) still yields significantly improved estimates of the PCA subspace. For uncorrelated measurement errors, MLPCA minimizes a weighted sum of squared residuals:

$$S^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(x_{ij} - \hat{x}_{ij})^2}{\sigma_{ij}^2} \tag{9}$$

In this equation, $x_{ij}$ is a measurement (an element of $\mathbf{X}$), $\hat{x}_{ij}$ is the maximum likelihood estimate of that measurement, and $\sigma_{ij}$ is the

(7) Paatero, P.; Tapper U. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 183−194.

corresponding measurement error standard deviation. (In practice, $\sigma_{ij}$ is typically replaced by its estimate, $s_{ij}$.) The MLPCA decomposition can be represented as

$$\mathbf{X} = \check{\mathbf{U}}\check{\mathbf{S}}\check{\mathbf{V}}^{\mathrm{T}} + \mathbf{E} = \check{\mathbf{T}}\check{\mathbf{V}}^{\mathrm{T}} + \mathbf{E} = \check{\mathbf{X}} + \mathbf{E} \tag{10}$$

where $\mathbf{X}$, $\check{\mathbf{X}}$, and $\mathbf{E}$ are $m \times n$, $\check{\mathbf{U}}$ and $\check{\mathbf{T}}$ are $m \times p$, $\check{\mathbf{S}}$ is $pxp$, and $\check{\mathbf{V}}$ is $n \times p$ for a $p$-dimensional model. The symbol "$\check{}$" above the matrix variables has been used here to distinguish the MLPCA solution from the truncated PCA solution. Although similar in their objectives, MLPCA and conventional PCA have some very significant differences.[5] In particular, while PCA requires *iid* normal measurement errors, the general MLPCA algorithm requires neither uniform variance nor independence to provide the maximum likelihood estimates. It is also important to note that, unlike PCA, MLPCA solutions are not nested; that is, the rank $p$ model cannot be obtained simply by truncating higher rank models. Instead, the dimensionality of the model needs to be specified before initiating the decomposition. Although this tends to make MLPCA more cumbersome to use, the superior results often make it worthwhile. Another difference is that, in conventional PCA, the estimate for any $1 \times n$ spectral vector, $\mathbf{x}_i$, is given by an orthogonal projection into the spectral subspace:

$$\hat{\mathbf{x}}_i = \mathbf{x}_i\tilde{\mathbf{V}}\tilde{\mathbf{V}}^{\mathrm{T}} \tag{11}$$

In contrast, the maximum likelihood estimate of $\mathbf{x}_i$ is given by a projection which is not generally orthogonal, but rather one which is weighted by the errors in the measurements:

$$\hat{\mathbf{x}}_i = \mathbf{x}_i\Sigma_i^{-1}\check{\mathbf{V}}(\check{\mathbf{V}}^{\mathrm{T}}\Sigma_i^{-1}\check{\mathbf{V}})^{-1}\check{\mathbf{V}}^{\mathrm{T}} \tag{12}$$

Here, $\Sigma_i$ is the $n \times n$ covariance matrix for $\mathbf{x}_i$ (note that any multiple of $\Sigma_i$ could also be used). For uncorrelated errors, this will be a diagonal matrix whose diagonal elements are the variances for the corresponding spectral measurements. It is clear that eq 12 will result in an orthogonal projection when all of the measurement errors are uncorrelated and have equal variances; i.e., the PCA projection is equivalent to a maximum likelihood projection under these conditions. For measurement errors which are correlated only within a spectrum (i.e., row correlations but no column correlations), eq 12 is still valid, but the function minimized by MLPCA is modified to

$$S^2 = \sum_{i=1}^{m}(\mathbf{x}_i - \hat{\mathbf{x}}_i)\Sigma_i^{-1}(\mathbf{x}_i - \hat{\mathbf{x}}_i)^{\mathrm{T}} \tag{13}$$

which reduces to eq 9 for uncorrelated errors. If measurement errors are correlated among both rows and columns, a somewhat modified version of MLPCA is needed. This is described elsewhere[5] and will not be treated here except to note that MLPCA can handle any measurement error covariance structure, provided that the error covariance matrix can be estimated. As noted in the introduction, uncorrelated measurement errors have been assumed throughout this work. Although this assumption is not generally valid for experimental data, the error covariance structure is rarely known in practical situations, so it is intended to reflect a realistic implementation of the methods described. The theoretical and practical aspects of dealing with correlated errors have been examined and will be presented elsewhere.

The regression model in MLPCR is developed in a manner analogous to that in PCR. Following from eq 4,

$$\hat{\mathbf{q}} = (\check{\mathbf{T}}^{\mathrm{T}}\check{\mathbf{T}})^{-1}\check{\mathbf{T}}^{\mathrm{T}}\mathbf{y} = \check{\mathbf{S}}^{-1}\check{\mathbf{U}}^{\mathrm{T}}\mathbf{y} \tag{14}$$

However, unlike conventional PCR, a maximum likelihood projection is used to determine the scores for the unknown sample, which are then used to estimate the concentration:

$$\hat{\mathbf{t}}_{\mathrm{unk}} = \mathbf{x}_{\mathrm{unk}}\Sigma_{\mathrm{unk}}{}^{-1}\check{\mathbf{V}}(\check{\mathbf{V}}^{\mathrm{T}}\Sigma_{\mathrm{unk}}{}^{-1}\check{\mathbf{V}})^{-1} \tag{15}$$

$$\hat{y}_{\mathrm{unk}} = \hat{\mathbf{t}}_{\mathrm{unk}}\hat{\mathbf{q}} \tag{16}$$

Note that, in MLPCR, there is no longer an analog to a universal regression vector, **b**, for all unknown samples, as defined in eqs 7 and 8 for PCR. This is because the projection matrix depends on the measurement error covariance matrix, which can be different for each unknown sample. This, however, is one of the main advantages of MLPCR, since the projection of the unknown sample onto the spectral subspace will exploit those measurements that have the smallest errors in order to obtain the best estimate of the scores.

To summarize, MLPCR improves the quality of the regression over PCR in two ways. First, it uses MLPCA in conjunction with measurement error information to obtain a more reliable estimate of the subspace containing the pure spectral vectors. Because measurements in the calibration set are appropriately weighted, a maximum likelihood estimate of the PCA model is obtained which is generally superior to that obtained by SVD. This is important because it is the determination of this initial space that ultimately affects the sensitivity of the calibration procedure. The second advantage of MLPCR derives from the projection of the measurements (calibration and unknown) onto the subspace determined by MLPCA. Because the projection is not orthogonal but rather optimized through the use of measurement uncertainties, the maximum information is extracted for the best estimation of the true measurements. These factors tend to lead to superior calibration models.

**Maximum Likelihood Latent Root Regression.** Although MLPCR can offer a significant improvement over conventional PCR, it is still not a "pure" maximum likelihood approach to calibration because of the final regression step. For this step to be optimal from a maximum likelihood perspective, the absolute uncertainties in the scores need to be much smaller than those in the concentration values. Since this will not always be true, it would be useful to develop a method which could accommodate an arbitrary error in the final regression step. This can be done by incorporating a variation of latent root regression (LRR).

