INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

RATIONAL APPROACHES TO DATA PREPROCESSING IN MULTIVARIATE CALIBRATION

by

Christopher David Brown

Submitted in partial fulfillment of the requirements
for the degree of *Doctor of Philosophy*
at
Dalhousie University
Halifax, Nova Scotia, Canada
June, 2000

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Rational Approaches to Data Preprocessing in Multivariate Calibration"

by Christopher David Brown

in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: August 18, 2000

External Examiner

Research Supervisor

Examining Committee

DALHOUSIE UNIVERSITY

June 2000

Author: **Christopher David Brown**

Title: **Rational Approaches to Data Preprocessing in Multivariate Calibration**

Department: Chemistry
Degree: Ph.D.
Convocation: October, 2000

Permission is herewith granted to Dalhousie University to circulate and have copied for non-commercial purposes, at its discretion, the above title upon the request of the individuals or institutions.

Signature of Author

TABLE OF CONTENTS

Table of Contents iv
List of Figures vii
List of Tables xvi
Abstract xvii
Abbreviations and Symbols xviii
Acknowledgements xxiv

Chapter 1
Introduction 1

1.1 Notation 4
1.2 Multivariate Calibration 4
1.2.1 General Modeling Theory and Philosophy 4
1.2.2 Simple Calibration 6
1.2.3 Multivariate Calibration 10
1.2.3.1 Inverse versus classical calibration 10
1.2.3.2 Inverse Multivariate Calibration 15
1.2.4 Principal Components Regression 18
1.2.4.1 Principal Components Analysis 18
1.2.4.2 Principal Components Regression 26
1.3 Noise Considerations in Regression 28
1.3.1 Characteristics and Representations of Noise 28
1.3.1.1 Measurement Error Attributes 29
1.3.1.2 Error Variance and Covariance Representations 34
1.3.2 Measurement Error Structure and Multivariate Calibration 38
1.4 Maximum Likelihood Principal Components Regression 42
1.4.1 Maximum Likelihood PCA 43
1.4.2 Maximum Likelihood PCR 46
1.5 Figures of Merit in Multivariate Calibration 48
1.5.1 Mixture Theory and the Net Analyte Signal 49
1.5.2 Multivariate Figures of Merit 53

Chapter 2
Digital Filtering and Preprocessing 57

2.1 Introduction 57
2.2 Digital Filtering 58
2.2.1 Calculation, and Expression of Digital Filter Coefficients 60
2.2.2 Frequency Response of Digital Filters 64

Chapter 3
Digital Smoothing and Multivariate Calibration 67

3.1 Introduction 67
3.2 Theoretical Considerations 69
3.2.1 Characteristics of Digital Smoothing Filters 69
3.2.1.1 Filter order 70
3.2.1.2 Filter Width 73
3.2.1.3 Correlation of Noise 74
3.2.2 Smoothing Filters and Calibration Theory 77
3.3 Experimental 83
3.3.1 Simulated Data Sets 83
3.3.2 Experimental Data Sets 85
3.3.3 Computational Aspects 86
3.4 Results and Discussion 86
3.4.1 Noise Level 90
3.4.2 Spectral Correlation 95
3.4.3 Spectral Bandwidth 98
3.4.4 Correlated Errors 99
3.4.5 Experimental Data 101
3.4.6 A Dissection of Prediction Error 102
3.5 Conclusions 106

Chapter 4
Drift Correction in Multivariate Calibration 109

4.1 Introduction 109
4.2 Theory 114
4.2.1 Derivative Filters 114
4.2.2 Sensitivity and Selectivity Considerations 119
4.2.3 Derivative Filters and Baseline Drift 121
4.2.4 Optimal Corrections for Baseline Drift 122
4.3 Experimental 128
4.3.1 Simulations 128
4.3.1.1 Controlled Spectral Data 128
4.3.1.2 Randomly Generated Spectral Data 129
4.3.1.3 Introduction of Correlated Measurement Errors 131
4.3.2 Experimental Data 134
4.3.3 Computational Details 135
4.4 Results and Discussion 135
4.4.1 Derivative Filtering and Signals 135
4.4.2 Derivative Filtering and Noise 142
4.4.3 Maximum Likelihood PCA and Drift Correction 145
4.4.3.1 MLPCR with the True Error Covariance Matrix 145

4.4.3.2 MLPCR With an Estimated Error Covariance Matrix 148
4.4.3.3 Experimental Data 153
4.5 Conclusions 160

Chapter 5
Future Directions and Concluding Remarks 162

5.1 Future Avenues of Investigation 162
5.2 Conclusions 172

References 175

LIST OF FIGURES

Chapter 1

1.1 A simple scatter plot of *y vs. x* along with a simple first order polynomial fit to the observed data.....8

1.2 Sample space representation of the standard univariate linear model. **a)** the two vectors **x**, and **y** oriented in sample space, and **b)** the orthogonal projection of **y** onto the space defined by **x**.....9

1.3 An illustration of the general layout of classical least-squares methods.....13

1.4 An illustration of the general layout of inverse least-squares methods.....14

1.5 An illustration of the least-squares solution with two mixture components. The least-squares solution is the orthogonal projection of **y** onto the space defined by the **x** vectors, $S_{\mathbf{X}}$.....16

1.6 **a)** The data in the fundamental coordinate system [0 1, 1 0] with the first and second loadings shown. **b)** The same data represented in the coordinate system defined by the first two principal components. The scores of the data on these axes are also shown.....21

1.7 **a)** The projected data after discarding in the direction of the second loading. **b)** The same data represented in the (now) unidimensional coordinate system defined by the first principal component.....22

1.8 **a)** variable representation of the projection of the data onto a rank 1 principal component analysis subspace. **b)** Sample space representation of the same.....23

1.9 Illustration of the fundamental difference between least-squares estimation, and PCA estimation of a one-dimensional model space. **a)** The least-squares approximation under the assumption that $\mathbf{x}_1$ is the true model space. **b)** The PCA approximation under the assumption that both $\mathbf{x}_1$ and $\mathbf{x}_2$ are corrupted by errors—the true model space ($\mathbf{t}_1$) must be estimated.....25

1.10 An observed signal, **x**, which can be considered to be the true signal, $\mathbf{x}^o$, corrupted by measurement noise **e**. 30

1.11 Noise power spectra (NPS) of **a)** white noise, and **b)** 1/*f* noise estimated from 50 repeat samples of the noise. Also shown are samples of the noise in the time domain. The power spectra have been normalized to a total power of 1. 33

1.12 Examples of error covariance matrices for measurement errors which are **a)** *iid*, and **b)** non-*iid*. The *iid* errors are characterized by an error covariance matrix which is diagonal, and a multiple of the identity matrix, whereas the non-*iid* errors are characterized by heteroscedasticity (non-equal diagonal elements) and/or correlated error (non-zero off-diagonal terms). Samples of the noise vectors which possess the indicated error structures are also shown (inset). 36

1.13 Illustration of the geometry of measurement errors resulting from the structure of the noise. **a)** *iid* noise (homoscedastic, and uncorrelated; white in the frequency domain), **b)** heteroscedastic noise, which has greater magnitude on channel *j* than channel *i* (uncorrelated, also white in the frequency domain), and **c)** noise which is heavily correlated, and heteroscedastic (1/*f* characteristics in the frequency domain) 38

1.14 An illustration of the likelihood implications of projecting $\mathbf{x}_l$ onto the model space (assumed known in this case) when the measurement errors corrupting the true $\mathbf{x}_l$ vector are *iid*. 39

1.15 An illustration contrasting the likelihood of the projected $\mathbf{x}_l$ when the measurement errors deviate significantly from *iid*. In this case the likelihood of the projection given the true value is *extremely* low because a simple orthogonal projection has been used, which is far from ideal when the errors are non-*iid*. 40

1.16 **a)** Standard PCA orthogonal projection of an **x**-vector onto the subspace estimated by PCA, while the measurement error structure is clearly deviating from *iid* conditions. **b)** MLPCA projection under the same circumstances. The measurement error structure provides a directional guide for the projection of the **x**-vector on the subspace. The resulting estimates in **a)** and **b)** are in the same space, but are very different in length. 46

1.17 An algorithmic summary of MLPCA, and its regression counterpart, MLPCR, with equal row covariance assumptions. 48

1.18 The embodiment of the linear mixture model, which models any observed mixture spectrum as the linear combination of pure-component contributions. 50

1.19 A series of mixture spectra when viewed from a mixture theory perspective. The mixture vectors lie in the "mixture space" which, in the absence of noise, will be coplanar with the space defined by the pure-component spectra of the active components. The contravariant vectors (shown at left) indicate the directions in the mixture space which are exclusively associated with their associated components 51

1.20 The net analyte signal vectors for each component, colinear with their associate contravariant vectors, are the orthogonal projections of the pure-component spectra at unit concentration onto the contravariant vectors. 52

1.21 An examination of the different issues of importance in multivariate signal-to-noise ratios. **Case A**: low *S/N* - Why? Negligible *SEN*. **Case B**: High *S/N* - Why? Large *SEN*, and low projection of the error covariance onto the contravariant vector. **Case C**: Low *S/N* - Why? Large *SEN*, but error covariance is greatest in the direction of the NAS, and therefore has a very large projection onto the contravariant vector 56

Chapter 2

2.1 Illustration of the convolution of a set of filter coefficients with the raw signal vector to yield a filtered vector of measurements 59

2.2 Illustration of the application of a seven-point moving-average filter. The points within the filter window are used to estimate a low-ordered polynomial approximation to the data. The estimated centre-point value on this fit is taken as the filtered estimate of the signal vector at the corresponding ordinal variable. 60

2.3 Transfer functions for a variety of Savitzky-Golay digital filters including a 31-point moving average, 11-point quadratic smooth and the 11-point quadratic second-derivative filter 65

Chapter 3

3.1 Transfer function for a 15-point moving-average (zero-order polynomial) Savitzky-Golay smoothing filter. The frequency axis has been normalized by the Nyquist frequency for generality..... 69

3.2 A comparison of the use of (right) a 7-point moving-average smooth, and (left) a 7-point quartic polynomial smooth. The moving-average filter achieves more substantial noise reduction due to its more aggressive attenuation of the signal in the frequency domain - only the very low frequencies are unattenuated. 71

3.3 Illustration of the importance of the local modeling ability of the polynomial model in minimizing signal distortion. Higher-order polynomial filters will invariably model the true signal better, however lower noise rejection results..... 72

3.4 Example of the increased distortion observed in the signal features with increased filter widths. 74

3.5 Transfer functions of zero-order smoothing filters with a variety of widths..... 74

3.6 **a)** A vector of uncorrelated measurement errors, and the calculated noise power spectrum for the given noise sequence (white noise). **b)** The vector of smoothed noise values from **a)** using a 15-point moving-average filter. The noise power spectrum for these smoothed values is also shown (coloured noise)..... 75

3.7 An example of the calibration spectra employed for the simulation studies under standard conditions. Also shown in the inset are the 3 pure-component spectra, with the middle (dotted) Gaussian band corresponding to the analyte of interest, component 2..... 85

3.8 Experimental data used to validate the simulation studies. The data shown consist of 128 UV-VIS mixture spectra for metal ion mixtures in nitric acid. 86

3.9 Theoretical and PCR-observed performance ratios for multivariate calibration and prediction under the influence of a moving-average (zero-order) Savitzky-Golay smoothing filter. The observed values are the resulting averages from 30 replicate trials (error bars represent ± 1s)..... 88

3.10 Theoretical, and PCR-observed performance ratios for multivariate calibration and prediction under the influence of a quadratic (second-order) Savitzky-Golay smoothing filter. 89

3.11 **a)** Performance ratios observed for PCR over a variety of filter sizes as the noise level of the data is systematically changed from 0.001 to 0.1 (σ) for standard simulation conditions. **b)** An example of the individual performance ratios observed with specific noise levels of 0.001 and 0.1 (σ). PR_{theo} (—), observed PR values for: σ=0.001 and 20 calibration samples (●), σ=0.1 and 20 calibration samples (▼), and σ=0.1 and 40 calibration samples (σ). 91

3.12 Plot of the mean angle between the true and PCA-estimated pure-component subspace (25 repeat measurements) as a function of the level of the error corrupting the spectral vectors. 92

3.13 **a)** observed performance ratios for components 1 (●), 2 (O) and 3 (π) at a noise level of 0.1 (σ_{noise}). **b)** Theoretical *S/N* curves as calculated for all 3 components. **c)** Theoretical *S/N* curves determined for the filtered data based on the subspace as estimated by PCA. 94

3.14 **a)** Performance ratios observed over all filter widths for spectral angles varying between 10° and 85° for standard simulation conditions. **b)** Observed performance ratios for calibration systems comprised of pure-component spectral overlap of 10° and 76°. 95

3.15 Illustration of the dual effects of noise level (shaded area indicates a 1σ level) and spectral angle on the precision of the calibration subspace estimation. **a)** A representation of a large subspace angle, with a noise level *N*. The orientation of the calibration subspace is uncertain, but the uncertainty in the estimate is relatively small. **b)** A representation of a similar system corrupted by a noise level, *N*, but with very small subspace angles. Because the pure-component spectra are very close together, the noise contributes a much greater uncertainty in the subspace orientation. 97

3.16 Comparison of the *PR*'s (both theoretical, and observed) for spectral data with varying frequency content. Small σ_{peak} values correspond to higher frequency spectral features (more readily distorted by smoothing filters). 98

3.17 MLPCR observed performance ratio compared to the PR_{obs} for PCR with increasing smoothing filter width. The depicted values are

averages of 20 replicate trials (error bars indicate ± 1s). Also shown are the theoretical performance ratios..... 100

3.18 **a)** Performance ratios observed after applying smoothing filters to the experimental data set discussed in **Section 3.2**. The dashed line indicates the PR_{obs} = 1 (no change) mark. **b)** Actual *RMSEP*'s for each of the three analytes of interest as a function of applied smoothing filter width. 102

3.19 Results of the simulation studies examining the effect of each of the three effects of filtering in isolation. The performance ratios are displayed for an "ideal filter" (one which achieves noise variance reduction only), a filter which introduces error covariance effects, and a filter which introduces signal distortion. The observed performance ratio is also shown for the application of a real smoothing filter (all effects present). 104

Chapter 4

4.1 Transfer function for a 'true' derivative filter. 114

4.2 **a)** Transfer functions for a variety of realistic derivative filters including a difference filter, as well as 3- and 13-point linear Savitzky-Golay derivative filters. In **b)** the transfer functions for 13-point linear first- and quadratic second-derivative filters are compared, showing the change in bandpass associated with higher derivatives. 116

4.3 **a)** A simulated spectrum corrupted by drift noise, and the resulting derivative spectra from applying **b)** difference, **c)** 3-point linear first-derivative, **d)** 13-point linear first-derivative, and **e)** 13-point quadratic second-derivative filters (magnified by a factor of 10 for clarity) to the original spectrum. 117

4.4 An illustration the concept of optimal filter application. **a)** original spectrum exhibiting heteroscedasticity and error covariance characteristics, **b)** rotation of the spectrum by some angle to eliminate error covariance, and **c)** scaling of the scores on each axis to eliminate heteroscedasticity. 125

4.5 An algorithmic summary of the procedures required to perform MLPCR when derived as an optimal filtering method. This derivation implicitly assumes that the error covariance structure of the

prediction data is not significantly different than the error covariance structure of the calibration data. 127

4.6 An example of 20 noise-free calibration spectra generated using controlled criteria for the spectral correlations and frequency composition. The inset shows the 3 pure-component spectra generating this observed set of mixtures. 129

4.7 Two examples of pure-component spectra generated with random features in the spectral domain (—), and the Gaussian bands summed to generate them (·····). **a)** Pure-component spectrum generated using 4 randomly located Gaussian bands with width σ_{peak}=2 channels, and **b)** a pure-component spectrum generated using 4 randomly located Gaussian bands of width σ_{peak} =25 channels. 130

4.8 An example of 20 noise-free calibration spectra generated using random criteria for the spectral correlations, and Gaussian bands of width σ_{peak} =25 channels. The inset shows the 3 pure-component spectra generating these mixture spectra. 131

4.9 A typical example the controlled calibration data when corrupted by drift noise introduced with a 111-point moving-average filter. 133

4.10 73 NIR diffuse reflectance spectra constituting the experimental data set. 135

4.11 Figures of merit studies on sets of pure-component spectra. **a)** An example of the pure-component spectra generated randomly with very broad features. **b)** and **c)** depict the results of 500 replicate *SEL* and *S/N* measurements on data with these characteristics both before, and after filtering with a 11-point quadratic second-derivative filter. **d)** Sample pure-component spectra used (narrow features) in an identical study of **e)** *SEL* and **f)** *S/N* changes as a result of derivative filtering. 137

4.12 Multivariate *S/N* studies with varying spectral characteristics and derivative filter treatments both before, and after derivative preprocessing. The grid-layout of the 9 plots is as follows: travel down the vertical axis corresponds to increasing broadness of the pure-component spectra (σ_{peak} values were 5, 25, and 55 channels), travel from left to right on the horizontal corresponds to increasing the second-derivative filter width (filter widths were also 5, 25, and 55 channels) 138

4.13 Simulation of two distinct cases for derivative filtering in which the filter performance is highly dependent on the frequencies of importance in calibration. In **Case One**, the information is contained at high frequencies, while in **Case Two**, the information resides at low frequencies. 140

4.14 **a)** A noise sequence showing significant levels of drift noise. **b)** A derivative spectrum of a using a 5-point quadratic second-derivative filter. **c)** A derivative spectrum using a 13-point quadratic second-derivative filter. **d**, **e**, and **f)** Error covariance matrices corresponding to each noise sequence determined experimentally from 50 replicate measurements of the noise sequences..... 143

4.15 Noise power spectra (NPS) showing the frequency content of a raw noise sequence corrupted with drift, and the resulting NPS's after treatment of that noise with a difference filter and 13-point quadratic second-derivative filter. 144

4.16 An illustration of the drift correction power of MLPCA compared to conventional PCA. **a)** 20 calibration spectra generated under controlled conditions and corrupted with substantial drift noise using a filter width of 95 channels (the noise has been scaled up to σ=0.05 to illustrate the point more clearly). **b)** The calibration data reconstructed from a rank 3 principal component subspace, and **c)** The calibration data reconstructed from a rank 3 space using MLPCA..... 146

4.17 Simulation results comparing the performance of PCR to derivative PCR and MLPCR. **a)** PCR, MLPCR and linear first-derivative PCR, and **b)** PCR, MLPCR and PCR with quadratic second-derivative preprocessing..... 147

4.18 *RMSEP* for MLPCR as a function of the number of replicates of each calibration sample used to estimate the *pooled* error covariance structure. The performances of PCR and derivative PCR are shown for reference..... 150

4.19 Performance of MLPCR as a function of the number of replicates used to estimate the error covariance structure of the data (*no pooling*). The performance of MLPCR on the same data when the error covariance estimate was smoothed (25-point block smooth) prior to use in MLPCR. Plots **a**, **b**, and **c** correspond to low, medium and high levels of drift (drift introduced with smoothing filter widths of 19, 59, and 99 channels)..... 151

4.20 The estimated error covariance structure for the experimentally obtained NIR diffuse reflectance spectra. 154

4.21 A visual comparison of the effects of drift correction on the 73 calibration spectra using a 13-point quadratic second-derivative filter and MLPCA. 156

4.22 A close-up illustration of drift correction on 5 repeat spectra of the same sample using derivative preprocessing (13-point quadratic second-derivative) and MLPCA drift correction. 157

Chapter 5

5.1 **a)** The effect of multiplicative error in the spectral domain, **b)** the effect of offset errors shown in a complimentary fashion, and **c)** the combined effects illustrated geometrically. 164

5.2 An illustration of the least-squares fitting procedure used to correct each of the mixture spectra in a calibration or prediction set to a mean scattering level. The actual mixture spectrum, and the mean spectrum are shown in the inset. 165

5.3 **a)** Simulated raw spectral data (25 spectra) corrupted with additive, multiplicative, and white noise. **b)** Those same spectra after MSC application. 166

5.4 Geometric interpretation of MSC in two dimensions. 167

5.5 A visual comparison of the MSC treated NIR data to the raw spectral data, and the previously discussed MLPCA correction. 169

LIST OF TABLES

Chapter 4

4.1 *SEN* and *SEL* values for **Case One** and **Case Two** both before (d^2 - no), and after (d^2 - yes) treatment with a 13-point second-derivative filter. The results are shown for all three components. **Case One** shows an enhancement in the *SEL*'s upon differentiation, while **Case Two** shows a substantial decrease in *SEL*. 141

4.2 Summary of calibration performances for PCR, MLPCR and various forms of derivative preprocessing used in conjunction with PCR. 'Filter condition' (OD) refers to 'O' the polynomial order of the SG filter, and 'D' the first (1) or second (2) derivative. (LV: latent variables determined to be optimal, *PR*: performance ratio of *RMSECV* for MLPCR to the *RMSECV* of the method under inspection) 159

Chapter 5

5.1 Summary of results for cross-validation studies of PCR and MLPCR models generated from unprocessed ABS polymer data, and from MSC corrected data. (LV = number of latent variables, *RMSECV* = root mean-squared error of cross-validation) 168

ABSTRACT

Multivariate calibration, used in conjunction with multichannel instrumental techniques, has been vital in making convenient, rapid and cost-effective chemical analysis possible. Numerical preprocessing techniques, which are intended to recondition the measurement data to a form which is better suited for chemometric methods, often play a key role in multivariate calibration. In some cases, the use of preprocessing techniques improves the precision of the analytical result. In other cases, meaningful results are altogether impossible *without* preprocessing in some form. Despite the integral importance of preprocessing strategies in multivariate analysis and calibration, the theoretical impact of many of these numerical methods in calibration theory is unknown, leaving the analyst no other option than a trial-and-error approach. In this work, two of the most prominent preprocessing methods, digital smoothing and differentiation, are examined in depth from the perspective of calibration theory.