Unlike PCR, LRR[8−10] is not well-known among chemists. With this technique, the original calibration matrix of response variables

(8) Montgomery, D. C.; Peck, E. A. *Introduction to Linear Regression Analysis*; Wiley: New York, 1982; p 339.
(9) Sanchez, E.; Kowalski, B. R. *J. Chemom.* **1988**, *2*, 247−263.
(10) Vigneau, E.; Bertrand, D.; Qannari, E. M. *Chemom. Intell. Lab. Syst.* **1996**, *35*, 231−238.

is augmented by the corresponding concentration vector(s). PCA is then carried out on the augmented matrix. In this way, the reduction of dimensionality and the determination of the calibration model are performed simultaneously. Using the previous example for the estimation of the concentration of a single component, we have

$$[\tilde{\mathbf{X}}|\tilde{\mathbf{y}}] = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^{\mathrm{T}} \tag{17}$$

As before, $\mathbf{X}$ is a matrix of $m$ spectra measured at $n$ wavelengths, $\mathbf{y}$ is an $m \times 1$ vector of concentrations for the component of interest, and the tilde indicates that the SVD results are truncated to pseudorank $p$. If $\tilde{\mathbf{V}}$ (which now has dimensions $(n + 1) \times p$) is partitioned into the upper $\tilde{\mathbf{V}}_1$ ($n \times p$) and lower $\tilde{\mathbf{V}}_2$ ($1 \times p$), then the regression vector will be given by

$$\hat{\mathbf{b}} = \tilde{\mathbf{V}}_1(\tilde{\mathbf{V}}_1{}^{\mathrm{T}}\tilde{\mathbf{V}}_1)^{-1}\tilde{\mathbf{V}}_2{}^{\mathrm{T}} \tag{18}$$

such that the predicted concentration is $y_{\mathrm{unk}} = \mathbf{x}_{\mathrm{unk}}\hat{\mathbf{b}}$, where $\mathbf{x}_{\mathrm{unk}}$ is $1 \times n$.

LRR is similar to PCR in its approach to calibration, but for some reason it has been virtually ignored in chemistry. It is possible that it is simply more cumbersome and less intuitive than PCR, and in many cases it does not offer significant advantages. Another difference between the two methods is that the predictive ability of PCR is unaffected by changes in the scale of the $y$ variable. This is because the regression step in PCR implicitly assumes (for the maximum likelihood solution) that all of the error resides in $y$, so a vertical projection is always used. In contrast, LRR is consistent with a maximum likelihood solution if the absolute uncertainties in all of the measured quantities ($x$ and $y$) are the same (*iid* normal), leading to results that will change with the scale of $y$. The situation is exactly analogous to the differences between ordinary least-squares and PCA when used for modeling purposes.[11] In reality, neither set of assumptions is likely to be valid. It would, therefore, be useful to have a single-step modeling procedure like LRR which accounts for all of the uncertainties. Such a method is presented here as maximum likelihood latent root regression (MLLRR).

The procedure for MLLRR is similar to that for MLPCR, except that, as in LRR, an augmented matrix is used. In a manner analogous to eq 17, the augmented matrix is decomposed using MLPCA rather than PCA. This requires a companion matrix of measurement variances, also augmented to include the variances in the concentration values. In the absence of other information, measurement uncertainties are usually assumed to be uncorrelated, but correlated errors can be accommodated by MLPCA as well. Once MLPCA has been carried out, prediction is performed using an augmented spectral vector:

$$[\hat{\mathbf{x}}_{\mathrm{unk}}|\hat{y}_{\mathrm{unk}}] = [\mathbf{x}_{\mathrm{unk}}|0]\Sigma_{\mathrm{unk}}{}^{-1}\check{\mathbf{V}}(\check{\mathbf{V}}^{\mathrm{T}}\Sigma_{\mathrm{unk}}{}^{-1}\check{\mathbf{V}})^{-1}\check{\mathbf{V}}^{\mathrm{T}} \tag{19}$$

In this case, $\Sigma_{\mathrm{unk}}$ is the error covariance matrix of the augmented row vector for the unknown, and $\check{\mathbf{V}}$ is the loadings matrix obtained from applying MLPCA to the augmented calibration matrix. The equation is written so that it produces a row vector, since this is the manner in which the spectra appear in the original calibration matrix. Note that eq 19 is simply a maximum likelihood projection of the unknown spectrum into the MLPCA subspace. The key is that, since $y_{\mathrm{unk}}$ is the quantity sought, the last entry in the error covariance matrix, $\Sigma_{\mathrm{unk}}$, is set to be numerically equivalent to infinity, forcing this value to be predicted from the others. Thus, the last entry in the first row vector on the right-hand side is unimportant and is set to zero in the equation. Extension of eq 19 to additional components is easily accomplished by further augmentation of the calibration and prediction matrices. As with MLPCR, there is no universal regression vector for MLLRR unless the covariance matrices are identical for all of the spectra obtained.

MLLRR is more general in its treatment of measurement errors than MLPCR in that it includes uncertainties in the concentration values. It is an optimal modeling method in the maximum likelihood sense, subject to the usual restrictions (linear model of known pseudorank, normally distributed errors with a known covariance structure).

## EXPERIMENTAL SECTION

**Data Sets.** To examine the two new methods proposed here, three simulated and two experimental data sets were employed. The simulated data sets were used to test the methods under carefully controlled conditions to evaluate their potential. Each of these was generated from a model of a three-component mixture. The pure component spectra consisted of Gaussian profiles centered at 480, 500, and 520 nm (for components 1, 2, and 3, respectively) with widths ($\sigma$) of 20 nm. Spectral data points were generated at 5 nm intervals between 400 and 600 nm. Calibration and prediction data sets consisted of 20 and 100 samples, respectively, whose component concentrations were generated randomly from a uniform distribution between 0 and 1. In all simulations, normally distributed measurement errors were added using a Gaussian random number generator.

Data set 1 was characterized by wavelength-dependent noise which was essentially uniform near the center of the spectral range but amplified on either side. To accomplish this, a baseline noise level of $\sigma_0$ was first selected. This standard deviation was then multiplied by a wavelength-dependent function to give the standard deviation for a particular wavelength. The function used in this work was a "double-sigmoidal" mask, with a value close to unity near the center of the spectral region and values of $r_{\mathrm{max}}$ at the limits. The profile of this mask is shown along with the individual spectral profiles in Figure 1. Using this mask, the standard deviation at wavelength $\lambda$ is given by

$$\sigma(\lambda) = \left[1 + (r_{\mathrm{max}} - 1)\left(\frac{1}{1 + \mathrm{e}^{a(\lambda - \lambda_1)}} + \frac{1}{1 + \mathrm{e}^{a(\lambda_2 - \lambda)}}\right)\right]\sigma_0 \tag{20}$$

In this equation, $\lambda_1$ is the inflection point of the sigmoid on the left-hand side of the range and $\lambda_2$ is that on the right-hand side. The parameter $a$ determines the slope of the sigmoidal curves such that

$$a = \frac{2 \ln 9}{\Delta\lambda} = \frac{4.394}{\Delta\lambda} \tag{21}$$

where $\Delta\lambda$ is the 10%−90% rise range of the sigmoid. In this work, $\lambda_1 = 460$ nm, $\lambda_2 = 540$ nm, and $\Delta\lambda = 40$ nm. The standard

(11) Andrews, D. T.; Chen, L.; Wentzell, P. D.; Hamilton, D. C. *Chemom. Intell. Lab. Syst.* **1996**, *34*, 231−244.

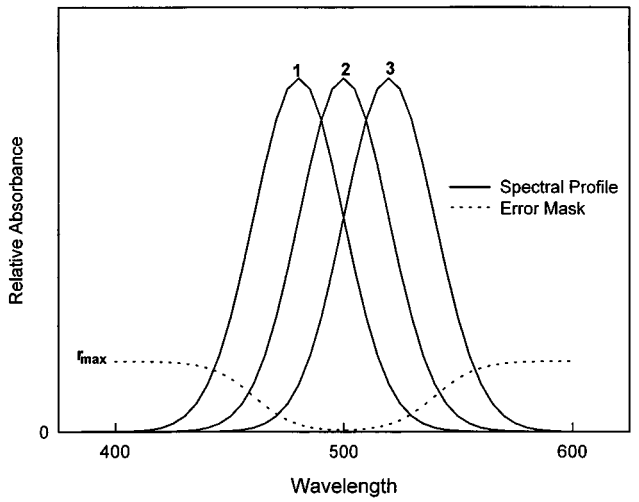


Figure 1. Spectral profiles for simulated three-component mixtures (data sets 1, 2, and 3) and error mask for data set 1.

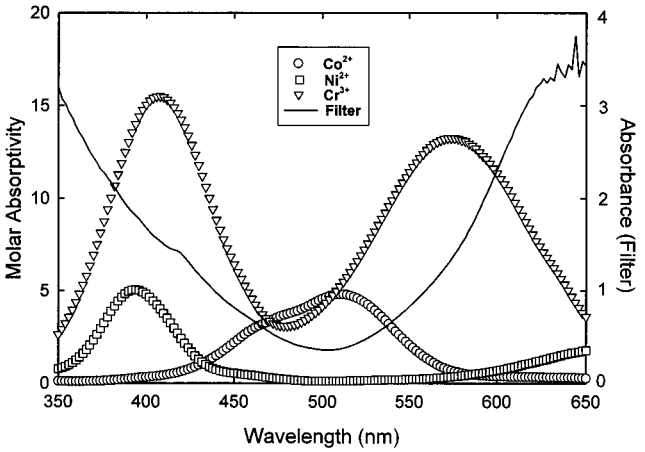


Figure 2. Pure component spectra for data set 4 and absorbance profile of band-pass filter applied to source beam for noise amplification.

deviation of the baseline noise level, $\sigma_0$, was taken to be 1% of the maximum absorbance in the noise-free calibration data. The noise amplification factor, $r_{max}$, was varied between 1 and 20 for this work. The concentration data used for calibration with data set 1 were assumed to be error free.

Data set 2 was the same as data set 1 except for the noise structure. In this case, both proportional and constant error were added to the signals to give measurement uncertainties that depended on signal amplitude. The formula used to calculate the standard deviation for a given absorbance, $A_{ij}$, was

$$\sigma_{ij} = \sqrt{\sigma_0^2 + (pA_{ij})^2} \qquad (22)$$

where $\sigma_0$ is the level of the constant noise component (in this case, 1% of the maximum signal in the calibration matrix), and $p$ is the level of proportional noise (varied between 0 and 0.20 in this work).

Data set 3, which was intended to exaggerate the differences between MLLRR and MLPCR, included errors in the calibration concentrations in addition to those in the spectral measurements. As for data set 2, the errors in the spectral measurements included both a constant term (1% of the maximum in the calibration set) and a proportional term (in this case fixed at a level of 2% of the pure signal). To simulate nonuniform errors in the calibration concentrations, proportional error was added to the reference concentrations. The proportional error had standard deviations that ranged from 0 to 20% of the true concentration.

To demonstrate the utility of the maximum likelihood calibration methods for practical applications, two experimental data sets were also examined. Data set 4, the first of these, was obtained through a carefully designed experiment involving three-component mixtures of metal ions (Co(II), Cr(III), and Ni(II)), a system suggested from the work of Osten and Kowalski.[12] Stock solutions of the nitrates were prepared with concentrations of 0.172, 0.0764, and 0.393 M for Co, Cr, and Ni, respectively, in 4% $HNO_3$. All chemicals used throughout this work were analytical reagent grade or better unless otherwise specified. A three-level, three-factor calibration design was used in which 1, 3, or 5 mL aliquots of the various stock solutions were combined and diluted to 25 mL with 4% $HNO_3$. Unfortunately, insufficient Ni stock remained for one solution (3:5:5 Co:Cr:Ni), so the calibration set consisted of 26 rather than 27 solutions. Final concentration ranges were 6.88−34.40 mM for Co, 3.06−15.29 mM for Cr, and 15.70−78.58 mM for Ni. Five replicate spectra were obtained for each sample using randomized blocks (i.e., five blocks of all 26 solutions, randomly ordered within each block). To minimize the effects of instrument drift, a reference spectrum was run prior to each new sample. Spectra were recorded over the range of 350−650 nm on an HP 8452 diode array spectrophotometer (Hewlett-Packard, Palo Alto, CA) using a standard 1 cm quartz cuvette. Measurements were made at 2 nm intervals with a 1 s integration time. In order to introduce nonuniform noise characteristics, a dichroic band-pass filter (green, no. 67) was placed between the source and the sample to decrease the source intensity at high and low wavelengths for all measurements. The spectra of the individual components and the optical filter are shown in Figure 2.

The second experimental data set employed in this work, data set 5, consisted of near-infrared spectra for three-component mixtures containing toluene, chlorobenzene, and heptane. The mixtures were prepared as part of a calibration transfer study by Scott Specialty Gases (Houston, TX) and consisted of 31 samples from an augmented three-level, three-factor factorial design. The concentrations varied between 20 and 70 wt % for toluene and chlorobenzene and between 2 and 10 wt % for heptane. The mixtures were sealed into standard 1 cm cuvettes, and spectra were obtained over the range 400−2500 nm on an NIRSystems grating spectrometer (NIRSystems, Silver Spring, MD) at intervals of 2 nm. The spectrometer employed a Si detector below 1100 nm and a PbS detector at longer wavelengths. Figure 3 shows a typical spectrum over the full range and standard deviations obtained from replicate scans. It is apparent that certain regions of the spectrum above 1600 nm are effectively opaque and, therefore, of little utility for analysis. Consequently, standard deviations in this region are high. The purpose of this data set was to demonstrate that MLPCR can utilize all of the available data to obtain superior predictive ability by extracting the optimum amount of information at each wavelength, provided measurement variance information is available. Unfortunately, standard deviation information for this data set was only available from replicate

(12) Osten, D. W.; Kowalski, B. R. *Anal. Chem.* **1985**, *57*, 908−915.

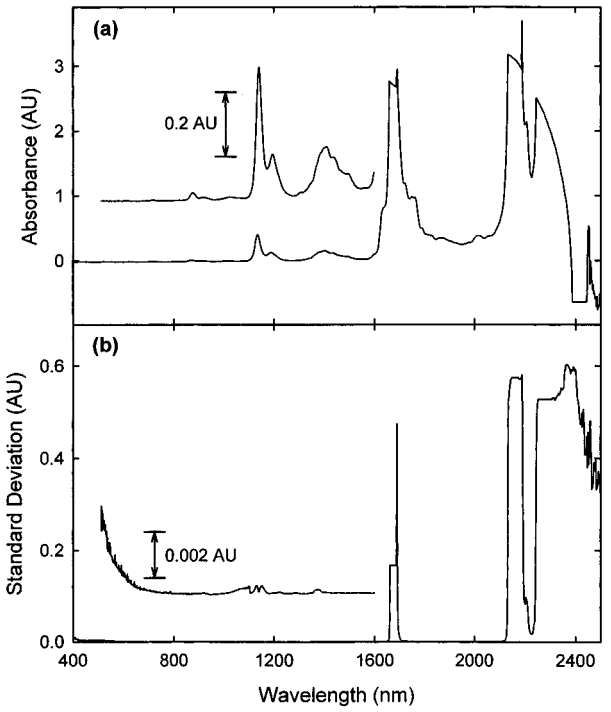


Figure 3. (a) Typical spectrum from data set 5 (toluene−chlorobenzene−heptane) and (b) corresponding measurement standard deviations from replicate scans. The region between 500 and 1600 nm has been enlarged (inset) for greater clarity.

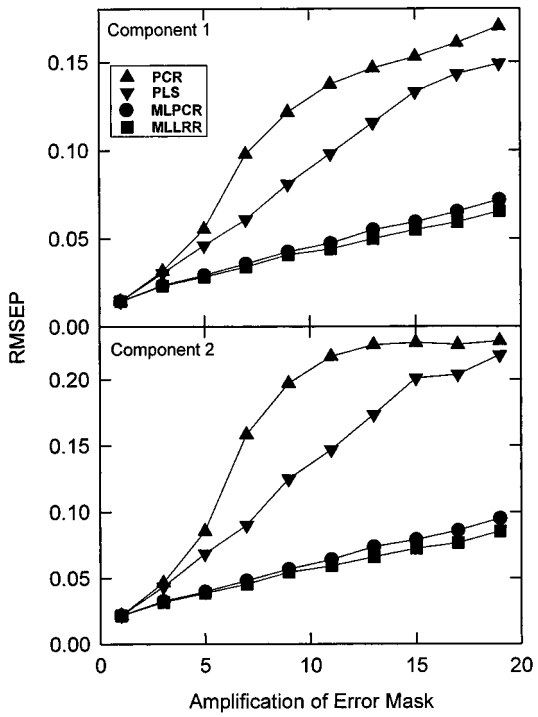


Figure 4. Comparison of calibration methods applied to simulated data subjected to a nonuniform error mask (data set 1).

scans of one sample. Not only will this fail to be a precise representation of the standard deviations from the remaining 30 samples, but it also does not reflect all of the sources of measurement error (e.g., cell positioning, sample preparation) for the sample for which it does apply. Nevertheless, it will be shown that MLPCR can utilize even this approximate information to provide better performance than conventional approaches.