Smoothing is very frequently performed with aspirations of enhancing the signal-to-noise ratio (*S/N*) of the measurement data. It is demonstrated here that, based on theoretical considerations, no enhancement in multivariate *S/N* or predictive ability can be anticipated from symmetric smoothing filter application. In practical studies, it is observed that gains can sometimes be made, although they are found to be consistently marginal, and attributable to substantial calibration model error. This leaves smoothing filters in multivariate calibration as little more than cosmetic devices which are more likely to obfuscate information than enhance it.

Derivative filters are widely employed for the alleviation of baseline drift, and other noise structures which contribute error covariance to the measurement data. Theoretical examinations of their operation reveal that drift reduction proceeds by attempted diagonalization of the error covariance matrix (and homogenization of the noise power spectrum), although this benefit is often offset by the deleterious side-effects of derivative filtering: potential signal degradation and loss of chemical interpretability. While derivative filters do relieve error covariance to some extent, they are suboptimal in their approach as no consideration is given to heteroscedasticity, error covariance, or the net analyte signal. It is shown that optimal drift correction methods can actually be derived by direct consideration of the error structure. It is further demonstrated that this optimal drift correction filter is a special case of maximum likelihood principal components analysis, a method recently introduced by this research group.

This work demonstrates that preprocessing and calibration strategies can be *logically* developed from careful consideration of the problem at hand. These rational approaches to calibration not only are often superior in performance, but also avoid the wildly empirical and inefficient approaches in widespread use.

ABBREVIATIONS AND SYMBOLS

In general, the conventions used in this thesis are as follows (also noted in **Section 1.1**). Matrices are represented as upper case bold letters and column vectors are represented as bold lower case letters. Normal face fonts (lower case) are used for scalars. Upper case italicized letters denote a vector space defined by the columns of a matrix (usually indicated by the subscript on the vector space symbol). Normal, Greek and script fonts are used with no particular pattern, but where possible, we have tried to adhere to symbols commonly used in the literature. Symbols that represent least-squares estimates of unknown quantities are designated with a caret ("^"). Symbols that represent truncated matrices (rank-reduced) are designated with other modifiers such as "~", or "∪". A matrix transpose is indicated by a superscript "T" and the Euclidean norm of a vector by ||o||. Matrix inverse notation will be dealt with in a context specific manner, although a superscript "-1" will consistently refer to a true matrix inverse.

List of abbreviations:

ABS	Acrylonitrile-Butadiene-Styrene
CLS	Classical Least-Squares
Co	Cobalt
Cr	Chromium
CV	Cross-Validation
iid	independent, and Identically Distributed (normally)
ILS	Inverse Least-Squares
LOD	Limit of Detection
MLPCA	Maximum Likelihood Principal Components Analysis
MLPCR	Maximum Likelihood Principal Components Regression
MLR	Multiple Linear Regression

MSC	Multiplicative Scatter/Signal Correction
NAS	Net Analyte Signal
Ni	Nickel
NIR	Near Infrared
NPS	Noise Power Spectrum
PCA	Principal Components Analysis
PCR	Principal Components Regression
PC	Principal Component
PLS	Partial Least-Squares
PR_{obs}	Observed Performance Ratio (**Chapter 3**)
PR_{theo}	Theoretical Performance Ratio (**Chapter 3**)
PR	Performance Ratio (**Chapter 4**)
R	Correlation Coefficient
RMSECV	Root Mean-Squared Error of Cross-Validation
RMSEP	Root Mean-Squared Error of Prediction
S/N	Signal-to-Noise
SEL	Selectivity
SEN	Sensitivity
SNV	Standard Normal Variate transformation
SG	Savitzky-Golay
SSR	Sums-of-Squares of the Residuals
SVD	Singular Value Decomposition

List of Symbols:

$\mathbf{0}$	null vector of suitable length
$\mathbf{1}$	one vector of suitable length
a	(i) molar absorptivity (**Chapter 1**)
	(ii) cosine multiplier in filter transfer function equation (**Chapter 2**)

	(iii) offset coefficient (**Chapter 4**, and **5**)
a_{ij}	elements of the matrix **A**
$\mathbf{a}_i$	vector of filter coefficients from the *i*th row of **A**
A	matrix of filter coefficients for polynomial least-squares filters
l	pathlength
b	(i) regression parameter
	(ii) sine multiplier in filter transfer function equation (**Chapter 2**)
$\hat{b}$	least-squares estimate of b
$\mathbf{b}$, $\mathbf{b}_i$	regression vector
$\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_i$	least-squares estimate of $\mathbf{b}$, $\mathbf{b}_i$
B	matrix of regression vectors
$\hat{\mathbf{B}}$	least-squares estimate of **B**
c	(i) concentration
	(ii) filter coefficients (**Chapters 2**, **3**, and **4**)
$\bar{c}$	mean concentration
$\mathbf{c}$	column vector of concentrations
C	matrix of concentrations
$\hat{\mathbf{C}}$	least-squares estimate of **C**
e, $e_{1,i}$	error observed on the first channel of the response vector, on the *i*th repetition
$\mathbf{e}$	column vector of error terms
$\hat{\mathbf{e}}$	least-squares estimate of $\mathbf{e}$
$\tilde{\mathbf{e}}$	residual error vector between $\mathbf{x}$ and $\tilde{\mathbf{x}}$
f	frequency
f_N	Nyquist frequency
$\mathbf{F}$	filter matrix
$\mathbf{F}_o$	optimal filter matrix
$\mathbf{i}_i$	information vector for the *i*th spectrum in MSC
$\mathbf{I}_n$	n x n identity matrix

k	number of wavelengths selected in MLR
m	number of samples
n	number of response channels in a vector measurement
p	(i) number of active chemical components in a mixture (ii) dimension or rank of PCA, or MLPCA
$\hat{\mathbf{p}}$	PCA regression vector from regression on scores formulation
r	instrument response
$\bar{r}$	mean response
$\mathbf{r}$	column vector of instrument responses
$\mathbf{R}$	matrix of instrument responses
s_1^2	variance of the sample on channel 1
s_{12}	covariance of the sample on channels 1 and 2
$\mathbf{s}$	(i) column vector of pathlength-normalized molar absorptivities (ii) pure-component instrument response vector at unit concentration
$\mathbf{s}_i$	*i*th pure-component instrument response at unit concentration
$\mathbf{s}_i^o$	true *i*th pure-component instrument response at unit concentration
$S_{\mathbf{X}}$	vector space spanned by the column vectors of $\mathbf{X}$
$S_{\tilde{\mathbf{X}}}$	vector space spanned by the column vectors of $\tilde{\mathbf{X}}$
$S_{\mathbf{X}}^o$	true vector space spanned by the column vectors of $\mathbf{X}^o$
$S_{\tilde{\mathbf{X}}}^o$	true vector space spanned by the column vectors of $\tilde{\mathbf{X}}^o$
$\mathbf{S}$	(i) matrix of pure-component instrument responses (ii) matrix of singular values from SVD (see below)
$\hat{\mathbf{S}}$	least-squares estimate of $\mathbf{S}$ (for (i) above only)

t_{ij}	element of the matrix **T**
$\mathbf{t}_i$	score vector (columns of **T**) for the *i*th loading, or PC
$\mathbf{T}$	matrix of scores
$\tilde{\mathbf{T}}$	truncated **T** in PCA
$\mathbf{U}, \mathbf{S}, \mathbf{V}^T$	matrices returned by SVD, PCA, or MLPCA
$\tilde{\mathbf{U}}, \tilde{\mathbf{S}}, \tilde{\mathbf{V}}^T$	truncated $\mathbf{U}, \mathbf{S}, \mathbf{V}^T$ in PCA
$\check{\mathbf{U}}, \check{\mathbf{S}}, \check{\mathbf{V}}^T$	truncated $\mathbf{U}, \mathbf{S}, \mathbf{V}^T$ in MLPCA
$\mathbf{v}_i$	(i) *i*th loading vector (ii) contravariant vector for the *i*th component
$\mathbf{v}_i^\circ$	true contravariant vector for component *i*
$\mathbf{V}$	matrix of loading vectors
$\tilde{\mathbf{V}}$	truncated **V** in PCA
x_{ij}	element of the matrix **X**
$\mathbf{x}, \mathbf{x}_i$	(i) column vector of independent variables in general regression formulae (ii) column vector of **X**, or row vector of **X** (context)
$\hat{\mathbf{x}}$	MSC corrected spectrum
$\mathbf{x}^\circ, \mathbf{x}_i^\circ$	true value of $\mathbf{x}$, $\mathbf{x}_i$ observable in the absence of measurement errors
$\mathbf{X}$	matrix of independent variables in general regression formulae
$\mathbf{X}^\circ$	matrix of true values of **X** observable in the absence of measurement error
$\tilde{\mathbf{x}}, \tilde{\mathbf{X}}$	PCA estimated $\mathbf{x}$, and **X**
y_i	(i) single instance of the dependent variable in general regression formulae (ii) signal measurement (**Chapter 2**)
y'	first derivative of y (ii)

$\mathbf{y}$	column vector of dependent variables in general regression formulae
$\mathbf{Y}$	matrix of dependent variables in general regression formulae
$\hat{\mathbf{y}}$	least-squares estimate of $\mathbf{y}$
$\hat{\mathbf{Y}}$	least-squares estimate of $\mathbf{Y}$
$\mathbf{Z}_F$	$\mathbf{X}$ data under rotation by the optimal filter
$\tilde{\mathbf{Z}}_F$	rank reduced $\mathbf{Z}_F$ by PCA
σ_i^2	variance of the population for *i*
σ_{ij}	covariance of the population for *i* and *j*
Σ	error covariance matrix
ϕ	phase component of filter frequency response
ω	angular frequency
Λ	matrix of eigenvalues from optimal filter design ($\Lambda\Lambda = \mathbf{S}$ from (ii), **Chapter 4**)

ACKNOWLEDGEMENTS

There are a number of people to whom I am indebted for contributions to both my personal, and professional growth during my doctoral studies at Dalhousie. Dr. Peter Wentzell has exhibited what I consider to be the very highest attribute of all great supervisors and teachers—he has made himself progressively unnecessary. To borrow words from Yeates, he has succeeded in 'lighting a fire, rather than filling a bucket', and, from the initiation of my graduate work, adapted his style to my bull-headedness as an autodidactic. Helpful interactions with committee members (Drs. Lou Ramaley, Robert Guy, and Charles Warren), and Drs. Randy Pell and Mary Beth Seasholtz (Dow Chemical) are also greatly appreciated.

Financial support for the research discussed in this thesis was provided principally by Dow Chemical Company (Midland, MI), and the Natural Sciences and Engineering Research Council (NSERC) of Canada. I am also grateful for graduate fellowships and scholarships awarded by Dalhousie University, the Walter C. Sumner Foundation, and AOAC International.

Family have been great in making all of this possible from the beginning. Dad not only tolerated, but fueled my constant curiosities in how's and why's, and Mum taught me to stand up and sing. Grandma's willingness to feed a hungry undergraduate went well beyond tomato soup and egg sandwiches. During our frequent and highly anticipated lunches she instilled in me an appreciation for questions (even when they may not have answers), and grayed the lines between science, art and religion previously established by inculcation.

Natasha has steadfastly supported me during the last several months, despite very late nights at work and little time at home; my true gratitude for her understanding and unswerving dedication would run on for pages. While carrying this unwieldy burden, it has been of immense importance to know I am standing on something solid.

> The troubles of our proud and angry dust
> Are from eternity, and shall not fail.
> Bear them we can, and if we can we must.
> Shoulder the sky, my lad, and drink your ale.
>
> *A. E. Housman*, in *Last Poems*, no. 9 (1922)

1. Introduction

The theory and practice of multivariate analysis in chemical problems has come a long way since the inception of the word "chemometrics" (originally the Swedish 'kemometri') in 1972 [1, 2]. The early applications of what came to be called chemometric methods involved pattern recognition, classification, and regression. These novel uses of multivariate mathematical methods with chemical experimentation marked the beginning of what is now a reasonably mature field, with two dedicated chemometrics peer-reviewed journals, and chemometrics-related research articles regularly appearing in scores of other chemistry publications. The applications of these mathematical and statistical techniques have spanned virtually all sub-disciplines of chemistry; however, it is analytical chemistry which has most rapidly evolved and accommodated chemometric methodologies. As Harald Martens succinctly put it [3], this chemometric movement in analysis was, at least initially, motivated by the problem of simply having "too much data".

Analytical chemistry—the science of chemical instrumentation and measurement [4]—was perhaps predestined to benefit most substantially from chemometrics—the science of extracting information from chemical measurements using mathematical methods. The advantages of chemometric methods are dramatically pronounced when this information is not readily attainable by conventional means, or when the information seems hopelessly obscured by interfering phenomena such as meddlesome chemical species and noise. While classical univariate analytical methods require full selectivity for functionality, multivariate chemometric methods are much more flexible, requiring only fractional selectivity in a multivariate context. As a result, sample preparation procedures can often be dramatically reduced, saving the analyst both time and money.

Analytical instrumentation is constantly improving, but the rapidly increasing complexity and nature of the systems of interest often pushes analytical instruments far from their ideal operating state of full analyte selectivity. Aside from the anticipated low analyte selectivities in complex samples, the desire to avoid extensive sample preparation also often leads to bulk sample properties that can be problematic for many analytical methods. An example of this is the now widely used near infrared (NIR) reflectance spectroscopy of solid samples. In the majority of cases, other analytical procedures could be undertaken, however they would require extensive preparatory work. In contrast, NIR spectroscopy often requires little sample pretreatment. This preparatory relief is achieved with expense, however. The data acquired in these experiments are often heavily corrupted by multiplicative scattering effects arising from pathlength variations at the photonic level (it has been said that as much as 99% of the variance observed in NIR reflectance is unrelated to chemical composition [5]). Situations like this can lead to chemical measurements that exhibit rather unconventional (and less than desirable) properties, and, consequently, are notoriously problematic in standard chemometric methods.

In the past, numerous mathematical manipulations and transformations of the measurements have been proposed to massage the data prior to multivariate analysis, a procedure that is generally referred to as 'preprocessing'. Preprocessing methods are typically applied in order to make the measurement data more amenable to the mathematical modeling methods most commonly employed in chemometrics, such as principal components analysis (PCA) and principal components regression (PCR). Preprocessing potentially encompasses everything from the elementary centering, normalization, and linearization operations, to the more complex procedures such as digital filtering, transforms, and projection methods. While many of these types of preprocessing procedures are widely used, the theoretical knowledge of how the more complex methods affect multivariate calibration models is surprisingly incomplete. The selection of the preprocessing techniques to employ in a particular circumstance is largely a

guess-and-check procedure which is wildly empirical, not to mention costly and time-consuming. Some general guidelines do exist for the simpler preprocessing methods such as centering and scaling, but more complex preprocessing operations such as digital filtering have never been explored using the theoretical concepts entrenched in multivariate calibration theory.

In this work an attempt has been made to examine in depth the effects of several more advanced, but commonly employed preprocessing methods on multivariate calibration models. Following two brief introductory chapters on calibration theory, and digital filtering, **Chapter 3** explores the efficacy of digital smoothing filters as a preprocessing tool in multivariate calibration. It is demonstrated by both mathematical derivation and experimentation that enhancements in the multivariate signal-to-noise ratio are very rarely achieved using these smoothing filters. In **Chapter 4**, common preprocessing tools for drift reduction are examined, with the emphasis placed on the very popular Savitzky-Golay (SG) derivative filters. From theoretical insights, it will be evident that derivative filters are suboptimal in reducing drift noise. An optimal preprocessing procedure for the elimination of drift noise is developed and discussed. While this optimal filter is designed from the perspective of digital filtering, it is shown to be equivalent to the recently introduced maximum likelihood PCA (MLPCA) when used in conjunction with projections on principal components. **Chapter 5** briefly discusses the scatter and drift correction method multiplicative scatter/signal correction (MSC), and illustrates its performance compared to MLPCA. The chapter closes with a discussion of the context and importance of this thesis research in contrast to current philosophical approaches to chemometric modeling and preprocessing. Possible avenues for future investigations are also discussed.

1.1 Notation

In order to retain consistent representations of the mathematical concepts in this work, the notation is standardized in all chapters. Italicized lowercase symbols will represent scalar-valued quantities (*e.g.*, x), while lowercase bold characters will denote vector-valued quantities (*e.g.*, $\mathbf{x}$). Unless otherwise indicated, these vectors will be column vectors. Matrices will be indicated by boldface uppercase characters (*e.g.*, $\mathbf{X}$), and the space defined by the column or row vectors of a matrix (the *vector space* of the matrix) will be implied when uppercase italicized characters are used (*e.g.*, S). In order to distinguish between 'true values' which would presumably be observed under noise-free, ideal conditions, and the experimentally observed values, a superscript 'o' will be used. $\mathbf{X}^{o}$ would therefore represent the matrix of *true* values, while $\mathbf{X}$ would represent the *observed* experimental values. Standard modifying characters also include a vector or matrix transpose, indicated by a superscript 'T' (*e.g.*, $\mathbf{X}^{T}$), and the matrix inverse, indicated by a superscript '-1' (*e.g.*, $\mathbf{X}^{-1}$). Several descriptive characters will also be used through the text, such as a superscript '^' over matrices or vectors (*e.g.*, $\hat{\mathbf{y}}$), which indicates a least-squares estimate of the respective parameter. Other modifying characters of this form will be discussed further upon initial introduction.

1.2 Multivariate Calibration

1.2.1 General Modeling Theory and Philosophy

The overwhelming majority of applications in chemometrics utilize mathematical and statistical techniques to develop models which are intended to make accurate predictions. At the heart of this chemometric modeling process is the development of a useful relation between some readily measurable qualities (such as temperature, or absorbance), and properties of interest that are inconvenient (if not impossible) to directly observe (*e.g.*, concentration). Ideally,

a strong correlation exists between the easily measured qualities and the properties of interest, making it feasible to make inferences about samples with unknown properties. The modeling process is simplified if this correlative relation is linear in nature (with respect to the model parameters), although this is certainly not a necessity. The model, therefore, provides a method of converting or mapping information about one set of variables to information about another set of variables.