**Computational Aspects.** The calculations performed in this work utilized a variety of computational platforms including (1) 486 and Pentium-based personal computers, (2) a Digital Equipment Corp. 3000/300X workstation with a 175 MHz clock speed and 96 MB of memory, and (3) a Sun Microsystems Sparc Server 1000 with 230 MB of memory and four 50 MHz SuperSPARC CPUs. All calculations were performed in Matlab (The MathWorks, Natick, MA).

## RESULTS AND DISCUSSION

**Simulated Data.** Initially, simulated spectroscopic data sets were used to assess the new calibration methods. Data set 1 was used to examine the effects of measurement errors whose standard deviation varies as a function of wavelength but is constant at any given wavelength. Such situations commonly arise when source intensity or detector sensitivity changes with wavelength, or when there is a strongly absorbing (constant) background component. Even when they are not coincident with the regions of the spectrum containing relevant information, noisy measurements can still influence the analysis, since the variance still needs to be accounted for by PCA. Although wavelength selection can often reduce this problem, the task of selection becomes difficult when noisy regions overlap regions of spectral significance, since the selection then relies on choosing the correct balance between the signal and noise retained in a given measurement. Maximum likelihood methods simplify the analysis by extracting the appropriate amount of information from each measurement.

Figure 4 shows the results of a comparison among PCR, PLS, MLPCR, and MLLRR for data set 1. Results are presented in terms of a root-mean-square error of prediction for components 1 and 2 in the three-component mixture (by symmetry, the results for component 3 will be statistically equivalent to those for component 1). The RMSEP is calculated as

$$\mathrm{RMSEP} = \sqrt{\sum_{i=1}^{N_{\mathrm{pred}}} (y_i^{\mathrm{pred}} - y_i^{\mathrm{true}})^2 / N_{\mathrm{pred}}} \qquad (23)$$

where $y_i^{\mathrm{pred}}$ and $y_i^{\mathrm{true}}$ are the predicted and actual concentrations of the analyte in prediction sample $i$, respectively, and $N_{\mathrm{pred}}$ is the number of prediction samples (100 in this case). In carrying out this calculation, the optimum number of latent variables was taken to be three for PCR, MLPCR, and MLLRR, since this should be the pseudorank of the calibration matrix by the constraints of the simulation. To permit greater flexibility for PLS, the optimum number of latent variables was selected by cross-validation (below an amplification factor of 8, the optimum number of latent variables was 3; from 10 to 12, it was 2; and above 12, only 1 was needed). For MLPCR and MLLRR, the standard deviation values known from the simulation were used for the measurement error estimates. In actual practice, these standard deviations would likely be determined from experimental replicates and would, therefore, be known with less accuracy, but the true values were used here to avoid introducing the number of replicates as a variable and also to examine a best-case scenario. For comparison, however, the MLPCR and MLLRR simulations were also run using variances estimated from five replicates. The results under these conditions were virtually identical, with all of the prediction errors falling within 3% of the values obtained when known

variances were used. Thus, at least in this case, the use of estimated variances did not have a large impact. Experimental data presented later also illustrate a case where standard deviations are estimated.

From Figure 4, it is apparent that, when the noise amplification factor, $r_{max}$, is unity (uniform noise), all of the methods perform equally well. This is expected since the maximum likelihood methods reduce to PCR under these conditions and there is unlikely to be any advantage of PLS over PCR. As the noise amplification factor is increased so that the variance on either side of the spectral range is amplified, the performance of all methods declines (prediction error increases). This is also expected, since increasing the noise decreases the information content of the data and increases the uncertainty. As the noise level is increased, the prediction error for the maximum likelihood methods remains significantly smaller than that of either PCR or PLS, illustrating the advantages of these techniques. It should be pointed out that, even at the limits of this study, the amplified noise on the wings of the spectrum represents only about 20% of the maximum signal in the calibration set, and this is not an unrealistic level. Nevertheless, the prediction errors obtained by the maximum likelihood methods are a factor of 2 to 3 smaller at this point than those obtained by the conventional methods. Comparison of the conventional methods indicates that PLS performs somewhat better than PCR in this case. This is due, in part, to the selection of an optimum number of latent variables for PLS, but a more important factor is likely to be the fact that PLS places some significance on correlation with the *y* variable in extracting latent variables, and so is not entirely based on *x*-variance.

It will also be noted that MLLRR consistently performs better than MLPCR in this example, although the difference is not substantial. The difference arises from the regression step in MLPCR, which assumes that the errors in *y* are uniform and much greater than the errors in the scores (for maximum likelihood estimation). In this example, however, the *y* values were generated with no errors, so the situation is exactly opposite of the second assumption, and MLLRR produces superior results. In a real calibration problem, it is likely that *y* will be determined by a reference method which has a significant uncertainty, so the assumptions of MLPCR may be more valid. It has been observed throughout this work that MLLRR generally yields results superior to those produced by MLPCR (because the most appropriate weighting of *x* and *y* is used) but that the two methods rarely give large differences.

A final point worth noting here is that the magnitudes of the errors are comparable for the two components in this example. In general, one might expect significantly larger prediction errors for component 2, since it is overlapped by two interferences (as opposed to one for components 1 and 3) and, therefore, should give a smaller net analyte signal.[13] However, the lower sensitivity of the method for component 2 is offset by the lower noise level near the center of the spectral range, so the prediction errors turn out to be comparable. The results presented here represent a limited study, and an infinite number of variations (spectral resolution, noise profiles, etc.) are, of course, possible. However, for all of the cases of this type that were examined, the maximum likelihood methods gave lower prediction errors than the conventional methods.

(13) Booksh, K. S.; Kowalski, B. R. *Anal. Chem.* **1994**, *66*, 782A−791A.

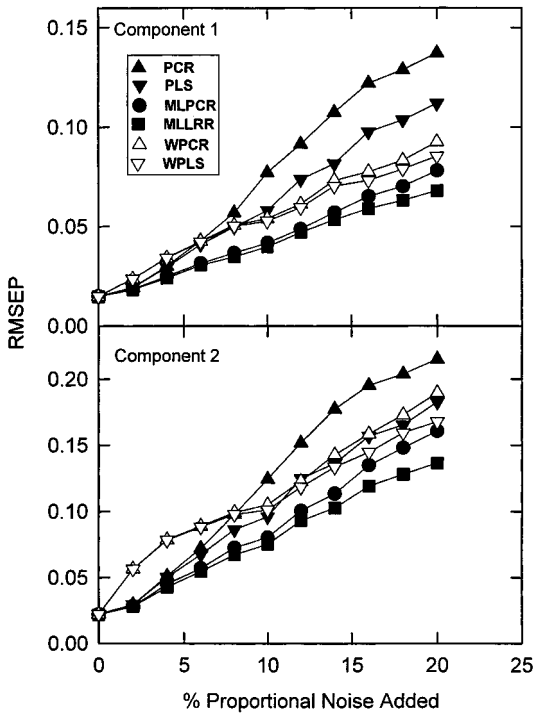


Figure 5. Comparison of calibration methods applied to simulated data with constant + proportional errors in spectra (data set 2).

Although data set 1 clearly shows the advantages of MLPCR and MLLRR, results comparable to those obtained with MLPCR for this data set could have been obtained simply by scaling each wavelength channel by its corresponding standard deviation prior to performing PCR. In other words, because the standard deviation matrix is rank 1, optimal scaling is possible. For this data set, scaling would be the preferred approach for reasons of computation speed, but optimal scaling is not possible in cases where the noise depends on signal magnitude. For this reason, data set 2, which contains both proportional and constant components of error, was employed for further comparison. A combination of errors was used to make the simulation more realistic, since purely proportional errors are rarely encountered.

Figure 5 provides a comparison of the prediction errors for the same four calibration methods applied to data set 2, as well as for two additional methods described in the following paragraph. Results are shown as a function of the level of proportional noise added to the data. Again, component 3 is omitted because of statistical equivalence, and, again, all methods are equivalent in the presence of uniform noise (0% proportional error). For PLS, the optimum number of latent variables was three up to 4% proportional noise and four thereafter. As before, the maximum likelihood methods show a significant improvement over the conventional methods, with the same order of performance. The improvement for component 1 is more striking than that for component 2 in this case, possibly because the central region of the spectrum remains more uniform in magnitude and, therefore, more uniform in noise. As with data set 1, the use of estimated variances with MLPCR and MLLRR gave only small differences in results (<7%).

To demonstrate that simple scaling is not sufficient to provide an improvement equivalent to the maximum likelihood methods for cases where the noise depends on signal amplitude, data set 2 was also examined using "weighted" PCR and PLS, designated

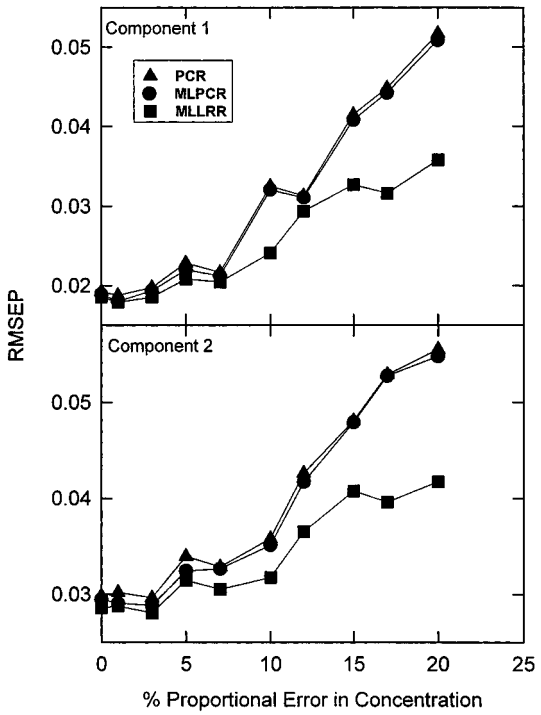


Figure 6. Comparison of calibration methods applied to simulated data with errors in both spectra and concentrations (data set 3).

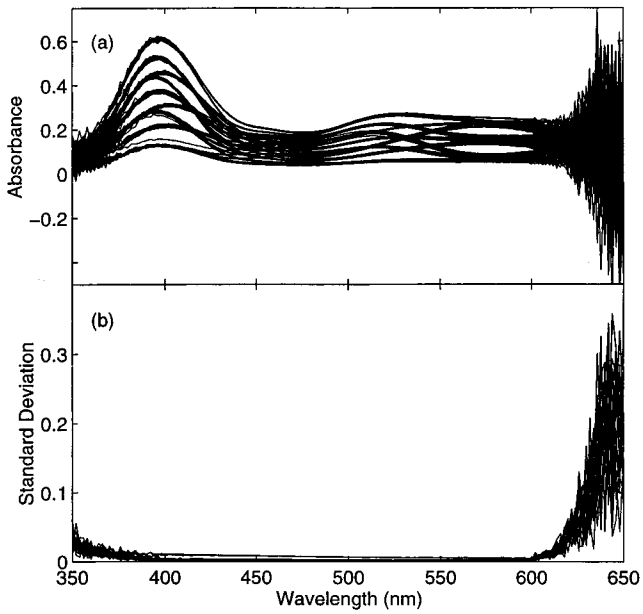


Figure 7. (a) Spectra for metal ion mixtures (data set 4) and (b) corresponding standard deviations.

as WPCR and WPLS in Figure 5. The data sets were scaled by the inverse of a pooled standard deviation at each wavelength. As the figure shows, this often results in smaller prediction errors compared to those obtained with PCR and PLS (with notable exceptions for small proportional errors), but the extent of improvement is less than what is achieved with the maximum likelihood methods. Although such suboptimal scaling may provide satisfactory results in some cases, it is our contention that MLPCR and MLLRR are preferrable because of their optimal performance in the general case, regardless of the error structure.

In the first two data sets, comparable performance was observed for MLPCR and MLLRR. It was speculated that differences would be exaggerated if significant nonuniform errors were added to the concentrations in the calibration set. For this reason, proportional errors were added to the reference concentrations in data set 3. This data set is also more realistic in the sense that multivariate calibration methods often use a reference method to determine concentrations in the calibration mixtures and such measurements are prone to uncertainty. Figure 6 shows the prediction errors for components 1 and 2 using PCR, MLPCR, and MLLRR as the level of proportional error in the reference concentrations is increased from 0 to 20%. The plot shows the actual prediction errors, i.e., the errors from the true concentrations in the prediction set rather than concentrations with errors added. As anticipated, the differences between MLLRR and MLPCR become more pronounced as the errors in the calibration concentrations increase, with MLLRR always providing superior results. In this example, there are only marginal differences between MLPCR and PCR, since the level of proportional noise is small enough to make the spectral measurements close to uniform error.

At this point, a comment should be made regarding the augmented error covariance matrix used for MLLRR. Throughout this work, it has been assumed that the errors in the calibration concentrations are uncorrelated with the errors in the spectral measurements. Strictly speaking, in a "designed" experiment, this may not be true. A designed experiment is one in which the mixtures are prepared by adding known amounts of the analytes to the calibration mixture. As such, there is no reference measurement used other than the gravimetric or volumetric data in the preparation. An error in these measurements can be considered to be correlated with the true error in the spectral measurements since it will affect these measurements proportionately. However, in practical circumstances, instrumental measurement errors are often much greater than the preparation errors, and the correlation can be considered insignificant. More importantly, multivariate calibration procedures more often employ "natural" calibration, where the concentrations in the calibration set are determined by a reference method which should be uncorrelated to errors in the spectral measurements.

**Experimental Data.** The first of the experimental data sets examined, data set 4, consisted of mixtures of Co, Cr, and Ni ions in dilute nitric acid. Spectra for these mixtures are shown in Figure 7a, with the corresponding measurement standard deviations for each set of replicates in Figure 7b. Increased noise levels are apparent at either end of the spectral range as the result of the optical filter placed in the light path. Two groups of samples were also found to have inordinately high standard deviations near the center of the spectrum, a problem that was traced to two of the samples out of the 130 which appeared to have an offset. The questionable samples were excluded from subsequent analysis, although it was found that their inclusion did not greatly affect the results. In the analysis of this data set, a diagonal error covariance matrix consisting of the variances for each measurement was used (i.e., uncorrelated errors were assumed). Although this assumption is known to be invalid, it was made to demonstrate the enhanced performance of maximum likelihood methods even when the error covariance information is unavailable.

To examine the predictive ability of various calibration methods, the technique of leave-one-out cross-validation was employed.