From a broad outlook, the analytical calibration process consists of two crucial steps: (1) establishing an estimate of the response space in which chemical variation of interest is expected to occur in the absence of other interfering signals, and (2) projecting observed signals onto that estimated space. A third step, determining regression coefficients that map one set of observations to another, is also crucial but the estimation of the regression coefficients is heavily dependent on the success of (1) and (2). If subspace estimation is successfully achieved, and the projections of the signals are accurate, then we can anticipate accurate, and hence powerful models of the relationship between the variables of interest.

In chemistry, we are often fortunate to have solid physical understandings and theories regarding these models. It is standard practice, for instance to automatically resort to using a first-order linear model when we are dealing with spectroscopic measurements, since the well known Beer's (or Bouguer-Lambert-Beer's) Law is presumed to sufficiently describe the theoretical relationship between analyte concentration and spectroscopic absorbance. It could be said, however, that there are two classes of models – those that are designed to aid in the description of the underlying physical principles governing the system, and those that are designed to predict future behavior. In analytical applications of chemometric modeling, the utility of the model in making future predictions on unknown samples is most often of paramount importance. Since the physical framework of the model is technically unimportant if prediction is the *sole* motivator, a host of different numerical methods could potentially be employed

with some success. Chemists, however, have the ability to operate on the fence, so to speak—with good theoretical knowledge at their disposal, good predictive models can be constructed that also have physical rationale. It could be said that this knowledge provides the distinction between a chemometrician, and a numerical analyst, or statistician, and oftentimes, the distinction between a useful and meaningful model, and a nonsensical construction of mathematical operations.

The following sections are intended as a concise introduction to the machinery of calibration. The discussion of multivariate calibration modeling will begin with the familiar univariate calibration and extend to multivariate calibration and calibration on principal components. With these technical details reviewed and established, **Section 1.5** will conclude the chapter, moving beyond the particulars and into the important concepts of theoretical analytical chemistry, and multivariate figures of merit for calibration. Although multivariate calibration can easily perform simultaneous multi-component analysis, in the discussions that follow we will focus on the analysis of a single component (the analyte of interest), with the other species simply acting as interferences.

1.2.2 Simple Calibration

Most likely, our first exposure to simple calibration in chemistry came in a lecture or laboratory on the concept of Beer's law, and the interactions between electromagnetic radiation and matter. Beer's law can be expressed mathematically in the well known form

$$r = alc + e \tag{1.1}$$

where a is the molar absorptivity at the measured wavelength, l is the pathlength of the incident radiation, c is the analyte concentration and r is the corresponding instrument response in absorbance units. The measurement error in the observed response is taken to be e. For multiple samples this equation can be expressed in vector notation as

$$\mathbf{r} = a\mathbf{l}\mathbf{c} + \mathbf{e} \tag{1.2}$$

where **r** is now an *m* x 1 column vector of responses corresponding to the *m* samples whose analyte concentrations are contained in the vector **c** (*m* x 1). Each of the responses observed in **r** is assumed to be corrupted by the measurement errors in the vector **e**. In univariate regression and calibration, expressions of this type are most often represented in more general vector notation as

$$\mathbf{y} = \mathbf{x}b + \mathbf{e} \tag{1.3}$$

x is typically referred to as the independent variable (concentrations in the Beer's law scenario, an *m* x 1 vector, *m* is the number of samples), **y**, the dependent variable (instrumental readings, *m* x 1), *b*, the regression parameter, and **e** (*m* x 1) the unobservable error or disturbance terms. (The reader will note that **Equation 1.3** disregards any intercept term that may contribute to the observed response. Without loss of generality, it will be assumed in the following discussions that an intercept (if present) has already been removed by some means.) Traditionally, the information summarized by **Equation 1.3** has been graphically represented as a plot of **y** *vs.* **x**, similar to the one shown in **Figure 1.1**. The process of generating a working model for the relation described in **Equation 1.3** then requires an estimate of the parameter, *b*. This least-squares estimate of *b* can be obtained by the minimization of the sums-of-squares of the residuals (SSR), represented by the objective function

$$f_{obj} = \sum_{i=1}^{m} (y_i - x_i b)^2 = (\mathbf{y} - \mathbf{x}b)^{\mathrm{T}} (\mathbf{y} - \mathbf{x}b) \tag{1.4}$$

which corresponds to solving the following linear equation for the estimate, $\hat{b}$.

$$\begin{aligned} \hat{b} &= (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{y} \\ &= \mathbf{x}^{+}\mathbf{y} \end{aligned} \tag{1.5}$$

The superscript '+' indicates a *pseudoinverse*, so named because the inverse of a non-square matrix does not truly exist. When $\mathbf{x}^{+} = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}$, as above, it is

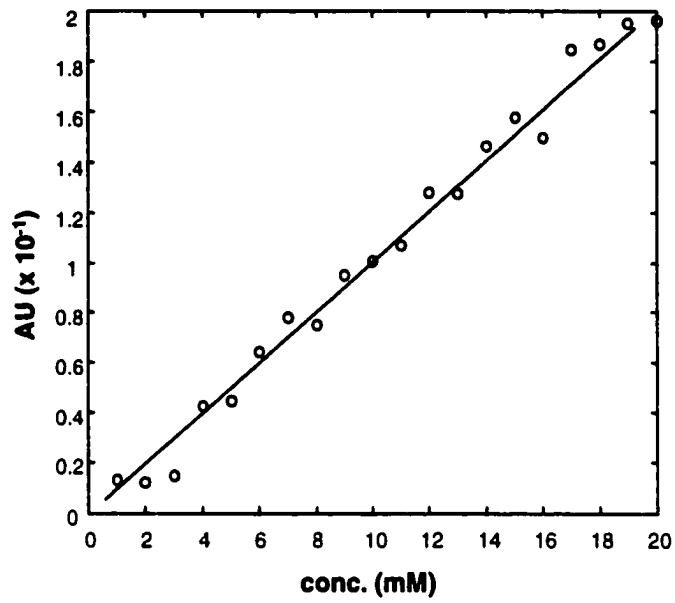


Figure 1.1 A simple scatter plot of y *vs.* x along with a simple first order polynomial fit to the observed data.

referred to as the Moore-Penrose pseudoinverse. Corresponding estimates $\hat{\mathbf{y}}$, and $\hat{\mathbf{e}}$ may subsequently be obtained using

$$\hat{\mathbf{y}} = \mathbf{x}\hat{b} \tag{1.6}$$

$$\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} \tag{1.7}$$

While the scatterplot in **Figure 1.1** is the traditional means of representing the concepts of univariate calibration, several crucial attributes of the method are obscured in this portrayal which will become instrumental in more complicated multivariate calibration schemes.

The characteristic of the scatterplot shown in **Figure 1.1** that makes it useful, namely that it yields information about the relationships of the samples, also severely limits it, since this 'variable space' representation often obscures information about the relationship of the variables $\mathbf{x}$ and $\mathbf{y}$. An alternative representation of the data in 'sample space' uses the samples as the axes of the plot, rather than the variables. The model from **Equation 1.3** then indicates that

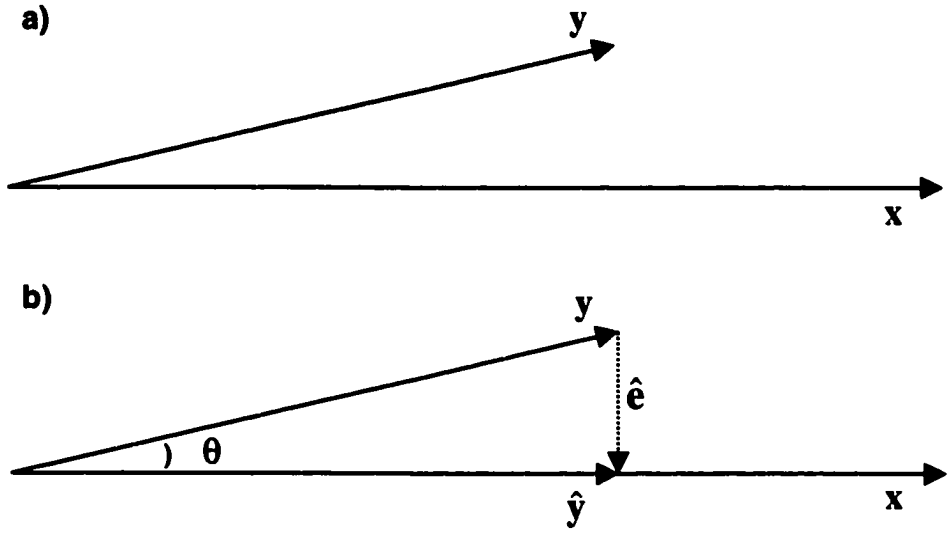


Figure 1.2 Sample space representation of the standard univariate linear model. **a)** the two vectors **x**, and **y** oriented in sample space, and **b)** the orthogonal projection of **y** onto the space defined by **x**.

the variable vector **y** should simply be a scalar multiple of the vector **x**, *i.e.*, the two vectors should point in essentially the same direction in sample space. Because it is presumed that **y** is corrupted with measurement errors, it is invariably displaced from being exactly colinear with **x**. This is illustrated in **Figure 1.2a**. Since the least-squares solution for b, $\hat{b}$ yields a minimum for the objective function in **Equation 1.4**, it must also minimize the length of the error vector, $\hat{\mathbf{e}}$.

$$f_{obj} = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 = (\mathbf{y} - \hat{\mathbf{y}})^{\mathrm{T}} (\mathbf{y} - \hat{\mathbf{y}}) \tag{1.8}$$

$$f_{obj} = \sum_{i=1}^{m} (\hat{e}_i)^2 = \hat{\mathbf{e}}^{\mathrm{T}} \hat{\mathbf{e}} = \|\hat{\mathbf{e}}\|^2 \tag{1.9}$$

$\|\cdot\|$ denotes the Euclidean norm (length) of the vector. Therefore, the least-squares solution in sample space is very clearly the solution that yields the shortest $\hat{\mathbf{e}}$ vector from the tip of **y** to any point on **x**. The shortest $\hat{\mathbf{e}}$ vector will

uniquely result when $\hat{\mathbf{e}}$ is orthogonal to $\mathbf{x}$, making $\hat{\mathbf{y}}$ apparent as the orthogonal projection of $\mathbf{y}$ on $\mathbf{x}$, and $\hat{b}$ the scalar that satisfies **Equation 1.5**. A reexamination of **Equation 1.6** makes this all the more evident, since $\mathbf{x}\mathbf{x}^+$ is an orthogonal projection matrix.

$$\hat{\mathbf{y}} = \mathbf{x}\mathbf{x}^+\mathbf{y} \tag{1.10}$$

These concepts are all presented in **Figure 1.2b**. Other standard calibration measures are also easily visible in this representation. The correlation coefficient, R, is the cosine of the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$, and the familiar F-statistic is proportional to the ratio of the squared length of $\hat{\mathbf{y}}$ to the squared length of $\hat{\mathbf{e}}$. (The constant of proportion is the ratio of the degrees of freedom of $\hat{\mathbf{e}}$ to $\hat{\mathbf{y}}$.)

While the simplicity of univariate calibration is an attractive attribute, there are several fundamentally limiting properties of the approach. Univariate calibration methods naturally require full selectivity for the analyte of interest. Interferences can therefore only be handled in the rather naïve case in which the amount of the interferent is constant in all calibration and prediction samples. This severe limitation mathematically precludes doing calibration in the presence of interferences, and simultaneous multicomponent analysis. A host of advantages are to be realized if one moves from the univariate realm of single measurements into the realm of multiple measurements, or, multivariate calibration.

1.2.3 Multivariate Calibration

1.2.3.1 Inverse versus classical calibration

In previous sections it was assumed that the vector of instrumental measurements, $\mathbf{y}$, was to be projected onto the vector of concentrations, $\mathbf{x}$, to extract the least-squares solution for the regression parameter. The projection is often performed in this manner because it is assumed that the errors in the

absorbance measurements, **y**, are substantially larger than the errors in concentrations, **x**, and that the models as given in **Equations 1.2** and **1.3** are sound approximations. It is also possible to 'invert' this classical representation and make the absorbance measurements **x**, and the concentrations **y**, implying the errors in the concentration values are significantly larger than the response errors. In the chemometrics literature, this modeling approach is called *inverse calibration*. In contrast to the more classical calibration setup outlined in the previous section, univariate inverse calibration uses least-squares to project the vector of concentrations (now **y**) onto the vector of responses (now **x**). In univariate calibration there are, in truth, only minor differences between classical and inverse calibration, however in multivariate calibration, the distinction becomes important in both theoretical and practical considerations.

While **Equation 1.2** gave an extension of Beer's law when multiple samples are involved, it is additionally possible to express Beer's law with multiple wavelength measurements. For a single-component system with *m* samples, this relation becomes

$$\mathbf{R} = \mathbf{c}\mathbf{s}^{T} \tag{1.11}$$

where **s** is an *n* x 1 vector of the pathlength-normalized molar absorptivities for each of *n* wavelengths, and **R** is now an *m* x *n* (samples x wavelengths) matrix of absorbances, with each row corresponding to a spectral measurement for a different sample. The matrix of spectra arises from the simple outer product of a the concentration vector, and the pure-component spectral vector for the component. (Although **R** and **c** will be referred to as matrices of spectra and concentrations, this is certainly not a necessity. **R** may well represent voltammograms, and **c** refractive indices. To avoid unnecessary abstractions, however, they will be referred to as spectra and concentrations without loss of generality.)

If *p* different spectroscopically active components are assumed to be present in the mixtures, then **Equation 1.11** can be trivially extended to

$$\begin{aligned}\mathbf{R} &= \mathbf{c}_1\mathbf{s}_1^T + \mathbf{c}_2\mathbf{s}_2^T + \cdots + \mathbf{c}_p\mathbf{s}_p^T \\ &= \mathbf{C}\mathbf{S}^T\end{aligned} \tag{1.12}$$

where **C** (*m* x *p*) is the matrix of concentrations (columns representing the concentrations of each of the *p* components), and **S** is the matrix of pure-component spectra at unit concentration. The response matrix will still be *m* x *n*, however it will now result from the spectral contributions of all *p* components. Of course, in the classical calibration scenario concentrations and responses are known, and so calibration modeling in this case involves solving **Equation 1.12** for an estimate of **S**, $\hat{\mathbf{S}}$ (a process referred to as *indirect* calibration). If **S** happens to be known, then calibration can proceed immediately without estimation (*direct* calibration). This classical multivariate calibration model, referred to in the chemometrics literature as classical least-squares (CLS), is extremely powerful when all of the requisite information is known, *i.e.*, we know **R** (measured spectra for each mixture sample), and we have access to the concentrations of every spectrally active component in the calibration samples, **C** (which must include all components which may be in future samples). This calibration scenario is illustrated in **Figure 1.3**. Oftentimes, however, we wish to calibrate a particular analyte in the absence of knowledge of the other interfering components in the system, which is an impossibility in this classical multivariate calibration approach. Additional drawbacks of this classical method are mathematical in nature, namely that the inverses of ($\mathbf{C}^T\mathbf{C}$) and ($\mathbf{S}^T\mathbf{S}$) must exist.

The form of **Equation 1.12** suggests that we could also write

$$\mathbf{C} = \mathbf{R}\mathbf{B} \tag{1.13}$$

which is the inverse calibration expression for the proposed model. As in **Section 1.2.2**, we will revert to the more general notation to distinguish the inverse approach from the CLS approach. (Again, **X**, or **x** will be referred to in terms of spectra, and **Y**, or **y** as concentrations without loss of generality) Thus, we have

$$\mathbf{Y} = \mathbf{X}\mathbf{B} \tag{1.14}$$

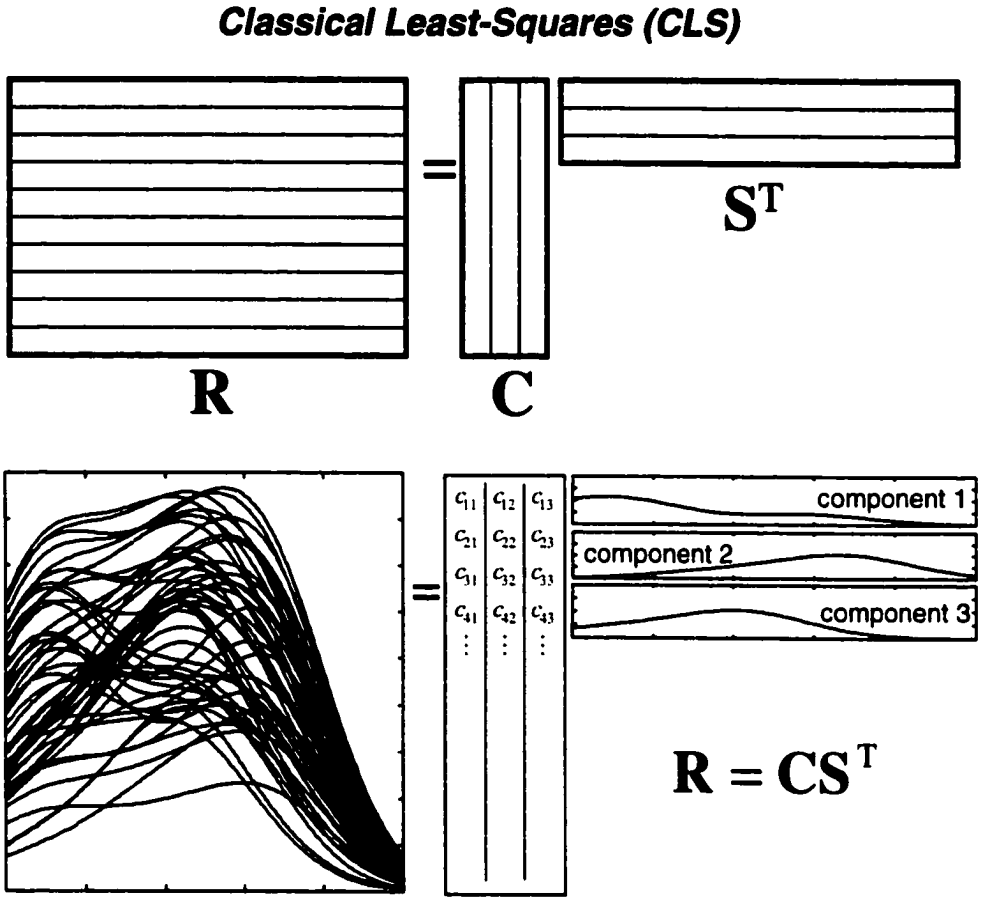


Figure 1.3 An illustration of the general layout of classical least-squares methods.

where **Y** is the *m* x *p* matrix of concentrations, **X** is the *m* x *n* matrix of instrument responses, and **B** is the *n* x *p* matrix of regression coefficients mapping **X** to **Y** (**Figure 1.4**). In inverse regression, the structure of the model itself relaxes the need for complete component knowledge, since **Equation 1.14** is not chemically bilinear in origin. It is therefore possible to restrict interest to whichever analytes are desired without regard for the other analytes active in the mixtures.

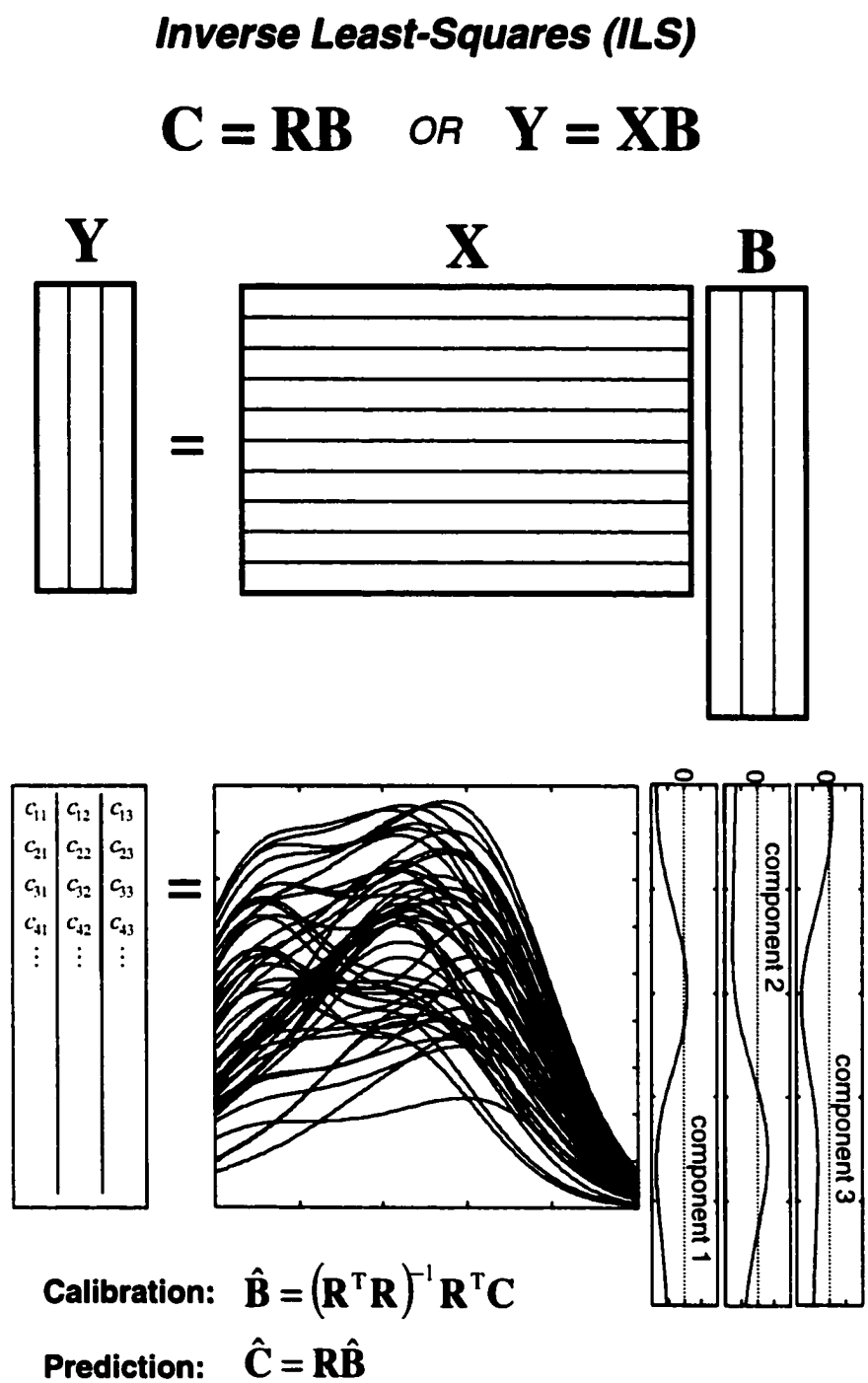


Figure 1.4 An illustration of the general layout of inverse least-squares methods.

This allows accurate calibration and quantitation in the presence of a host of interfering species. The pure-component spectra need not be known, and the concentrations of the interfering species in the mixture samples are also unnecessary.

1.2.3.2 Inverse Multivariate Calibration

In inverse multivariate calibration (hereafter referred simply to as multivariate calibration), the least-squares solution for the calibration of a single analyte of interest in a mixture of other interfering species is given by

$$\hat{\mathbf{b}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \tag{1.15}$$

where, as above, $\mathbf{y}$ is the vector of concentrations for the analyte, and the rows of $\mathbf{X}$ contain the measured mixture spectra. Like the complimentary univariate expression, **Equation 1.15** is also sometimes expressed as

$$\hat{\mathbf{b}} = \mathbf{X}^{+}\mathbf{y} \tag{1.16}$$

Although it is difficult to express this high-dimensional operation geometrically, it is analogous to the univariate scenario discussed previously. The estimated vector of concentrations, $\hat{\mathbf{y}}$, is the orthogonal projection of $\mathbf{y}$ onto the subspace $S_{\mathbf{X}}$ spanned by the column vectors of $\mathbf{X}$, in a similar fashion to the univariate counterpart in **Equation 1.10**.

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \\ &= \mathbf{X}\mathbf{X}^{+}\mathbf{y}\end{aligned} \tag{1.17}$$

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}} \tag{1.18}$$

An illustration of the multivariate least-squares solution is given in **Figure 1.5** assuming that there are only two spectrally active components in the mixture samples (the analyte, and an interferent).

With the regression parameters estimated from the calibration procedure, predictions for an unknown sample can be easily made via

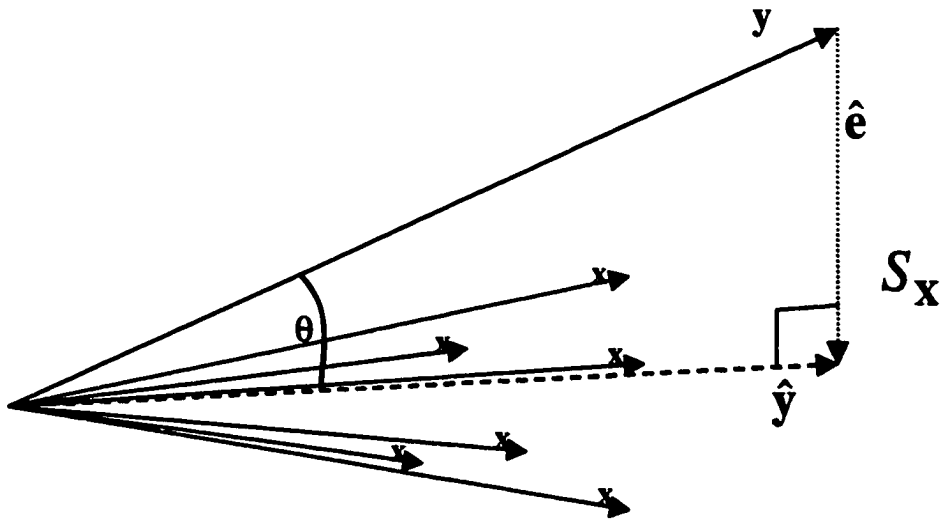


Figure 1.5 An illustration of the least-squares solution with two mixture components. The least-squares solution is the orthogonal projection of **y** onto the space defined by the **x** vectors, S_X.

$$\hat{y}_{unk} = \mathbf{x}_{unk}\hat{\mathbf{b}} \tag{1.19}$$

and, for multiple samples,

$$\hat{\mathbf{y}}_{unk} = \mathbf{X}_{unk}\hat{\mathbf{b}} \tag{1.20}$$

While inverse calibration side-steps the difficulties associated with classical calibration, it suffers from one premier difficulty, namely, the necessity of inverting the matrix ($\mathbf{X}^T\mathbf{X}$) in **Equation 1.17**. For this matrix to be conveniently inverted, the number wavelength channels in the spectral domain must be less than the number of samples used to build the calibration model (*n* must be less than *m* in the dimensions of **X**). With current instrumentation and spectrometers automatically yielding hundreds (or thousands) of measurements in the wavelength domain, standard inverse calibration in this fashion would require hundreds or thousands of calibration samples, a demand that is often prohibitively costly. An additional contributor to the inversion problems with $\mathbf{X}^T\mathbf{X}$

is the natural colinearity that is often observed in $\mathbf{X}$. It can be ascertained from **Equation 1.12** that, if the response matrix is generated by an assumed underlying system of p independent components, $\mathbf{X}^{\circ}$ will contain (at most) p independent rows or columns in the absence of measurement errors. In mathematical terminology, $\mathbf{X}^{\circ}$ would be said to be "rank p". Since the contribution of measurement errors will normally make the matrix full rank (min(m,n)), $\mathbf{X}$ is often referred to as "pseudorank p". In any case, the inverse of $\mathbf{X}^{T}\mathbf{X}$ will still be ill-conditioned (numerically unstable) because the matrix is close to being singular. These natural linear dependencies tend to complicate the inversion of the already ill-conditioned matrix $\mathbf{X}^{T}\mathbf{X}$.

These annoyances can be circumvented to some degree using wavelength selection methods in which only k spectral channels are selected to use in the calibration modeling process, where $p \leq k < m$ (and $k << n$). Provided a proper number of wavelengths have been selected, then singularity problems with the inversion of $\mathbf{X}^{T}\mathbf{X}$ are largely avoided, and inverse calibration can proceed without difficulty. In cases in which $\mathbf{X}$ is naturally suited for inverse calibration ($n < m$), or wavelength selection methods are used to precondition $\mathbf{X}$, the calibration procedure is referred as multiple linear regression, or MLR. Unfortunately for the analyst, conventional spectrometers yield data matrices that are not typically suited for MLR, and wavelength selection is a less-than trivial task. As a result, considerable effort has been expended in answering the question of how best to select the wavelength channels for inverse calibration without discarding important information. Although a discussion of these wavelength selection methods is outside of the interests of this work, other numerical methods have proven extremely useful in alleviating the numerical difficulties associated with **Equation 1.17**, without relinquishing the multichannel advantage.

1.2.4 Principal Components Regression

1.2.4.1 Principal Components Analysis

The method of principal components analysis (PCA) was devised by Cauchy in 1829 [6], and formalized by Pearson in 1901 [7]. It was discussed in the chemical literature (although not by name) as early as 1878 [8], although in reference 8 it was conceptualized as a simple method to perform regressions with errors in both an **x** and **y** variable. By the late 60's and early 70's computing technology had evolved sufficiently to make PCA a realistic endeavor in chemical analysis, and consequently, the employment of the method in solving multivariate chemical problems generally dates to this era. Since these early attempts, PCA has proven useful in a variety of applications including mixture analysis, pattern recognition (and classification and discrimination), curve-resolution and multivariate calibration. While PCA is useful in a broad range of chemical problems, at its core, it is simply a method of determining a concise coordinate system for expressing the variations in the data.

When presented with a bivariate data matrix, $\mathbf{X}$ (m x 2), like the one given below, it is standard practice to pictorially represent the structure of the data in a Cartesian coordinate system with the assumed axes of the coordinate system defined by the orthogonal basis vectors [1 0] and [0 1]. In usual fashion, these basis vectors are of unit length. We can envision the data matrix, then as being two sets of scores, $\mathbf{T}$, on these fundamental axes, $\mathbf{V}^T$

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{m1} & x_{m2} \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \\ \vdots & \vdots \\ t_{m1} & t_{m2} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{1.21}$$

or

$$\mathbf{X} = \mathbf{T}\mathbf{V}^T \tag{1.22}$$

where the *i*th column of **T** contains the scores on the *i*th basis vector (the *i*th row of $\mathbf{V}^T$). Since the fundamental axes are typically assumed to be the vectors [1 0] and [0 1] in this bivariate case, it is unnecessary that they be explicitly discussed, and in this case **T** = **X**. It is possible to express the same data as new scores in a different coordinate system, however, such as

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{m1} & x_{m2} \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \\ \vdots & \vdots \\ t_{m1} & t_{m2} \end{bmatrix} \cdot \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \tag{1.23}$$

The scores are now referring to the axes given by the rows of $\mathbf{V}^T$, [0.7071 0.7071] and [-0.7071 0.7071] (a coordinate system oriented at a 45° angle to the standard axes), and they will be substantially different than the scores in **Equation 1.21**. There are, of course an infinite number of coordinate systems in which to express the data; however, some orientations are much more useful than others.

Principal components analysis is one method of generating a new coordinate system, commonly referred to as the loadings or principal components (PC's), and the corresponding scores on those PC's for a data matrix, **X**. This new coordinate system is selected in the following manner:

(1) Define the first PC, PC_1 (first row of $\mathbf{V}^T$) as the direction in the subspace defined by the data matrix, **X**, which can account for the most observed variation.
(2) The second PC must be orthogonal to the first, and be oriented in the direction which can account for the most variance not accounted for by PC_1.
(3) The third PC must be orthogonal to all previously defined loading vectors, and be oriented in the direction which can account for the most variance not accounted for by the other two loadings
(4) Continue until all directions are accounted for.

As an example of this procedure, some bivariate data are tabulated and plotted in **Figure 1.6a** and **Figure 1.6b**, including the orientation of the (two) loading vectors, and the scores on these loadings. Since the data essentially fall on a straight line, a large proportion of the variation in the data can be described using the first PC, as is immediately apparent in **Figure 1.6b**. Indeed, examination of the data in terms of the scores on PC_1 and PC_2 indicates that the scores on the second loading are all considerably smaller than the scores on the first loading. When looking at the data matrix in this frame of reference (the principal component axes in **Figure 1.6b**), then, it is evident that the direction of PC_2 is of minor importance in describing the structure of the data.

Although it cannot readily be appreciated with simple two-dimensional data, PCA is an extremely powerful way to visualize the variations and relationships in high-dimensional data, which are typically difficult to view effectively with standard representations. In addition to this visualization benefit, PCA allows the analyst to discard dimensions of the data that appear to be of little importance. In the example shown in **Figures 1.6a** and **1.6b**, the second PC can be discarded by eliminating the second column of $\mathbf{T}$, and the second row of $\mathbf{V}^T$, and regenerating an approximation to the data matrix using these truncated matrices. To distinguish these truncated matrices from the full matrices, an overstrike "~" will be used. Sometimes a subscript is also used to indicate the number of PC's that have been used in the reconstruction, as in the rank *p* expression,

$$\tilde{\mathbf{X}}_p = \tilde{\mathbf{T}}_p \tilde{\mathbf{V}}_p^T \tag{1.24}$$

We will omit the subscript unless it is necessary for clarity. The geometric impact of the elimination of the second PC in the example given is shown in **Figures 1.7a** and **1.7b** for the standard coordinate axes, and the principal component axes. In both representations, it is clear that the truncation corresponds to orthogonal projections of the data points onto the first PC.

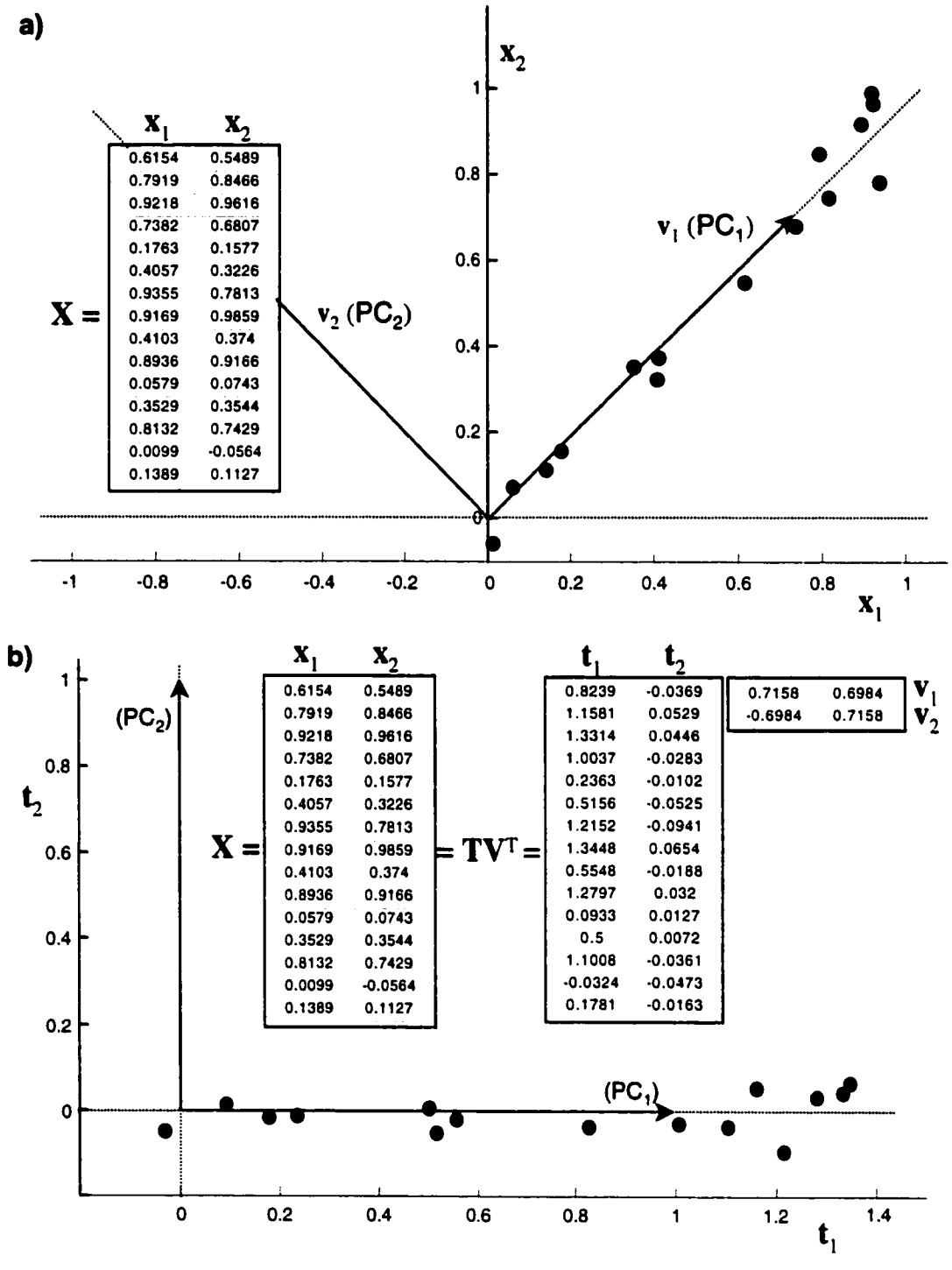


Figure 1.6 **a)** The data in the fundamental coordinate system [0 1, 1 0] with the first and second loadings shown. **b)** The same data represented in the coordinate system defined by the first two principal components. The scores of the data on these axes are also shown.