In this approach, the calibration model is first constructed for a particular analyte using all but one sample. The concentration of the analyte in the excluded sample is then predicted using the model, and the deviation from the expected concentration is measured. This process is repeated so that each of the 128 calibration samples is excluded once, and a root-mean-square error of cross-validation (RMSECV) is calculated by

$$\mathrm{RMSECV} = \sqrt{\sum_{i=1}^{N_{\mathrm{cal}}} (y_i^{\mathrm{pred}} - y_i^{\mathrm{ref}})^2 / N_{\mathrm{cal}}} \qquad (24)$$

where $y_i^{\mathrm{pred}}$ and $y_i^{\mathrm{ref}}$ are the predicted and reference concentrations, respectively, of the analyte in the excluded sample, and $N_{\mathrm{cal}}$ is the number of calibration samples. The RMSECV was calculated for each of the three analytes in the mixtures. An overall or total RMSECV was also calculated from

$$\mathrm{RMSECV}_{\mathrm{tot}} = \sqrt{(\mathrm{RMSECV}_{\mathrm{Co}}{}^2 + \mathrm{RMSECV}_{\mathrm{Cr}}{}^2 + \mathrm{RMSECV}_{\mathrm{Ni}}{}^2)/3} \qquad (25)$$

The RMSECV values calculated in this way give an indication of the predictive ability of the model. However, it should be pointed out that, for the PCR methods, two different approaches can be used for cross-validation. In what will be referred to as "leave-one-sample-out" cross-validation, PCA or MLPCA (as appropriate) is carried out on the subset of 127 calibration samples, and the results are used for calibration. In "leave-one-score-out" cross-validation, all 128 samples are used for PCA or MLPCA, and these results are retained for all subsequent calibrations, leaving the appropriate score out when building the calibration models by regression. In other words, the basis set is developed using all 128 samples, which are then projected onto the basis to obtain the scores. The regression is carried out on the scores for each combination of 127 samples, leaving one set of sample scores out in each case for cross-validation. This approach is faster, since PCA or (especially) MLPCA is only performed once (or once for each model dimensionality in the case of MLPCA). Although leave-one-score-out cross-validation can be considered a legitimate approach in that it does not employ concentration information about the prediction sample in the calibration procedure, purists may argue that it is not as valid as the leave-one-sample-out approach, which is completely blind to the prediction sample. For this reason, both approaches are included in the results presented here. As expected, the differences are very small, and the time savings of the leave-one-score-out method is a factor of $N_{\mathrm{cal}}$, an important consideration with MLPCR, which is substantially slower than PCR.

The results for data set 4 are presented in Table 1, which shows the RMSECV for each method and analyte as a function of the number of latent variables. The appropriate number of latent variables for this data set should be three, but since experimental realities such as offsets and nonlinearities can affect the optimum number of latent variables, results are given for up to six factors. The results are presented in tabular rather than graphical format because the range of values and number of methods would obscure a conclusive graphical interpretation. In addition to PCR, PLS, MLPCR, and MLLRR, results are also given for weighted PCR and PLS, using pooled standard deviations at each wavelength as weighting factors. For all of the methods examined, the predictive ability is poor when one or two latent variables are used, as expected. The maximum likelihood methods generally reach a performance plateau around three latent variables, where the prediction errors level off, although there is some marginal improvement with additional factors. For PCR and PLS, the plateau is less distinct, with additional factors continuing to bring further improvement. However, even with the addition of more latent variables than are shown in the table, the cross-validation errors for PCR and PLS did not reach the level of those for the maximum likelihood methods (the minimum total error for both methods was 0.38 mM, attained at 16 latent variables for PCR and 10 for PLS). The differences between the maximum likelihood methods and the conventional techniques are dramatic. Compared to PCR, the cross-validation errors for the maximum likelihood methods are more than an order of magnitude smaller in most cases. PLS fares somewhat better, but the RMSECV values are still substantially higher. Among the maximum likelihood methods, the results for MLPCR and MLLRR are very similar in most cases for this application. It will be also be noted that there is little difference between the leave-one-score-out and leave-one-sample-out cross-validation methods, as expected. In this example, the weighted regression methods (WPCR and WPLS) perform almost identically to the maximum likelihood methods, but this is expected, since the variances are primarily

Table 1. Comparison of Calibration Methods for Data Set 4 (Mixtures of Co, Cr, and Ni)[a]

| calibration method[b] | species | number of latent variables | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| MLPCR | Co | 10.68 | 6.34 | 0.32 | 0.32 | 0.19 | 0.17 |
| | Cr | 3.07 | 3.11 | 0.11 | 0.11 | 0.07 | 0.07 |
| | Ni | 24.35 | 17.98 | 0.38 | 0.37 | 0.33 | 0.33 |
| | total | 15.45 | 11.15 | 0.29 | 0.29 | 0.23 | 0.22 |
| MLPCR* | Co | 10.67 | 6.30 | 0.32 | 0.32 | 0.17 | 0.16 |
| | Cr | 3.07 | 3.09 | 0.11 | 0.11 | 0.07 | 0.07 |
| | Ni | 24.37 | 17.93 | 0.38 | 0.37 | 0.33 | 0.32 |
| | total | 15.46 | 11.12 | 0.29 | 0.29 | 0.22 | 0.21 |
| MLLRR | Co | 10.89 | 7.08 | 0.33 | 0.17 | 0.16 | 0.16 |
| | Cr | 3.47 | 3.43 | 0.11 | 0.07 | 0.07 | 0.07 |
| | Ni | 24.48 | 16.00 | 0.38 | 0.35 | 0.35 | 0.36 |
| | total | 15.60 | 10.30 | 0.30 | 0.23 | 0.23 | 0.23 |
| PCR | Co | 11.53 | 8.47 | 8.39 | 8.94 | 6.07 | 2.62 |
| | Cr | 3.51 | 3.11 | 3.15 | 3.29 | 2.44 | 0.85 |
| | Ni | 20.69 | 11.51 | 11.73 | 12.42 | 8.03 | 3.44 |
| | total | 13.82 | 8.44 | 8.52 | 9.04 | 5.98 | 2.54 |
| PCR* | Co | 11.53 | 8.42 | 8.27 | 8.33 | 5.80 | 2.46 |
| | Cr | 3.50 | 3.11 | 3.11 | 3.09 | 2.33 | 0.79 |
| | Ni | 20.69 | 11.48 | 11.57 | 11.53 | 7.63 | 3.18 |
| | total | 13.82 | 8.41 | 8.40 | 8.41 | 5.69 | 2.37 |
| PLS | Co | 11.58 | 9.43 | 1.72 | 1.49 | 0.63 | 0.60 |
| | Cr | 3.55 | 2.87 | 0.79 | 0.58 | 0.46 | 0.42 |
| | Ni | 20.41 | 8.83 | 2.35 | 1.09 | 0.97 | 0.70 |
| | total | 13.70 | 7.64 | 1.74 | 1.12 | 0.72 | 0.58 |
| WPCR | Co | 10.14 | 5.40 | 0.32 | 0.29 | 0.20 | 0.16 |
| | Cr | 3.08 | 3.02 | 0.11 | 0.10 | 0.08 | 0.07 |
| | Ni | 25.36 | 19.76 | 0.42 | 0.36 | 0.34 | 0.32 |
| | total | 15.87 | 11.95 | 0.31 | 0.27 | 0.23 | 0.21 |
| WPLS | Co | 10.14 | 5.43 | 0.32 | 0.29 | 0.23 | 0.16 |
| | Cr | 3.08 | 3.03 | 0.11 | 0.10 | 0.08 | 0.07 |
| | Ni | 25.37 | 19.85 | 0.42 | 0.37 | 0.36 | 0.32 |
| | total | 15.87 | 12.01 | 0.31 | 0.28 | 0.25 | 0.21 |

[a] Values given are the root-mean-squared errors of cross-validation (RMSECV) in millimolar. [b] Asterisk indicates leave-one-score-out cross-validation as opposed to leave-one-sample-out cross-validation.

dependent on the wavelength channel in this absorbance range. As demonstrated for data set 2, differences from the weighted methods are more obvious with errors that depend on signal magnitude. For this application, we would expect to see a greater effect at higher absorbance values. In any case, the utility of the maximum likelihood methods is that they guarantee an optimal estimation of the PCA subspace, which is not always assured with scaling.