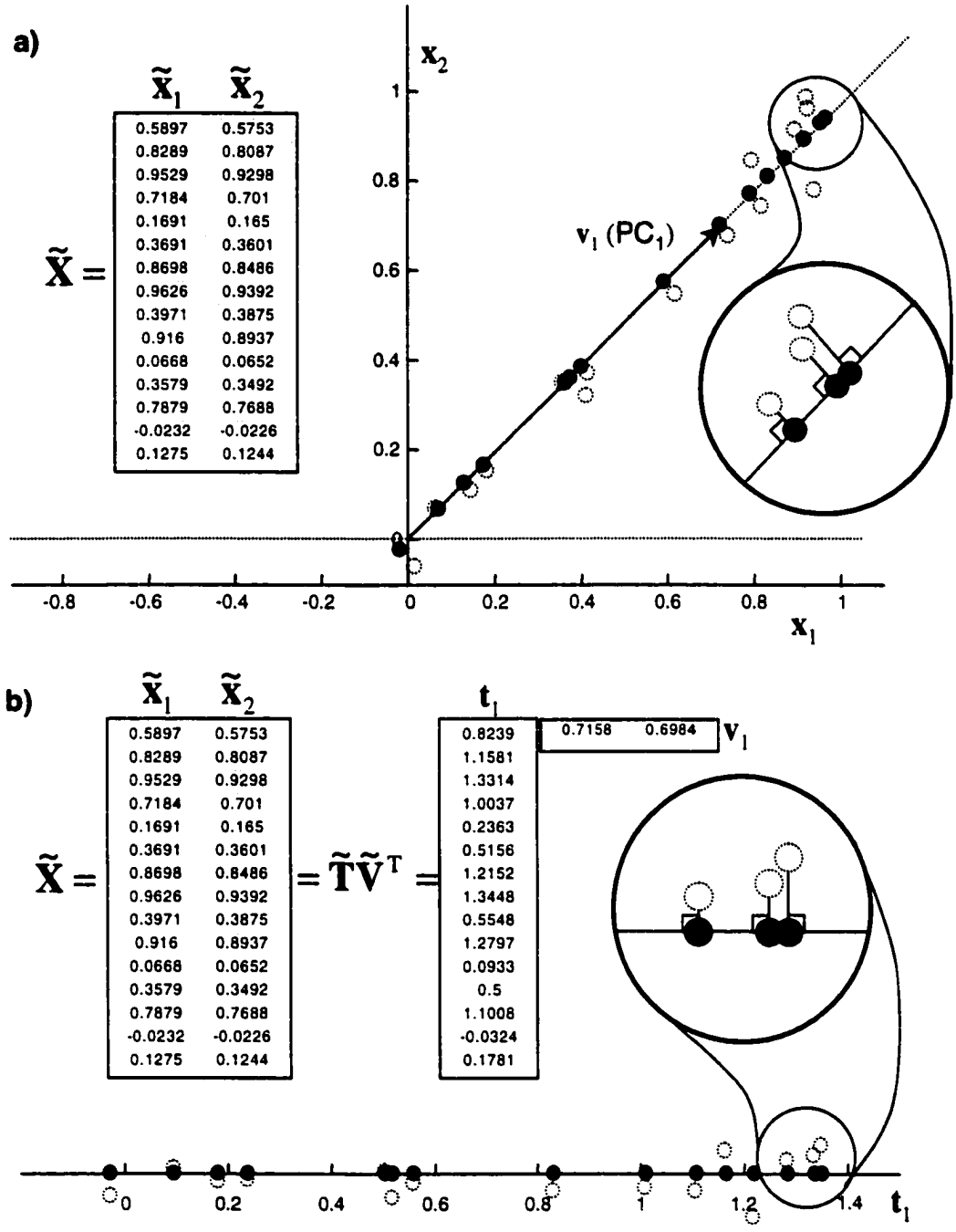


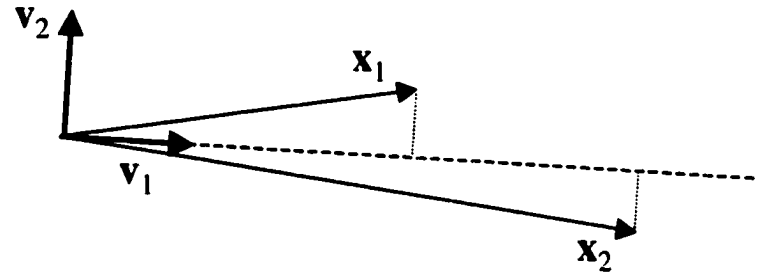
Figure 1.7 **a)** The projected data after discarding in the direction of the second loading. **b)** The same data represented in the (now) unidimensional coordinate system defined by the first principal component.

PCA is also illustrated from a vector perspective in **Figure 1.8** for both variable and subject space representations.

Operationally, principal components analysis can be performed using a variety of numerical methods. Singular value decomposition (SVD) is a very powerful matrix diagonalization procedure which can accommodate rectangular matrices of the sort that are frequently encountered in chemical applications. In SVD notation, the matrix **X** is decomposed as the product of three matrices, **U**, **S**, and $\mathbf{V}^T$.

a) ***variable space - loading vectors***

(x vectors are the rows of **X**)



b) ***sample space - score vectors***

(x vectors are the columns of **X**)

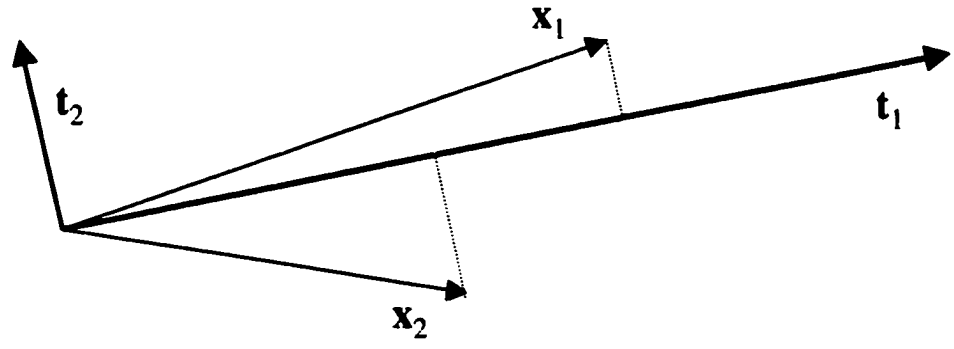


Figure 1.8 **a)** variable representation of the projection of the data onto a rank 1 principal component analysis subspace. **b)** Sample space representation of the same.

$$\mathbf{X} \xrightarrow{SVD} \mathbf{USV}^{\mathrm{T}} \tag{1.25}$$

In the fully general case for an m x n dimensioned $\mathbf{X}$ matrix with $m < n$, $\mathbf{U}$ is m x m, $\mathbf{S}$ is m x m, and $\mathbf{V}^{\mathrm{T}}$ is m x n. The row vectors of $\mathbf{V}^{\mathrm{T}}$ are orthonormal (orthogonal to each other, and of unit length), as are the columns of $\mathbf{U}$, and the matrix $\mathbf{S}$ is a diagonal matrix with decreasing elements down the diagonal ($s_{ii} \geq s_{jj}, i < j$). This decomposition is particularly convenient for PCA, since, in the notation used in **Equation 1.24**, $\mathbf{T} = \mathbf{US}$, and $\mathbf{V}^{\mathrm{T}}$ is $\mathbf{V}^{\mathrm{T}}$. Truncation is easily achieved by discarding the requisite portions of the matrices.

$$\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^{\mathrm{T}} \tag{1.26}$$

As was alluded to above, this truncation operation can also be expressed as a simple orthogonal projection of $\mathbf{X}$ onto the subspace defined by the loadings in $\tilde{\mathbf{V}}^{\mathrm{T}}$.

$$\tilde{\mathbf{X}} = \mathbf{X}\left(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^{\mathrm{T}}\right) \tag{1.27}$$

Considering the representations of the example data in **Figure 1.6a** and **1.7a**, it might appear that the first PC models the data in a manner similar to traditional least-squares regression; however, several important differences exist between these two methods. The most crucial difference is that in least-squares the vector $\mathbf{x}_2$ would be projected onto the vector $\mathbf{x}_1$, since it is implicitly assumed that $\mathbf{x}_1$ is essentially error free, and represents the true model space. In contrast, PCA assumes that the errors corrupting $\mathbf{x}_1$ and $\mathbf{x}_2$ are of approximately the same magnitude, and thus *both* $\mathbf{x}_1$ and $\mathbf{x}_2$ are used to estimate a model space, PC_1 in this case, and estimates of both $\mathbf{x}_1$ and $\mathbf{x}_2$ are achieved via **Equations 1.26** or **1.27**. In variable space, this difference corresponds to a minimization of the sums-of-squares of the residuals in the *vertical* direction for least-squares, and a minimization of the sums-of-squares of the residuals in an *orthogonal* direction in PCA. In the subject space (illustrated in **Figure 1.9**), the error vector in the least-squares sense is necessarily orthogonal to the model space, $\mathbf{x}_1$, while in PCA, the 'error vector' (presumed to correspond to the rather uninformative PC_2) is

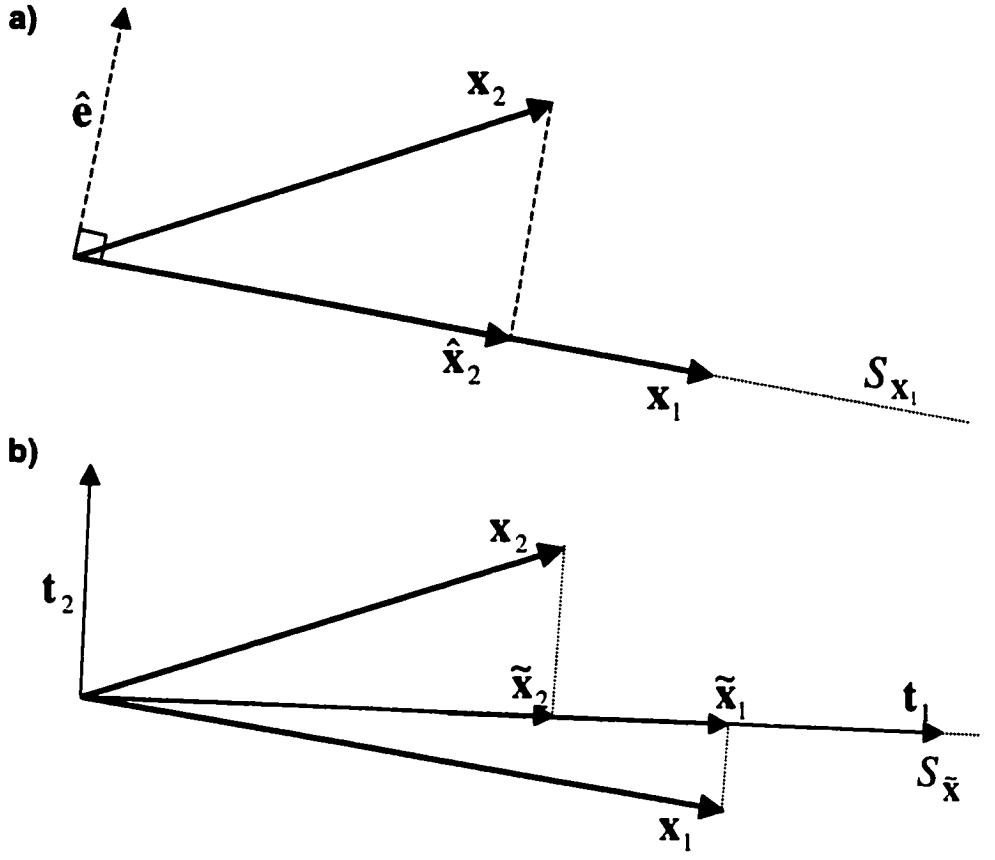


Figure 1.9 Illustration of the fundamental difference between least-squares estimation, and PCA estimation of a one-dimensional model space. **a)** The least-squares approximation under the assumption that $\mathbf{x}_1$ is the true model space. **b)** The PCA approximation under the assumption that both $\mathbf{x}_1$ and $\mathbf{x}_2$ are corrupted by errors — the true model space ($\mathbf{t}_1$) must be estimated.

also orthogonal to the model space, PC_1. While the method is nicely illustrated in this simple bivariate case, its full utility is more readily appreciated in truly multivariate methods. In these cases, a matrix of very high-dimensionality may be reduced to a model of only a few dimensions, and geometrically, points which are spread out over many, many dimensions in space can be approximated by a relatively simple *p*-dimensional subspace which can still describe the important variations in the data, but without using the full, very high-dimensional space.

At the end of **Section 1.2.3.2** it was noted that the colinearity/singularity problem was the bane of full-spectrum inverse calibration, since there are typically more wavelength channels than samples, and natural colinearities contribute to linear dependencies in the data matrices. As we will see, PCA is a very efficient method of reducing the impact of these problems in multivariate calibration, and this use of PCA to stabilize the data matrix in conjunction with inverse calibration is referred to as principal components regression, or PCR.

1.2.4.2 Principal Components Regression

While the discussion of principal components analysis involved aspects of regression, PCA is a procedure performed entirely on a data matrix $\mathbf{X}$, and in that sense, PCA is a regression of $\mathbf{X}$ onto a lower-dimensional estimate of itself. Principal components regression involves a PCA of the instrumental response matrix, followed by an additional regression of the properties of interest onto the PCA truncated instrument responses. The numerical convenience of PCR is derived from the fact that PCA allows the simplification of the variations in the instrumental responses to just a few primary directions (latent variables). Since the $\mathbf{X}$-$\mathbf{y}$ regression can just as easily be performed using these latent variables, the colinearity issues that are problematic in inverse calibration procedures using MLR are remedied. In addition, PCR can take advantage of the full spectrum, achieving a significant degree of signal averaging and eliminating the need for variable selection routines.

An outline of PCR follows for a single component property of interest (a univariate $\mathbf{y}$), although it is trivially extendable to simultaneous multi-component calibration and prediction. With a principal component analysis of the instrumental responses determined (**Equation 1.25**), the desired number of principal components, p, can be selected giving $\tilde{\mathbf{X}}$. In order to regress $\mathbf{y}$ onto the space of $\tilde{\mathbf{X}}$, an orthogonal projection for $\mathbf{y}$ is needed as in **Equations 1.14** and **1.15**. This is readily achieved using the convenient properties of $\mathbf{U}$ and $\mathbf{V}$, and thus

$$\tilde{\mathbf{X}}^{+} = \tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^{T} \tag{1.28}$$

The orthogonal projection of $\mathbf{y}$ in PCR, then, is simply

$$\begin{aligned}\hat{\mathbf{y}} &= \tilde{\mathbf{X}}\tilde{\mathbf{X}}^{+}\mathbf{y} \\ &= \left(\tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^{T}\right)\left(\tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^{T}\right)\mathbf{y}\end{aligned} \tag{1.29}$$

Using the orthonormal properties of $\tilde{\mathbf{V}}$ and the fact that $\tilde{\mathbf{S}}$ is diagonal, **Equation 1.29** reduces to

$$\hat{\mathbf{y}} = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^{T}\mathbf{y} \tag{1.30}$$

Similar to **Equation 1.16**, the PCR regression vector estimate is given by

$$\hat{\mathbf{b}} = \tilde{\mathbf{X}}^{+}\mathbf{y} \tag{1.31}$$

$$\hat{\mathbf{b}} = \tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^{T}\mathbf{y} \tag{1.32}$$

As noted above, this in effect performs the regression of $\mathbf{y}$ on $\mathbf{X}$ using only the scores, since we could just as easily choose to define another form of the estimated regression vector, $\hat{\mathbf{p}}$, as

$$\hat{\mathbf{p}} = \tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^{T}\mathbf{y} = \tilde{\mathbf{T}}^{+}\mathbf{y} \tag{1.33}$$

As in MLR, with the regression parameters estimated, new unknown samples can be predicted from the simple formulae,

$$\begin{aligned}\hat{y}_{unk} &= \mathbf{x}_{unk}\hat{\mathbf{b}} \\ \hat{\mathbf{y}}_{unk} &= \mathbf{X}_{unk}\hat{\mathbf{b}}\end{aligned} \tag{1.34}$$

Or, using the scores formulation of **Equation 1.33**,

$$\hat{\mathbf{y}}_{unk} = \tilde{\mathbf{T}}_{unk}\hat{\mathbf{p}} \tag{1.35}$$

While our discussions of least-squares regression, MLR and PCA/PCR have involved measurement errors, the reader has most likely noted that discussions of the error assumptions underlying these statistical methods were avoided. This may appear to be careless error, however the omission was

intentional. In many ways, the order in which standard calibration theory and measurement errors are addressed in this chapter mimics the low priority typically given to the characteristics of the noise in practice. Unfortunately, the properties of the measurement errors can have an appreciable influence on the validity and performance of multivariate calibration models. The subsections that follow contain a discussion of the nature of measurement errors and their respective characteristics, and most importantly, how these characteristics can be expected to impact multivariate calibration models and their utility.

1.3 Noise Considerations in Regression

1.3.1 Characteristics and Representations of Noise

The precision of every instrumental measurement is hampered by the presence of measurement errors, which are often referred to with the generic term "noise". Noise can be broadly defined as any undesirable variations in a measured signal which obscure the measurement of the signal of interest – the true signal. Under this sweeping generality, noise may be introduced by everything from the sample presentation system, to other sensor-active chemical species in the sample. This definition will be narrowed somewhat for this work, and a distinction will be made between unwanted chemical variations in the signal ("chemical noise" arising from phenomena such as chemical interferences, sampling uncertainty, *etc.*), and noise attributable to non-chemical variations. Hereafter, "noise" will refer to the latter. The principal reason for disregarding chemical noise in subsequent discussions is that it can be negotiated reasonably well using multivariate calibration, an advantage discussed previously in **Section 1.2**.

Some notational remarks are in order, which should be considered as an addendum to those outlined in **Section 1.1**. The term "signal" will be used to refer to a measurement sequence which consists of a pure or undistorted signal corrupted by noise. This signal is implicitly measured as a function of some other

ordinal variable such as wavelength, or time. In the evolution of signal processing, the traditional ordinal variable was time. In order to maintain generality, then, references to the signal in the "time domain" will refer to the original measurement sequence even if the ordinal variable is not time. Likewise, the inverse domain of the ordinal variable will be referred to as the frequency domain, and representations of the signal in this domain will correspond to a Fourier transform of the signal in the time domain.

Given the complexity of modern instrumentation, it is obvious that measurement errors can arise from a plethora of sources and have a correspondingly complex range of properties and characteristics. In the interest of succinctness, only a few of the more general classifications of noise and their associated attributes will be discussed. In doing so, it is helpful if we imagine a discretely sampled signal vector, $\mathbf{x}$ (*e.g.*, a spectrum or chromatogram), from which we could subtract the pure signal component, $\mathbf{x}^{\circ}$ leaving only the noise, $\mathbf{e}$, as shown below by equation, and depicted in **Figure 1.10**.

$$\mathbf{e} = \mathbf{x} - \mathbf{x}^{\circ} \tag{1.36}$$

The properties of this error vector are, as we shall see, of fundamental importance in multivariate calibration. The following discussions in this section outline some of the more common properties of certain types of measurement errors.

1.3.1.1 Measurement Error Attributes

The noise in an instrumental signal can be classified in any number of ways, the most common of which include (1) its source, (2) its distribution, (3) its characteristics in the frequency domain, and (4) its characteristics in the time domain. Unfortunately, classifications based on these methods are not all mutually exclusive. Instead, we will strive to classify noise as being either independent and identically distributed normally (*iid*), or non-*iid*, since these two classifications are particularly relevant in multivariate calibration; however, we will frequently refer to some or all of the four considerations above.

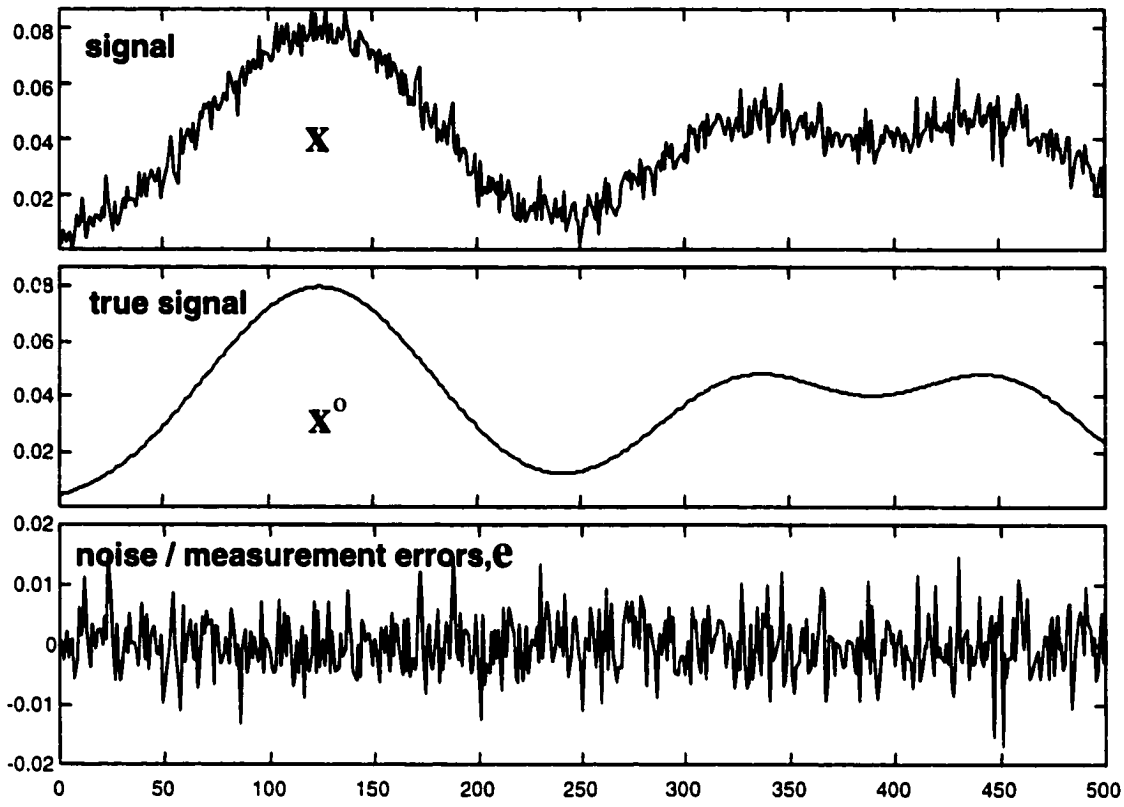


Figure 1.10 An observed signal, $\mathbf{x}$, which can be considered to be the true signal, $\mathbf{x}^o$, corrupted by measurement noise $\mathbf{e}$.

The term "*iid* noise" conveys a lot of information. The concept of *independence* with regard to measurement errors implies that the error observed at any one channel in the signal vector is independent or unrelated to the error observed at any other (different) channel in the signal vector. Independence in the measurement errors is also implied if the measurement errors are *uncorrelated*. *Identically distributed* implies a homogeneity in the error variance across all channels in the signal vector; *i.e.*, the error variance at every channel in the signal vector is the same. The terms *homoscedastic* and *heteroscedastic* are also often used to indicate whether measurement errors are identically distributed, or not. The *normal* condition simply refers to the normal distribution often assumed for the noise observed at one channel in the signal vector over many repeat measurements. Thus, measurement errors are said to be *iid*-

normal if all of the above conditions are met, and non-*iid-normal* if any of these conditions are violated. Assumptions regarding normality are seldom significantly violated to be of concern, and therefore, future references to *iid-normal* errors will often be shortened to simply "*iid*".

Instrument noise is frequently categorized based on its dominant source, which often implies certain distributional and frequency characteristics. *Johnson* or *thermal noise* originates from the thermal agitation of electrons or other charge carriers in resistive elements in the instrument. It is typically classed as fundamental noise, since it does not arise as a result of instrument or component deficiencies, and can never be totally eliminated. It is ubiquitous in resistive elements whether they are carrying current or not. *Shot*, or *quantum noise* is also fundamental, and arises from the random statistical nature of discrete events, since the rate at which they occur is subject to statistical fluctuations. The magnitude of shot noise is typically much lower than that of Johnson noise, although this may not be the case in measurements based on a very low number of events, such as molecular fluorescence. $1/f$ or *pink noise* can arise from a variety of sources and is recognizable by an inverse proportionality between the magnitude of the noise fluctuations and the frequency of the signal (f) being observed (see below). It is commonly considered nonfundamental noise and generally arises from longer term (lower frequency) flicker or drift of instrumental components. For this reason the term $1/f$ noise is often considered synonymous with *flicker noise*, and *drift noise*. Other lesser known noise sources include such things as detector noise, read-out noise, quantization noise, and noise from environmental sources (interference noise).

The distribution of the noise can be observed if a histogram of the noise magnitudes for a large number of measurements can be obtained. By far the most common noise distribution (assumed or measured) for analytical measurements is the normal, or Gaussian, distribution. The reason for this normal distribution is the Central Limit Theorem which, simply put, states that if a measurement is the sum of a series of values drawn from arbitrary distributions,

the distribution of the measurement will approach a normal distribution as the length of the series approaches infinity. Since, in an analytical instrument, the observed noise is a consequence of many smaller random events, the Central Limit Theorem can be rationalized to hold. Other noise distributions (*e.g.* uniform, log-normal) are also observed but are much less common. One other type which is common, however, is the Poisson distribution, which is observed in cases where the signal arises from a collection of discrete events, such as photons striking a photomultiplier tube (resulting in shot noise). In practice, however, the Poisson distribution can simply be considered a special case of the normal distribution (with the mean equaling the variance), and questions regarding the distributional assumption of normality in regression are, as a consequence, rarely of significance.

The frequency domain representation of the noise corrupting a signal is usually referred to as the noise power spectrum (NPS), and it conveys very important information about the time domain correlations of the noise. From a frequency perspective, noise may either be classified as *white noise*, or *coloured noise*. White noise, by analogy to white light, contains equal contributions in noise power at all frequencies, and thus the NPS tends to look flat. Johnson and shot noise are examples of white noise in the frequency domain. In contrast, coloured noise is characterized by the dominance of particular frequencies in the NPS. Perhaps the most prevalent type of coloured noise is $1/f$ noise as outlined above, although additional noise types, such as interference noise may also contribute to colour in the NPS. Myriad origins for $1/f$ noise are possible, including everything from source lamp flicker and ripple voltage in particular instrument components, to temperature fluctuations in the laboratory. Examples of the NPS for samples of white noise, and $1/f$ noise are shown in **Figure 1.11a** and **b**. Of course in reality, instrument noise is the superposition of many different types of noise, and thus the NPS will reflect the relative contributions of all noise sources with their associated characteristics.

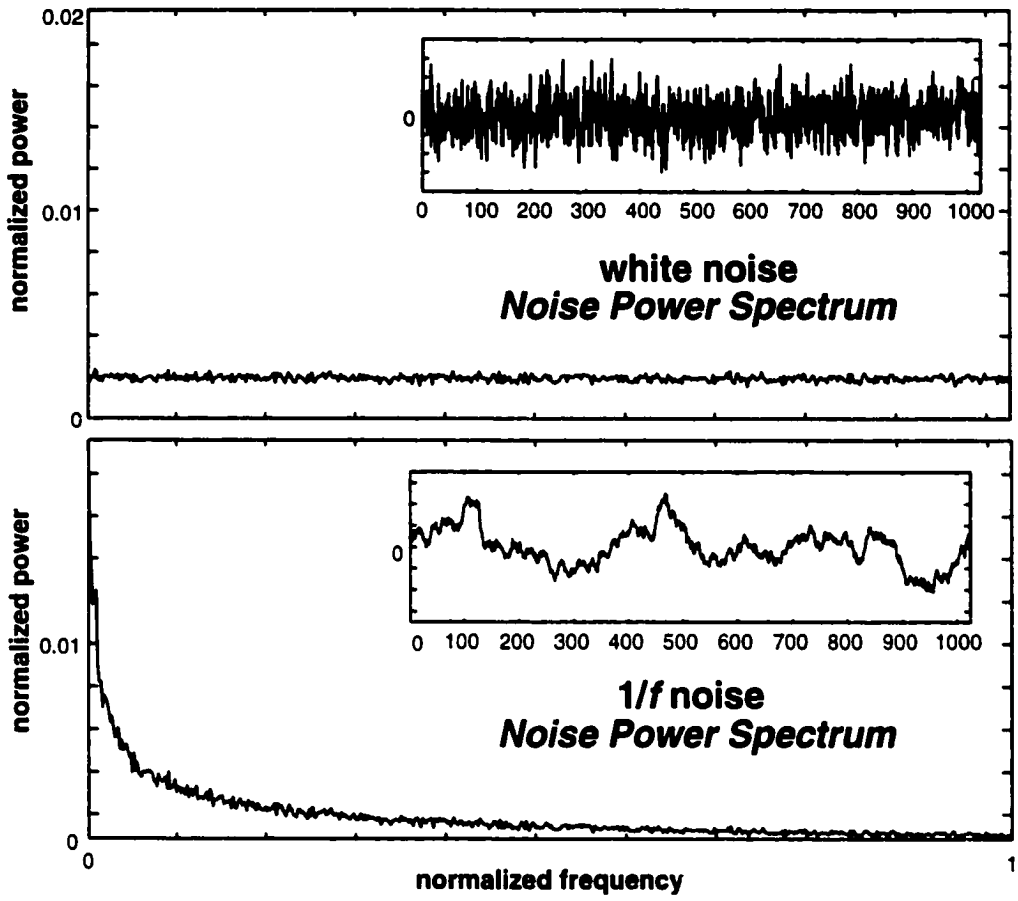


Figure 1.11 Noise power spectra (NPS) of **a)** white noise, and **b)** 1/f noise estimated from 50 repeat samples of the noise. Also shown are samples of the noise in the time domain. The power spectra have been normalized to a total power of 1.

The NPS is extremely useful in giving insight regarding the condition of independence of the measurement errors. White noise in the NPS corresponds to uncorrelated (independent) measurement errors, while coloured noise manifests itself in the time domain as a dependence among the errors (error covariance). Unfortunately the examination of the NPS to ascertain the dependence/independence in the measurement errors is at best only semi-quantitative, since the frequency domain yields no information about the localization of the error correlation structure in the time domain.

1.3.1.2 Error Variance and Covariance Representations

Time domain analysis of the error covariance structure can reveal time-localized characteristics of the measurement errors, and the structural correlations in the noise are much more transparent. The estimated measurement error variance for the first channel in the noise vector $\mathbf{e} = [e_1 \quad e_2 \quad \cdots \quad e_n]$ is given by the well-known formula

$$s_1^2 = \frac{\sum_{i=1}^{m} (e_{1,i})^2}{m-1} \tag{1.37}$$

where m is the number of repeat measurements used in the estimate, or alternatively

$$s_1^2 = \frac{\sum_{i=1}^{m} (e_{1,i})(e_{1,i})}{m-1} \tag{1.38}$$

While this equation gives a quantitative estimate of the magnitude of the error variance at channel one, it says nothing about the relation of the errors at channel one to errors at another channel. This correspondence can be quantified by calculating the *error covariance*, given for channels 1 and 2 by

$$s_{12} = \frac{\sum_{i=1}^{m} (e_{1,i})(e_{2,i})}{m-1} \tag{1.39}$$

where the summation product now includes the errors at different channels (1 and 2) over the m samples. The error covariance term is positive when the errors at channels i and j are correlated, negative when the errors are anticorrelated, and zero when the errors are independent of one another. The calculation of variances, and error covariances for every channel in a signal vector, then, allows one to map the structure of the variations in the measurement errors, and how they are correlated between channels. This structure is conveniently summarized in an error covariance matrix, Σ, a mapping

of the variance and covariance of the measurement errors, which has the general form for an *n*-channel signal vector of

$$\Sigma = \begin{bmatrix} s_1^2 & s_{12} & s_{13} & \cdots & s_{1n} \\ s_{21} & s_2^2 & s_{23} & \cdots & s_{2n} \\ s_{31} & s_{32} & s_3^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & \cdots & s_n^2 \end{bmatrix} \tag{1.40}$$

Since $s_{ij} = s_{ji}$, error covariance matrices are necessarily symmetric.

For *iid* noise, the error covariance terms should approach zero in the expectation, and if the noise is identically distributed, then all of the diagonal terms (variances) should be approximately the same. Expressed in another way using all of the elements of the *n*-element vector of errors, **e**,

$$\Sigma = E\left(\mathbf{e}\mathbf{e}^{\mathrm{T}}\right) = \sigma^2 \cdot \mathbf{I}_n \tag{1.41}$$

where E indicates the expectation value of the quantity in brackets. This structure arises because, in the *iid* case, the expectations of the individual variances, and covariances are

$$E\left(e_i^2\right) = E\left(e_i e_i\right) = \sigma_i^2 \tag{1.42}$$

$$E\left(e_i e_j\right) = \sigma_{ij} = 0 \tag{1.43}$$

Under *iid* noise conditions, then, the error covariance matrix should be diagonal. Deviations from the *iid* condition have easily recognizable influences on the error covariance matrix. The loss of the independence condition corresponds to error covariance terms being significantly different than zero, and Σ deviates from the diagonal form of **Equation 1.41**. Heteroscedasticity (loss of identical distribution at all channels) is characterized by unequal diagonal elements in Σ. Examples of error covariance matrices estimated from 100 replicate measurements for *iid*, and non-*iid* noise are given in **Figure 1.12.** (An associated noise sample is also shown.)

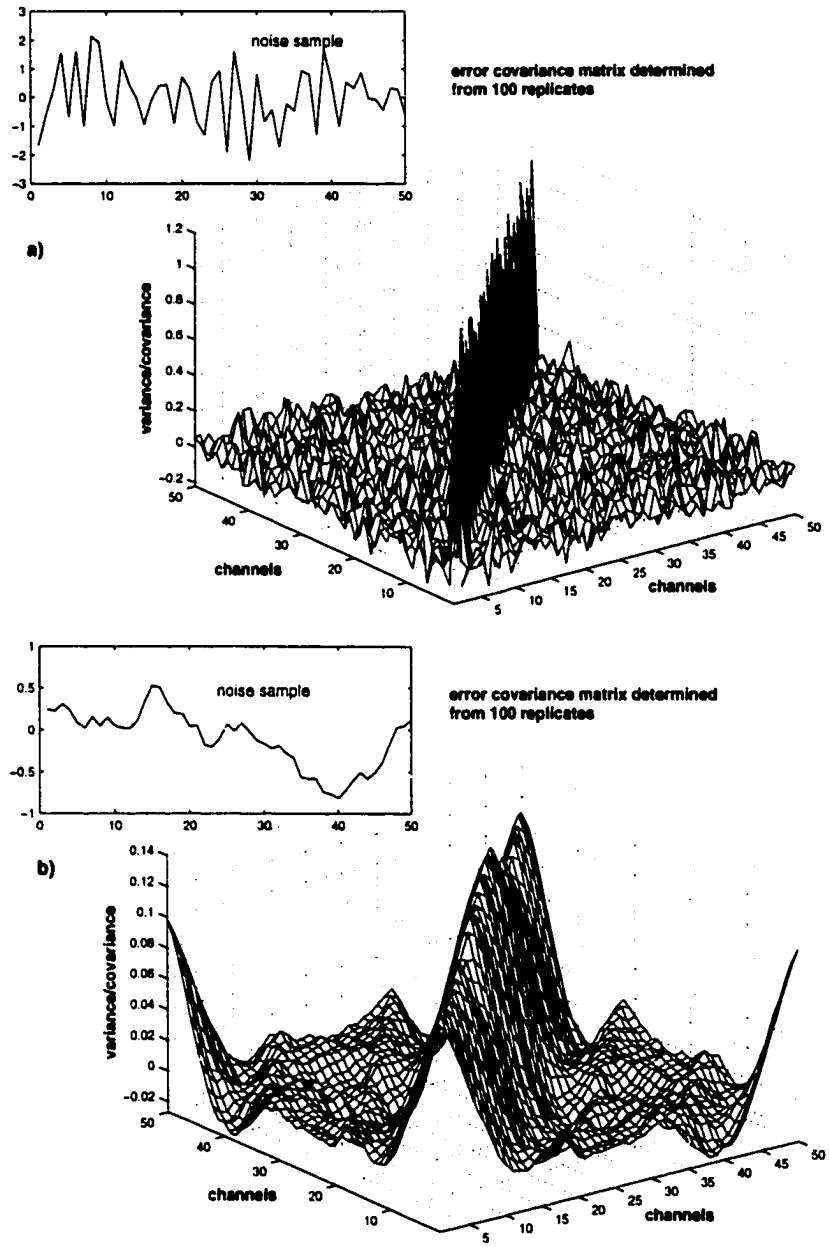


Figure 1.12 Examples of error covariance matrices for measurement errors which are **a)** *iid*, and **b)** non-*iid*. The *iid* errors are characterized by an error covariance matrix which is diagonal, and a multiple of the identity matrix, whereas the non-*iid* errors are characterized by heteroscedasticity (non-equal diagonal elements) and/or correlated error (non-zero off-diagonal terms). Samples of the noise vectors which possess the indicated error structures are also shown (inset).

1.3.1.3 The Geometry of Measurement Errors

Sections 1.3.1.1 and **1.3.1.2** have provided a broad overview of some measurement error attributes, and several useful characterization methods, however it is essential for later theoretical discussions that the mathematical implications of measurement error structure be considered. Since the mathematics of calibration theory is largely linear algebra, and linear algebra is largely geometry, we will pursue geometric interpretations of the two measurement error classes of interest to us: *iid* and non-*iid*. Once again, we will rely on the concept of a true signal vector, $\mathbf{x}^{\circ}$, and its observable counterpart, $\mathbf{x}$, which has been corrupted with noise, $\mathbf{e}$.

A vector of noise which arises under *iid* conditions will demonstrate the following behavior. Because the variance is equivalent in magnitude for every element in the vector (and hence, every dimension in space), the direction of the vector $\mathbf{e}$ will be completely random (around a null vector $\mathbf{0}$). The uncertainty associated with the true vector $\mathbf{x}^{\circ}$ will actually be a multivariate normal distribution which is perfectly symmetrical about $\mathbf{x}^{\circ}$ (provided the errors are *iid*). This is illustrated in **Figure 1.13a**. If the condition of identically distributed errors is relaxed, then the multivariate normal distribution associated with $\mathbf{x}^{\circ}$ can be stretched along the principal axes of the coordinate system, as is shown in **Figure 1.13b**, since it is no longer necessary that the variation in direction *i* be equivalent to the variation in direction *j*.

The final relaxation that can be made from *iid* conditions (while still retaining normality as a distributional assumption) is of course the allowance of error variance *and* covariance. Error covariance has the potential to introduce a directional skew in the multivariate gaussian distribution, since with positive error correlation on channels *i* and *j*, a positive error on channel *i* means that the error on channel *j* will also tend to be high. This sort of behavior is shown geometrically in **Figure 1.13c**.

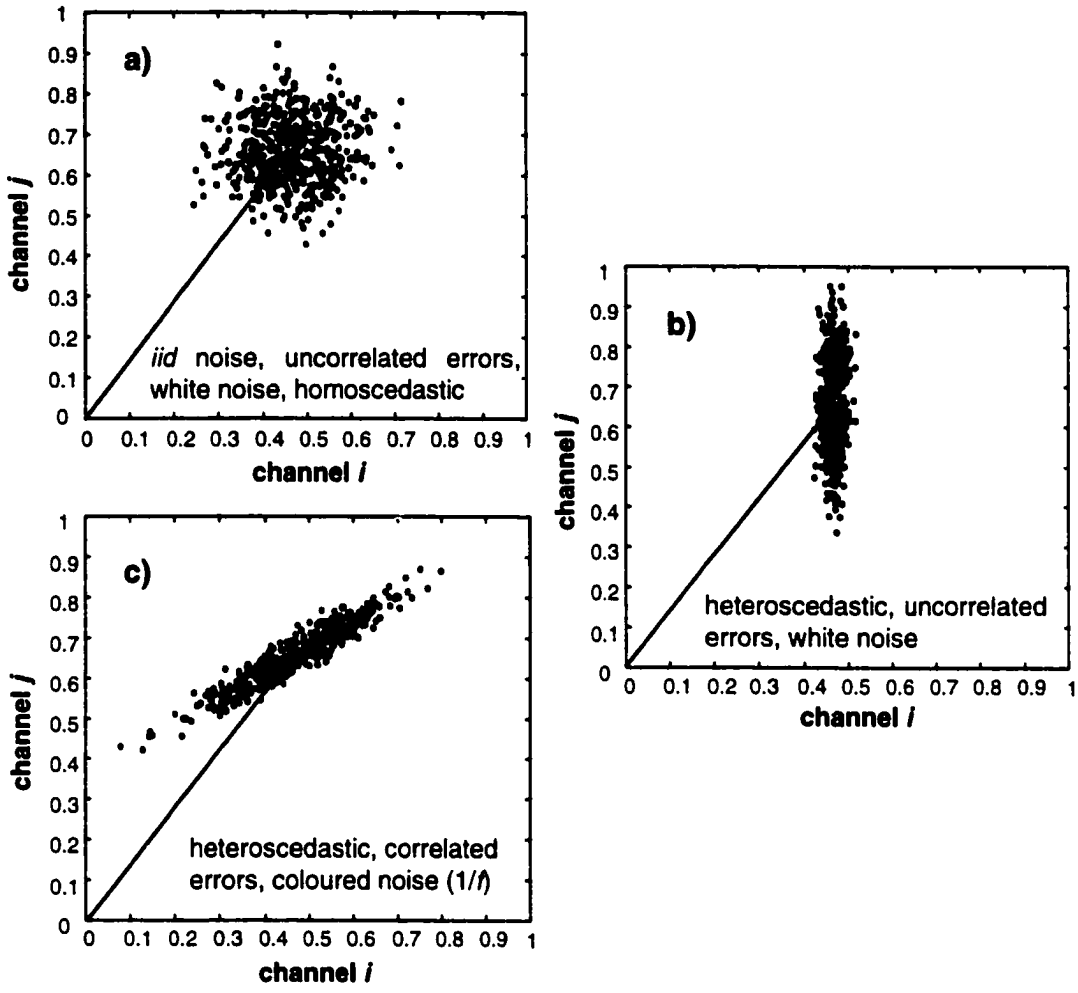


Figure 1.13 Illustration of the geometry of measurement errors resulting from the structure of the noise. **a)** *iid* noise (homoscedastic, and uncorrelated; white in the frequency domain), **b)** heteroscedastic noise, which has greater magnitude on channel *j* than channel *i* (uncorrelated, also white in the frequency domain), and **c)** noise which is heavily correlated, and heteroscedastic (1/*f* characteristics in the frequency domain).