The remarkably poor performance of PCR in this example motivated further examination of the reasons underlying the differences observed. In conducting this investigation, it was decided to focus on a comparison of PCR and MLPCR, since these two methods are the most complementary. Two of the most important factors influencing the performance of an analytical method are the sensitivity of the technique and the noise in the measurements. It is anticipated that, because of the nature of the geometric projections used, the uncertainty in the scores will be smaller for MLPCR than for PCR, but since such differences can be difficult to quantify in the general case, it was decided to focus on the sensitivity aspect. For first-order calibration methods, the sensitivity is related to the net analyte signal (NAS) by

$$\mathrm{SEN} = ||\mathbf{NAS}|| \tag{26}$$

where ||·|| indicates the Euclidean norm, or length, of the NAS vector.[13,14] The NAS for a given analyte is that part of the pure analyte spectrum that is orthogonal to the spectra of all other constituents in the mixture. The pure component spectra for a *p*-component mixture can be represented as vectors in an *n*-dimensional absorbance space (*n* = number of wavelength channels) and will define a *p*-dimensional subspace (hyperplane) within that space. If the vector representing the spectrum of analyte *i* (the analyte of interest) is now excluded, it is possible to identify a vector that is orthogonal to the remaining vectors and lies in the subspace defined by all *p* spectra. This vector is called the *contravariant* vector,[9] and it is the projection of the analyte spectral vector onto the unit vector in this direction that defines its NAS. Mathematically, if the pure component spectra of all constituents are known, the NAS is defined as

$$\mathbf{NAS}_i = (\mathbf{I} - \mathbf{R}_i(\mathbf{R}_i^{\mathrm{T}}\mathbf{R}_i)^{-1}\mathbf{R}_i^{\mathrm{T}})\mathbf{r}_i \tag{27}$$

where $\mathbf{R}_i$ is an $n \times (p-1)$ matrix whose columns consist of the pure component spectra for all constituents except the analyte, $\mathbf{r}_i$ is an $n \times 1$ vector containing the analyte spectrum (normalized to unit concentration), $\mathbf{I}$ is the $n \times n$ identity matrix, and $\mathbf{NAS}_i$ is the net analyte signal vector for analyte *i*.

An obvious problem with eq 27 is that the spectra of all constituents must be known. For situations where there are unknown constituents, methods such as PCR are used to estimate the NAS by regression against concentration. To make the problem mathematically tractable, PCA is used as the first step in PCR to identify the subspace of the pure component spectra. Calibration spectra are projected into this space and regressed against concentration. The NAS determined in the subspace, which will be designated as **NAS***, can then be transformed back to the original space. The important equations are

$$\mathbf{q}_i = \tilde{\mathbf{V}}^{\mathrm{T}}\mathbf{r}_i \tag{28}$$

$$\mathbf{Q}_i = \tilde{\mathbf{V}}^{\mathrm{T}}\mathbf{R}_i \tag{29}$$

$$\mathbf{NAS}_i^* = (\mathbf{I} - \mathbf{Q}_i(\mathbf{Q}_i^{\mathrm{T}}\mathbf{Q}_i)^{-1}\mathbf{Q}_i^{\mathrm{T}})\mathbf{q}_i \tag{30}$$

$$\mathbf{NAS}_i^{\mathrm{PCR}} = \tilde{\mathbf{V}}\cdot\mathbf{NAS}_i^* \tag{31}$$

In these equations, $\mathbf{q}_i$ ($p \times 1$) and $\mathbf{Q}_i$ ($p \times (p-1)$) are analogous to $\mathbf{r}_i$ and $\mathbf{R}_i$ in eq 27 and represent "abstract spectra" in the principal components space. $\mathbf{NAS}^*$ represents the $p \times 1$ net analyte signal vector in the subspace, and $\mathbf{NAS}^{\mathrm{PCR}}$ is the same vector in the original absorbance space. Note that $\mathbf{NAS}^{\mathrm{PCR}}$ is distinguished from the "true" **NAS** in eq 27 since they will only be identical in the ideal case. If, for example, PCA does not correctly determine the subspace of the component spectra, projection of individual spectra will result in a shorter vector and reduced sensitivity.

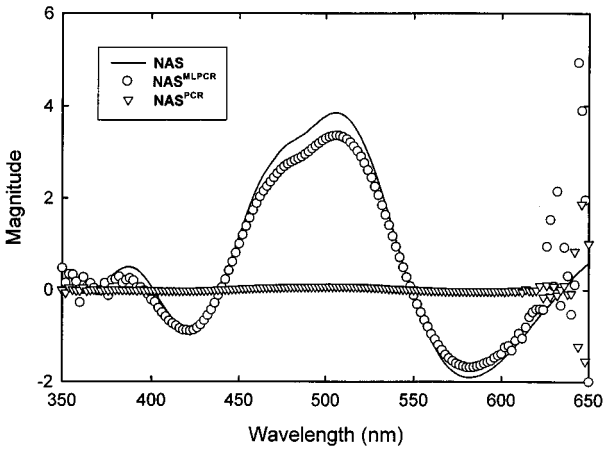


Figure 8. Comparison of net analyte signal vectors calculated for Co (data set 4) using different methods.

In the present study, pure component spectra are available for the three components in the mixture, and therefore, it is possible to obtain the NAS directly as well as by PCR and MLPCR. Figure 8 shows the results of this calculation for cobalt using three latent variables. Similar results were obtained for chromium and nickel, which are not shown. Note that **NAS** obtained from direct calculation and $\mathbf{NAS}^{\mathrm{MLPCR}}$ are very similar and have the expected shape. However, $\mathbf{NAS}^{\mathrm{PCR}}$ is much smaller in magnitude than the other two, and it is clear even without resorting to the calculation of eq 27 that the sensitivity of PCR will be much lower. These observations are consistent with results of Table 1. Note that the small NAS for PCR does not derive from the regression step, since none is used in this direct calculation method. Instead, it is believed that the spectral space is poorly estimated by PCA as compared to MLPCA, and subsequent projection into this space reduces the sensitivity of PCR. A comparison of eigenvectors produced by PCA and MLPCA is made in Figure 9, which shows the loadings (abstract spectra) for each of the first three factors. It will be noted that the first two factors are virtually identical for both methods, but there are radical differences in the third factor. While the third factor for MLPCA shows some meaningful structure in the spectrally active region, the PCA results are

(14) Lorber, A.; Kowalski, B. R. *J. Chemom.* **1988**, *2*, 67−80.

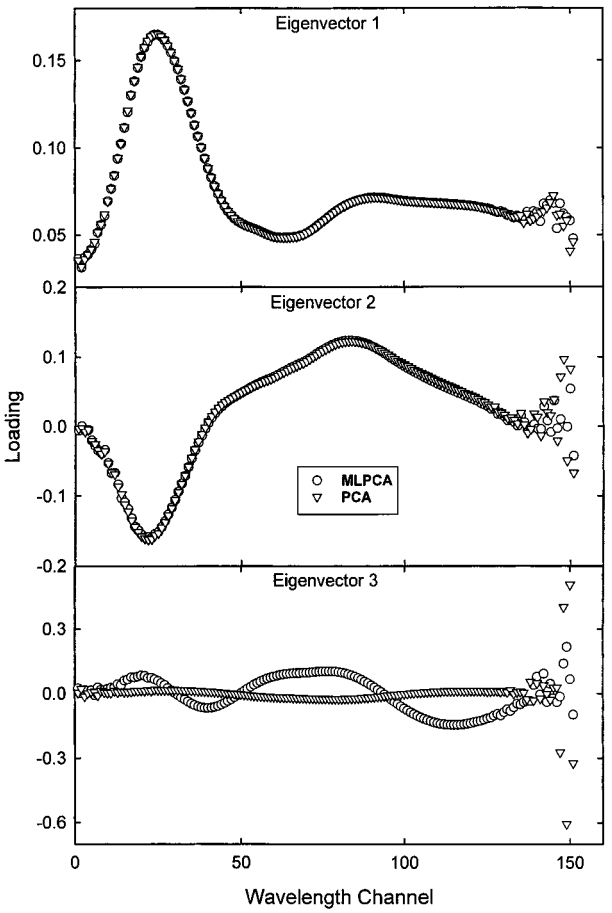


Figure 9. Comparison of eigenvector loadings for PCA and MLPCA applied to data set 4.

essentially flat in this region and show contributions mainly in the region dominated by noise. In other words, at the point at which the third principal component is extracted, the residual variance in the data set is dominated by the noise, and these are the regions modeled by PCA. MLPCA, on the other hand, is able to better account for the systematic variations. This is clearly indicated by the calibration results.