The difference in the behavior of *iid* noise, and non-*iid* noise is considerable. As we shall see, these marked differences can have an enormous impact on the validity of multivariate calibration models.

1.3.2 Measurement Error Structure and Multivariate Calibration

It was noted in the end of **Section 1.2.4.2** that several assumptions which are implicit in multivariate calibration were intentionally avoided. Here we will

examine the influence of measurement error structure in PCA and multivariate calibration (PCR), from the standard assumptions of *iid* noise to the implications of non-*iid* noise.

Principal components analysis can be applied to data with any noise structure, however the technique develops parameter estimates under the assumption that the measurement errors corrupting the data are *iid* (and normal). If the structure of the measurement errors is indeed *iid*, then PCA is ensured to provide *maximum likelihood* parameter estimates, which are, in essence, the *most likely* guesses at the true parameters, given the data at hand. The simple scenario of two spectral vectors, approximated by a rank one space (assumed known) is depicted in **Figures 1.14**. The representation of the spectrum $\mathbf{x}_1$ in the rank one space is the orthogonal projection of this vector onto the space $S^o_{\tilde{\mathbf{x}}}$ (as

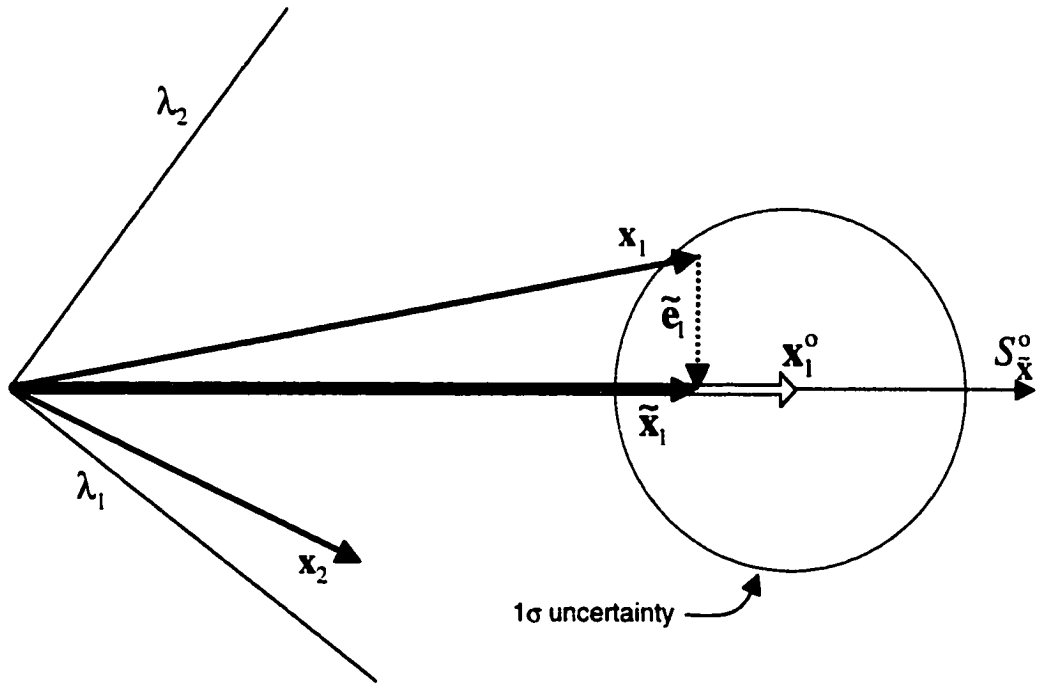


Figure 1.14 An illustration of the likelihood implications of projecting $\mathbf{x}_1$ onto the model space (assumed known in this case) when the measurement errors corrupting the true $\mathbf{x}_1$ vector are *iid*.

discussed in **Section 1.2.4.1**). This orthogonal (least-squares) projection naturally minimizes the length of the error vector between any $\mathbf{x}_i$ and $\tilde{\mathbf{x}}_i$ ($\mathbf{x}_i$'s projection onto $S^{\circ}_{\tilde{\mathbf{x}}}$), $\tilde{\mathbf{e}}_i$. This projection also yields the most likely estimate of $\mathbf{x}^{\circ}_i$, or, the *maximum likelihood* solution, under the assumption that the errors in the system are *iid*, and the true vector value is of course unknown. It is, in short, the best guess at $\mathbf{x}^{\circ}_i$ when the measurement errors are *iid*.

If the measurement errors are not equally distributed at all channels in the spectrum, then the uncertainty sphere shown in **Figure 1.14** will be stretched along the coordinate axes in some fashion, as is shown in **Figure 1.15**. The error variance on channel λ_1 is much greater than the variance at channel λ_2, causing the probability density functions to be longer along the λ_1 axis than the λ_2

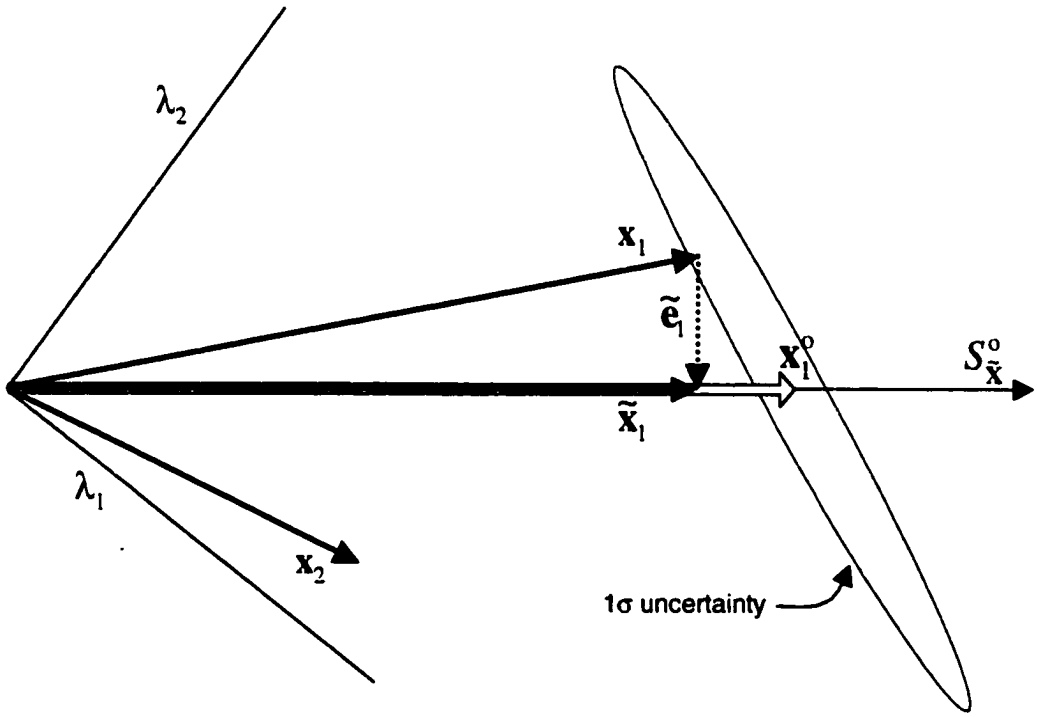


Figure 1.15 An illustration contrasting the likelihood of the projected $\mathbf{x}_1$ when the measurement errors deviate significantly from *iid*. In this case the likelihood of the projection given the true value is *extremely* low because a simple orthogonal projection has been used, which is far from ideal when the errors are non-*iid*.

axis. It is additionally apparent from **Figure 1.15** that the system is corrupted by error covariance, since the principal axes of the error ellipsoid are not aligned with the coordinate axes λ_1 and λ_2. If the standard PCA orthogonal projection onto $S_{\tilde{\mathbf{x}}}^{\circ}$ is used, the estimate of $\mathbf{x}_i^{\circ}$ remains the "least-squares" estimate (a guess), however it is no longer the *maximum likelihood* estimate of the true signal vector (in fact, in the case shown in **Figure 1.15**, it is an extremely *unlikely* estimate – the probability of the estimate $\tilde{\mathbf{x}}_i$ giving rise to $\mathbf{x}_1$ under these error conditions is < 0.1%).

Measurement errors can be correlated in several different ways in a data matrix $\mathbf{X}$. There may be correlations among channels in the calibration spectra (columns in the matrix), for instance. This is often the prevailing scenario in spectroscopic calibration, and can arise from a variety of conditions including sensor spatial correlations in the instrument, source flicker, and numerous signal processing techniques. Provided this row covariance structure is approximately equivalent from sample spectrum to sample spectrum, the PCA rank *p* subspace estimate is still typically quite good. This can be rationalized from the fact that since the error structure in all the calibration spectra is roughly the same, then the numerical averaging achieved by using many samples results in a reasonable estimation of the calibration space. However, marked errors will inevitably occur in the projections of the calibration (and prediction) spectra onto that rank *p* subspace, since, as illustrated above, the orthogonal projections have the potential to yield highly unlikely estimates in many cases. In principal components regression, then, the scores for the sample spectra will be highly inaccurate, a problem that will severely hamper any predictive utility for the model.

In addition to this row covariance, there may be correlations from sample to sample. This can be the case for multivariate calibration if, for example, samples are not run in random order, in which case low-frequency variations in the instrument (*e.g.*, temperature drift, source degradation) can become imbedded in the measurement error structure. This error covariance condition

tends to lead to the opposite condition observed for row-only covariance in that it is the accuracy of the subspace estimate that will be hampered, while the scores will be reasonably estimated. Nonetheless, this error structure also severely limits the utility of the calibration model, since the model space tends to be very poorly estimated.

The *iid* error assumption, pervasive in chemical modeling methods such as PCR, is largely a remnant of the age of univariate calibration. Because single channel measurements predominated in chemical calibration, the correlation of sample to sample errors was assumed to be minimal, and of course the correlation between the errors at different sensors was an impossibility. The mass movement to multichannel instrumentation, and now to multivariate calibration, however, has occurred without corresponding advances in the handling of measurement errors. The most widely used multivariate calibration methods of today (*e.g.*, PCR, partial least-squares (PLS)) typically assume error structures which are, in practice, rarely even approximately valid. Some scaling methods have been proposed which can assist in handling the simplest deviation from *iid* – heteroscedasticity. Unfortunately these scaling procedures can only properly adjust the error structure of the data in the rare case in which the rank of the matrix of measurement error standard deviations is unity [9]. While it is standard practice to go to great lengths in order to force the data to conform to the *iid* assumptions, a more direct approach is to make the model accommodate the data, as models are – by definition – designed to do.

1.4 Maximum Likelihood Principal Components Regression

Principal components analysis can be applied to data with any measurement error structure, however PCA generates estimates of the parameters (scores, and loadings) with the assumption that the errors corrupting the data are *iid*. If the errors are indeed *iid*, and the selected dimension of the subspace, p, is correct, then PCA is ensured to yield *maximum likelihood*

estimates for the data, which are, in essence, estimates of the true parameter values which are the *most likely* given the data at hand. As noted above, however, deviations from *iid* conditions seriously hamper the accuracy of standard PCA. With row-correlated measurement errors in particular, the estimations of the sample scores are most seriously hindered. Since the sample scores are crucial if PCA is to be used in PCR, it is clear that PCR with non-*iid* errors is bound to perform less than desirably. From the geometric discussion of measurement errors and regression, it is apparent that what is needed is a modeling method which accounts for measurement error structure in the estimation of the spectral subspace, and in the estimation of the sample scores. Such a method has been recently introduced as maximum likelihood PCA (MLPCA) [10], and, like PCA, it can be used in the context of multivariate calibration (maximum likelihood PCR, or MLPCR) [11].

1.4.1 Maximum Likelihood PCA

MLPCA can be considered to be a general extension of PCA to the situations in which the simple assumptions of *iid* errors are sufficiently violated. The MLPCA decomposition of a data matrix $\mathbf{X}$ is achieved so that the optimal p-dimensional subspace is found regardless of the structure of the measurement errors. The method can readily accommodate heteroscedasticity, and has been shown to be of great benefit in this area [11]. It can also handle the more difficult problem of error covariance [10, 12]. The most complex scenario for error structure in the data matrix, $\mathbf{X}$, is when error covariance exists among columns, and rows, and the error covariance structure is different for every sample in the data set, and every wavelength in the spectral domain. This full treatment of error structure, while theoretically feasible, is computationally burdensome. In first-order spectroscopic calibration (calibration on vector sample measurements) several simplifying assumptions can be made which ease this computational expense. It is often reasonable to assume that there is no error covariance between the samples; that is, the values of the measurement errors for one sample are unrelated to the values of the measurement errors for another

sample. An additional simplification which may or may not be viable is the assumption that the error covariance structure is roughly equivalent for all samples in the calibration set. This assumption may be tenuous if a significant portion of the error covariance structure arises from sample specific phenomena. Instrumental contributions to error covariance structure, however, generally result in similar noise characteristics independent of the sample under observation. Other effects, such as multiplicative scattering in diffuse reflectance, introduce error covariance structure which is highly dependent on the shape of the spectral profile. Provided the sample spectra are reasonably similar (as in many near-infrared applications), equal row-covariance assumptions can often safely be made. In all cases under examination in this work, the equal row error covariance structure was found to be approximately valid, and hence, it is this form of MLPCA, and MLPCR that will be discussed here.

Previously in **Section 1.2.4.1**, the PCA rank *p* subspace was estimated using a singular value decomposition of the data matrix (**Equation 1.25**). While an alternating least-squares optimization can be used to estimate the MLPCA model parameters in the general case [10, 12], an excellent approximation can be rapidly achieved if the simplifying assumptions discussed in the previous paragraph appear to be valid. In these scenarios, the loading vectors $\mathbf{V}^T$ are very well estimated by a simple SVD of the data matrix, and so this shortcut can be used without peril.

$$\mathbf{X} \xrightarrow{SVD} \mathbf{USV}^T \tag{1.44}$$

As we saw in **Section 1.3.2**, however, with error covariance contributing significantly to the error structure the estimated scores of the spectral vectors can be drastically wrong, even given the correct model subspace. Unlike PCA, in MLPCA the scores are not simply $\mathbf{US}$ from the above calculation, but must be determined from a more elaborate procedure.

The projection of a sample spectrum, $\mathbf{x}$, onto the rank *p* MLPCA subspace (defined by $\tilde{\mathbf{V}}^T$) is calculated using the following general operator

$$\breve{\mathbf{x}} = \mathbf{x}\boldsymbol{\Sigma}^{-1}\breve{\mathbf{V}}\left(\breve{\mathbf{V}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\breve{\mathbf{V}}\right)^{-1}\breve{\mathbf{V}}^{\mathrm{T}} \tag{1.45}$$

The overstrike "$\cup$" is used to distinguish the MLPCA parameters from the usual PCA parameters. (While $\breve{\mathbf{V}}$ (MLPCA) and $\tilde{\mathbf{V}}$ (PCA) are essentially identical under these assumptions, the scores and data estimates are not; to maintain uniformity in presentation, the "$\cup$" will be used throughout to indicate MLPCA-associated values.) Alternatively, all of the sample spectra can be projected onto the subspace by the extension of **Equation 1.45**,

$$\breve{\mathbf{X}} = \mathbf{X}\boldsymbol{\Sigma}^{-1}\breve{\mathbf{V}}\left(\breve{\mathbf{V}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\breve{\mathbf{V}}\right)^{-1}\breve{\mathbf{V}}^{\mathrm{T}} \tag{1.46}$$

This projection is general in the sense that it is not fixed to be orthogonal to the subspace defined by the loading vectors. Rather, the projection will occur obliquely in some direction defined by the principal components and the error covariance matrix. This oblique operation, in essence, guides the projections of the spectral vectors in the directions of greatest variation in the error structure – the direction of greatest uncertainty. An illustration of the difference between the standard PCA projections and the MLPCA projections is shown in **Figure 1.16**. The reader will also notice that when the error covariance matrix is a multiple of the identity matrix (the errors are *iid*) **Equations 1.45** and **1.46** reduce to the standard orthogonal projections of PCA,

$$\breve{\mathbf{X}} = \mathbf{X}\breve{\mathbf{V}}\breve{\mathbf{V}}^{\mathrm{T}} \tag{1.47}$$

as is to be expected. The MLPCA decomposition can be completed now by a second SVD of the (now rank p) projected data matrix $\breve{\mathbf{X}}$.

$$\breve{\mathbf{X}} \xrightarrow{SVD} \breve{\mathbf{U}}\breve{\mathbf{S}}\breve{\mathbf{V}}^{\mathrm{T}} \tag{1.48}$$

Only p of the scores and loading vectors are meaningful since $\breve{\mathbf{X}}$ was only rank p to begin with, as is intended to be conveyed by the notation. The second SVD is simply a tidy way of determining the scores with the MLPCA projection already achieved.

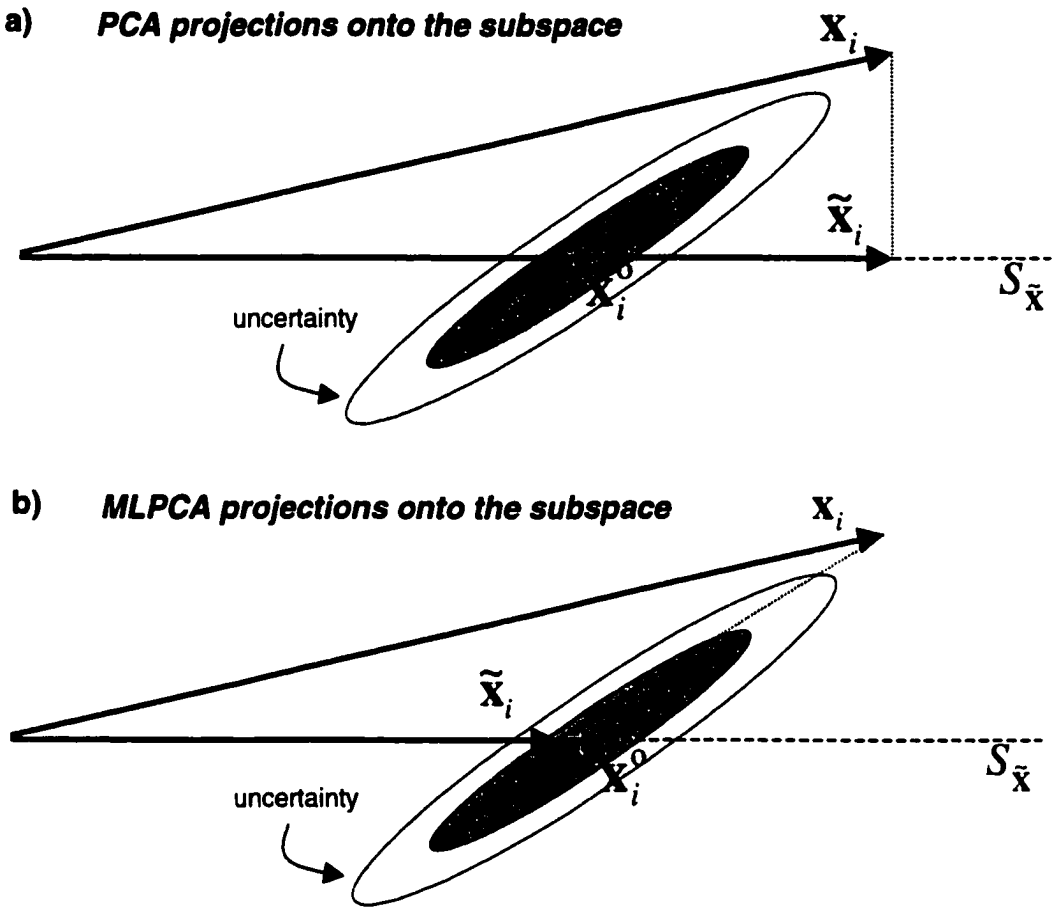


Figure 1.16 **a)** Standard PCA orthogonal projection of an **x**-vector onto the subspace estimated by PCA, while the measurement error structure is clearly deviating from *iid* conditions. **b)** MLPCA projection under the same circumstances. The measurement error structure provides a directional guide for the projection of the **x**-vector onto the subspace. The resulting estimates in **a)** and **b)** are in the same space, but are very different in length.