As a final illustration of the power of the maximum likelihood calibration methods developed here, consider data set 5. Based on the typical spectrum shown in Figure 3, one would normally choose to carry out PCR on a subset of the full spectral range, e.g., in the region of 700−1600 nm. If the high-noise regions are included, the PCR results are very poor due to the tendency of the PCA decomposition to model the noise variance. On the other hand, selecting a single region excludes other regions that may be useful for calibration. A more refined variable selection procedure could be used, but this normally relies on cross-validation and is extremely time consuming. A better approach is to apply MLPCR to the entire data set and allow the variance information to determine the importance of each channel.

A comparison of PCR and MLPCR for data set 5 is presented in Table 2 in terms of cross-validation errors (leave-one-score-out method). PCR was carried out over the region 700−1600 nm, while MLPCR was applied to the entire data set. For both methods, optimum performance occurs around three latent variables, as expected. It is clear that MLPCR generates models with significantly better predictive ability for all three components. Although it is not necessarily obvious from the spectra, it is apparent from the results that the inclusion of additional wavelength channels in the analysis improves the calibration model through MLPCR. This is because important information exists in the region above 1600 $cm^{-1}$ on the shoulders of peaks that saturate the detector. Thus, valuable information lost through suboptimal wavelength selection can be recovered through MLPCR.

Table 2. Comparison of PCR and MLPCR for Data Set 5 (Organic Mixture)[a]

| calibration method[b] | analyte | number of latent variables 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| PCR (700−1600 nm) | toluene | 16.60 | 7.68 | 0.61 | 0.60 | 0.60 | 0.46 |
| | chlorobenzene | 16.55 | 10.32 | 0.57 | 0.54 | 0.54 | 0.42 |
| | heptane | 3.13 | 2.66 | 0.15 | 0.14 | 0.14 | 0.14 |
| MLPCR (400−2500 nm) | toluene | 20.95 | 7.96 | 0.12 | 0.13 | 0.13 | 0.12 |
| | chlorobenzene | 13.08 | 10.32 | 0.13 | 0.11 | 0.11 | 0.11 |
| | heptane | 2.84 | 2.65 | 0.09 | 0.07 | 0.07 | 0.06 |

[a] Values given are the root-mean-squared errors of cross-validation (RMSECV) in weight percent. [b] Leave-one-score-out cross-validation was used for both methods.

It is also important to note that the results for data set 5 did not rely on precisely correct standard deviation estimates since, for all samples, these were based on 400 replicate scans for just one sample (so wavelength scaling could also have been used here). Correlations in the measurement errors, which are known to exist, were also ignored. Nevertheless, this approximation was sufficient to improve the calibration model. This suggests that even approximate information on measurement errors, such as that which might be provided by a skilled spectroscopist, can be used to advantage in multivariate calibration.

## CONCLUSIONS

It has been the objective of this work to describe the theoretical basis of maximum likelihood multivariate calibration methods (MLPCR and MLLRR) that are based on MLPCA and to present results demonstrating their ability to provide superior calibration models over conventional methods in certain cases. This objective has been accomplished through the use of both computer-generated and experimental data sets which showed that significant improvements over PCR and PLS can be realized by including measurement error information in the calibration procedure. In the majority of cases, MLLRR provided better results than MLPCR, but the improvement was often marginal for the cases examined here.

This study was not intended to be exhaustive in its investigation of the new methods and leaves open many issues concerning, for example, situations under which maximum likelihood methods should offer significant improvements, the relative merits of MLPCR and MLLRR under different measurement conditions, the role of measurement error covariance in the quality of a calibration model, and more extensive comparisons with other methods. Nevertheless, the underlying reasons for the improved results have been described from a fundamental perspective using standard figures of merit for multivariate calibration.

Two of the most common arguments against methods such as MLPCR and MLLRR relate to the requirement for measurement error variance estimates and the extended computation time necessitated by the algorithm. The first argument asserts that methods such as PCR require no variance information and are, therefore, more universally applicable. This argument is decep-

tive, since the use of PCR implicitly assumes that the measurement errors are uniform, so variance information is, in fact, required. In the absence of any knowledge of measurement error characteristics whatsoever, an assumption of uniform errors may be reasonable, but practitioners of PCR and similar methods should be aware of the limitations that such assumptions impose. It is the authors' contention that some instinct for measurement error characteristics on the part of the analyst is almost always present. Even if measurement error variances are not directly available, reasonable approximations of the error structure can be used effectively with the maximum likelihood techniques, as was demonstrated with data set 5. This should also be true even when the error distribution is only approximately normal, or when an exact covariance structure is not known. Finally, the results presented here support the case for designing instruments which provide measurement error information. Some instruments presently have this capability, but more often the information is unavailable, even when the instrument has the fundamental ability to provide it routinely from replicate scans (e.g., FT-IR spectrometers).

It is true that the maximum likelihood methods presented here are more computationally intense. However, the basic MLPCA algorithm[5] is quite simple to implement (about 30 lines of Matlab code) and converges reliably without the need for any "fine tuning" like many algorithms. Actual computation times vary with the size of the matrix and error structure and have been described elsewhere.[5] In this work, time for calculations ranged from several minutes to several days, with the longest times being observed for leave-one-sample-out cross-validation for MLPCR and MLLRR. As demonstrated here, leave-one-score-out cross-validation is generally equivalent for MLPCR and reduces computation time by a factor equal to the number of samples. This might typically take a few hours. Unfortunately, it is not possible to perform leave-one-score-out cross-validation for MLLRR because of the inclusion of concentration information, which is a drawback to this method. In any case, the time spent on calibration is still much less than that typically required to obtain the experimental data, and past history has demonstrated that computational barriers erode quickly with advancing technology.

Beyond the broad utility that these methods may find in practical situations, there is a more important aspect of their development. Whereas many new techniques are simply modifications of conventional methods designed to improve their utility, MLPCR and MLLRR are generalizations of PCR and LRR, respectively. In other words, PCR and LRR are special cases of the parent techniques that apply under conditions of uniform error variances. The development of general principles and methods for incorporating measurement uncertainties into the calibration process will allow the limitations and strengths of other calibration techniques to be appreciated from a wider perspective, a feature which is inherently valuable.

In the context of the preceding statement, the performance of PLS in the results presented here can be examined. Direct comparisons with PLS have been avoided until now because of basic differences in the fundamental philosophy toward the calibration process. It is generally viewed that, for systems with a well-defined rank, PLS should provide results comparable to those obtained with PCR when the correct number of latent variables is used (although PLS may provide better results than PCR when fewer latent variables are used). We have found this to be the case when uniform measurement errors prevail, but in cases where measurement errors are significantly nonuniform, PLS consistently performed better than PCR, although it performed worse than the maximum likelihood methods. This is likely because PLS uses correlation with concentration data to help exclude much of the noise variance. Based on this observation, one can speculate that the presence of nonuniform noise in many other applications may be partly responsible for the relative popularity of PLS over PCR in practical environments. This factor may also be important in the relative success of wavelength selection methods for some methods but not for others. Whatever the reasons for these observations, further investigation is warranted, and the maximum likelihood calibration methods presented here provide a unifying framework from which to better understand the application of multivariate calibration methods to chemical problems.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Center for Process Analytical Chemistry (CPAC) at the University of Washington.

Received for review October 9, 1996. Accepted March 14, 1997.[⊗]

AC961029H

[⊗] Abstract published in *Advance ACS Abstracts,* May 1, 1997.