1.4.2 Maximum Likelihood PCR

When used in the context of multivariate regression and calibration MLPCA provides a powerful calibration method to deal with non-*iid* noise conditions. While similar to PCA when used in multivariate regression, there are some subtle differences in the use of the scores and loadings of the calibration, and prediction data.

Equation 1.32 of **Section 1.2.4.2** gave the estimated regression vector for a single component in PCR as

$$\hat{\mathbf{b}} = \tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^{\mathrm{T}}\mathbf{y} \tag{1.49}$$

and the prediction equation for new samples as

$$\begin{aligned}\hat{\mathbf{y}}_{unk} &= \mathbf{X}_{unk}\hat{\mathbf{b}} \\ &= \mathbf{X}_{unk}\tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^{\mathrm{T}}\mathbf{y}\end{aligned} \tag{1.50}$$

Although not immediately apparent, the prediction step involves a projection of the spectra of the unknown samples onto the subspace estimated by PCA, which is more evident when **Equation 1.50** is written as

$$\hat{\mathbf{y}}_{unk} = \mathbf{X}_{unk}\left(\tilde{\mathbf{V}}\tilde{\mathbf{V}}^{\mathrm{T}}\right)\tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^{\mathrm{T}}\mathbf{y} \tag{1.51}$$

Because an oblique projection onto the subspace is required in MLPCR, this step is slightly more complicated, involving

$$\hat{\mathbf{y}}_{unk} = \left[\mathbf{X}_{unk}\boldsymbol{\Sigma}_{unk}^{-1}\check{\mathbf{V}}\left(\check{\mathbf{V}}^{\mathrm{T}}\boldsymbol{\Sigma}_{unk}^{-1}\check{\mathbf{V}}\right)^{-1}\check{\mathbf{V}}^{\mathrm{T}}\right]\check{\mathbf{V}}\check{\mathbf{S}}^{-1}\check{\mathbf{U}}^{\mathrm{T}}\mathbf{y} \tag{1.52}$$

where the reader will recognize the portion of the equation in the square brackets as the oblique projection operator from **Equation 1.46** above. With the inner product $\check{\mathbf{V}}^{\mathrm{T}}\check{\mathbf{V}}$ canceling, the prediction step for MLPCR becomes

$$\hat{\mathbf{y}}_{unk} = \mathbf{X}_{unk}\boldsymbol{\Sigma}_{unk}^{-1}\check{\mathbf{V}}\left(\check{\mathbf{V}}^{\mathrm{T}}\boldsymbol{\Sigma}_{unk}^{-1}\check{\mathbf{V}}\right)^{-1}\check{\mathbf{S}}^{-1}\check{\mathbf{U}}^{\mathrm{T}}\mathbf{y} \tag{1.53}$$

With the error covariance structure for the unknown samples being included in the prediction step, it is impossible to develop a general regression vector which would handle any future samples, unless of course the error covariance matrix never changes. Several other technical points are worthy of note. The error covariance matrix is a directional guide for the projections of the spectra onto the model subspace, and as a result, **Equation 1.53** is invariant to changes in the scale of $\boldsymbol{\Sigma}$. In addition, the obliqueness of the spectral projections indicates that MLPCA and hence, MLPCR do not generate *nested* models. In

MLPCA

$$\mathbf{X} \xrightarrow{SVD} \mathbf{USV}^{\mathrm{T}}$$

$$\breve{\mathbf{X}} = \mathbf{X}\boldsymbol{\Sigma}^{-1}\breve{\mathbf{V}}\left(\breve{\mathbf{V}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\breve{\mathbf{V}}\right)^{-1}\breve{\mathbf{V}}^{\mathrm{T}}$$

$$\breve{\mathbf{X}} = \breve{\mathbf{U}}\breve{\mathbf{S}}\breve{\mathbf{V}}^{\mathrm{T}}$$

MLPCR and Calibration

$$\mathbf{X} \xrightarrow{SVD} \mathbf{USV}^{\mathrm{T}}$$

$$\breve{\mathbf{X}} = \mathbf{X}\boldsymbol{\Sigma}^{-1}\breve{\mathbf{V}}\left(\breve{\mathbf{V}}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\breve{\mathbf{V}}\right)^{-1}\breve{\mathbf{V}}^{\mathrm{T}}$$

$$\breve{\mathbf{X}} = \breve{\mathbf{U}}\breve{\mathbf{S}}\breve{\mathbf{V}}^{\mathrm{T}}$$

$$\hat{\mathbf{y}}_{unk} = \mathbf{X}_{unk}\boldsymbol{\Sigma}_{unk}^{-1}\breve{\mathbf{V}}\left(\breve{\mathbf{V}}^{\mathrm{T}}\boldsymbol{\Sigma}_{unk}^{-1}\breve{\mathbf{V}}\right)^{-1}\breve{\mathbf{S}}^{-1}\breve{\mathbf{U}}^{\mathrm{T}}\mathbf{y}_{cal}$$

‡ $$\hat{\mathbf{y}}_{unk} = \mathbf{X}_{unk}\hat{\mathbf{b}}_{unk}$$

Figure 1.17 An algorithmic summary of MLPCA, and its regression counterpart, MLPCR, with equal row covariance assumptions.

(‡ The estimated regression vector is not general in the manner that a PCR regression vector is general, since it depends on the error covariance structure of the spectra for the unknown samples.)

PCR, the rank p model contains the rank p-1 model. This is not necessarily the case in MLPCR, and therefore each model must be calculated independently. An algorithmic summary of MLPCA and MLPCR is given in **Figure 1.17**.

1.5 Figures of Merit in Multivariate Calibration

The last several sections have outlined some of the machinery available to the analyst for performing multivariate calibration and subsequent predictions, including CLS, PCR and MLPCR. Aside from these operational details, a general theory of analytical chemistry exists which is extremely useful in

comparing calibration procedures independently of the methods used to build the models. In this section, the general theory is outlined, and some theoretical metrics for multivariate calibration are introduced. Since the modeling machinery is in place, it will be shown how these theoretical values can be estimated in practice.

1.5.1 Mixture Theory and the Net Analyte Signal

General calibration theory is reliant on the so-called linear mixture model, expressed previously (**Equation 1.12**) as

$$\begin{aligned}\mathbf{R} &= \mathbf{c}_1\mathbf{s}_1^{\mathrm{T}} + \mathbf{c}_2\mathbf{s}_2^{\mathrm{T}} + \cdots + \mathbf{c}_p\mathbf{s}_p^{\mathrm{T}} \\ &= \mathbf{C}\mathbf{S}^{\mathrm{T}}\end{aligned} \tag{1.54}$$

where, to review, $\mathbf{R}$ is an m x n matrix of instrument responses, $\mathbf{C}$ is an m x p matrix of the p-component concentrations, and $\mathbf{S}$ is the p x n matrix of pure-component spectra. When the spectrum of a mixture is obtained, it can be imagined to arise from the addition of p different pure-component contributions, as is shown in **Figure 1.18** for a simple two component system. Since any mixture spectrum in a p-component system can only arise from only p independent spectral contributions, all of the pure-component spectra, and the mixture spectra they can possibly generate lie in a p-dimensional subspace. Due to the ubiquity of measurement errors, these observed mixture spectra typically only *approximately* lie in a p-dimensional space, and therefore, this space must be estimated by some numerical method in multivariate calibration, such as PCA or MLPCA, for instance.

While this space is of fundamental importance in calibration, it is essential that a further dissection of the mixture space be achieved if we hope to quantify particular analytes. We must ascertain which directions *in this subspace* are related to the analyte of interest, and independent of the other interfering chemical species. The only direction in the mixture space which is independent of the interfering species will be the direction which is orthogonal to all of the interfering pure-component spectra. This direction is referred to as the

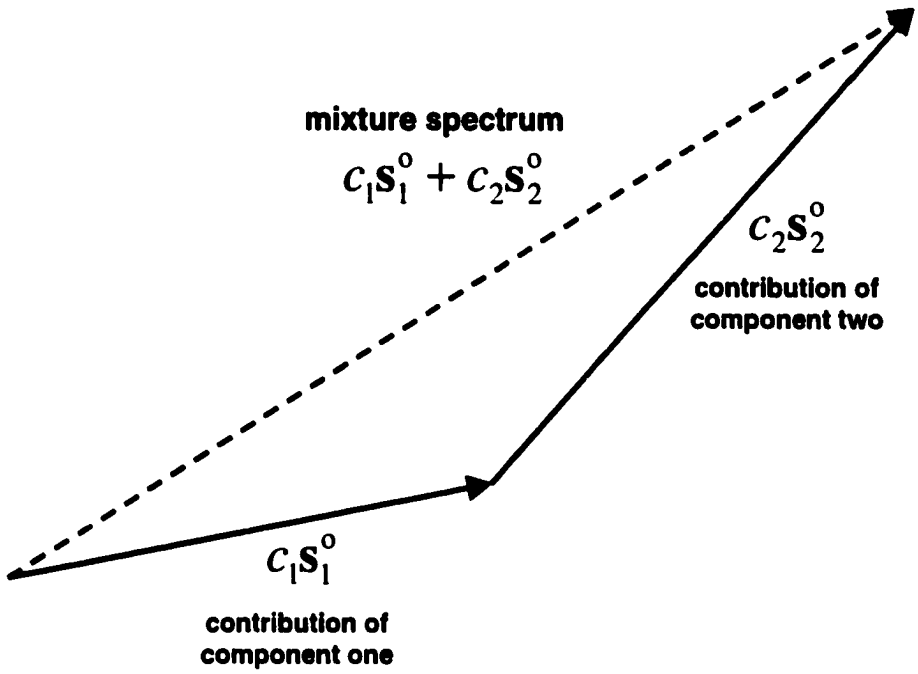


Figure 1.18 The embodiment of the linear mixture model, which models any observed mixture spectrum as the linear combination of pure-component contributions.

contravariant ('against the variation') domain of the interferences, and, for the *i*th analyte of interest, is indicated by the contravariant vector, $\mathbf{v}_i$ [13]. For our simple two component example, the two contravariant vectors are shown in **Figure 1.19**. Therefore, in an analysis for analyte 1, we can selectively remove the contributions of any interfering chemical species from a mixture spectrum by projecting this mixture spectrum onto the contravariant vector associated with component 1 (and orthogonal to all other interferences). We have therefore established a single direction which is *exclusively* related to the analyte of interest. It is perhaps most evident from the geometric expression of the contravariant vector that doubling the contribution of the analyte of interest to a mixture spectrum will double the length of the projection on the contravariant vector, while doubling the concentration of the interferent will leave the projection unchanged. To do accurate quantitation, it is necessary to have some reference

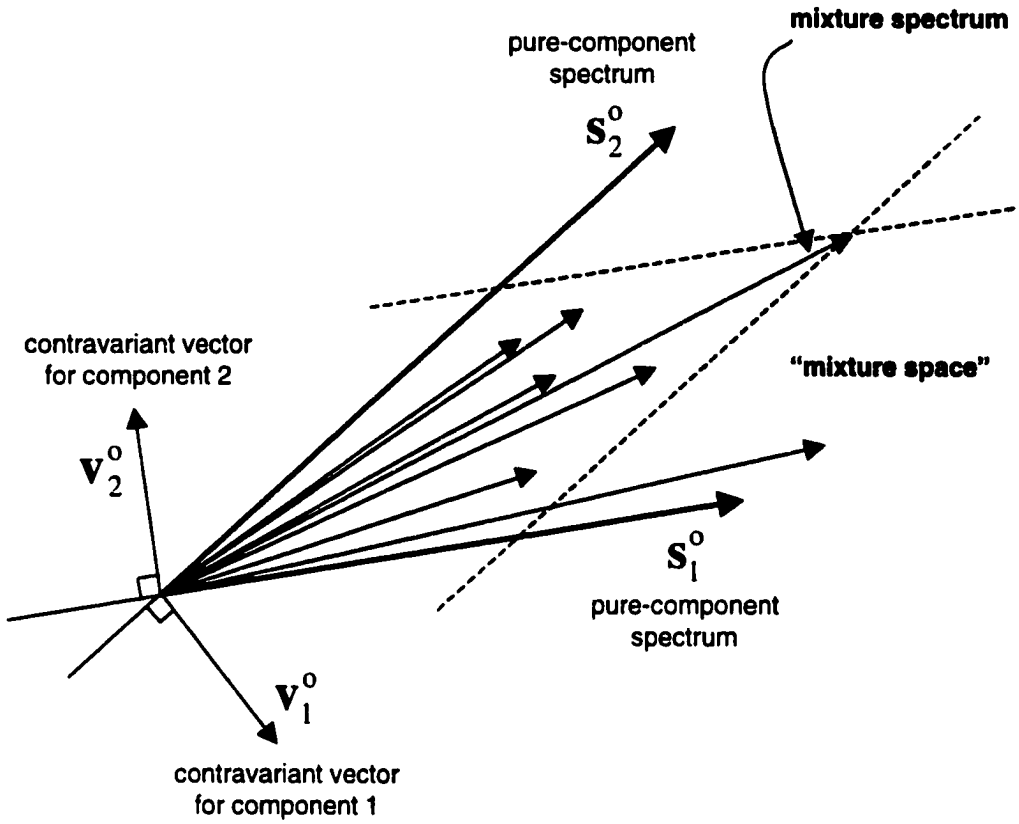


Figure 1.19 A series of mixture spectra when viewed from a mixture theory perspective. The mixture vectors lie in the "mixture space" which, in the absence of noise will be coplanar with the space defined by the pure-component spectra fo the active components. The contravariant vectors (shown at left) indicate the directions in the mixture space which are exclusively associated with their associated components.

value, a projection on the contravariant vector corresponding to a known analyte concentration. If the pure-component spectra were known and measured at unit concentration, this reference projection, called the net analyte signal (NAS) [14], would be given for the *i*th analyte as

$$\mathrm{NAS}_i = \left(\mathbf{I}_n - \mathbf{S}_{-i}\mathbf{S}_{-i}^{+}\right)\mathbf{s}_i \tag{1.55}$$

where columns of the matrix $\mathbf{S}_{-i}$ contain the pure-component spectra of the interferences in the mixture (and *excluding* the analyte of interest), and $\mathbf{s}_i$ is the pure-component spectrum of the analyte [15]. This concept is illustrated in

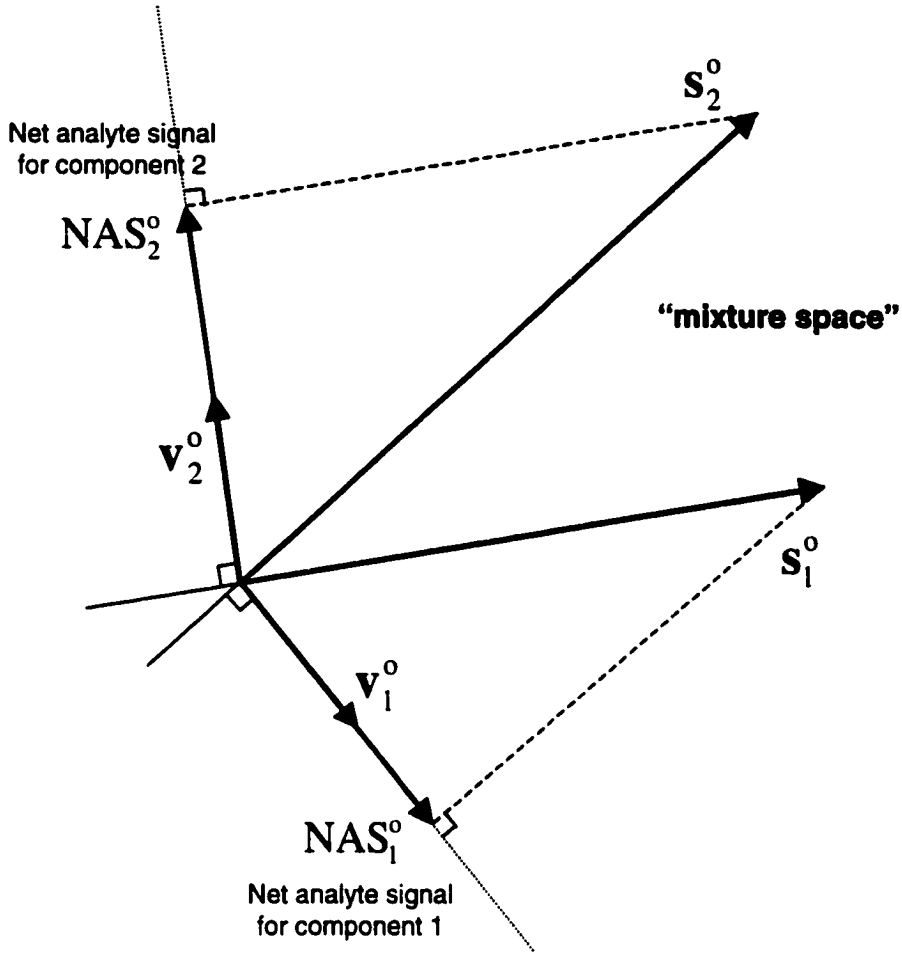


Figure 1.20 The net analyte signal vectors for each component, colinear with their associate contravariant vectors, are the orthogonal projections of the pure-component spectra at unit concentration onto the contravariant vectors.

Figure 1.20. Even though the NAS is a vector-valued quantity, we will refer to it in subsequent equations by its acronym to avoid introducing additional complicating symbols. With the NAS established, quantitative prediction can easily proceed by projecting the mixture spectrum onto the contravariant domain, and comparing the magnitude of this projection with the NAS.

It was noted that a particular issue with classical least-squares methods for multivariate calibration was that all of the pure-component spectra may not be available, or, more likely, several of the components in the mixture are unknown,

making the calculation of the NAS from **Equation 1.55** impossible. Fortunately, the NAS is intimately related to the regression vector obtainable via inverse calibration methods such as PCR and MLPCR. The NAS_i can be approximated from the estimated regression vector, $\hat{\mathbf{b}}_i$, using

$$\hat{\mathbf{b}}_i \cong \frac{NAS_i}{\|NAS_i\|^2} \tag{1.56}$$

and

$$\|NAS_i\| \cong \frac{1}{\|\hat{\mathbf{b}}_i\|} \tag{1.57}$$

1.5.2 Multivariate Figures of Merit

In standard univariate calibration, figures of merit such as the sensitivity (*SEN*), selectivity (*SEL*) and signal-to-noise ratio (*S/N*) are often used to describe the attributes of a univariate calibration model. Multivariate analogues of these univariate metrics exist and are perhaps even more insightful in the understanding of the strengths and limitations of particular multivariate calibration scenarios. These multivariate figures of merit are all related in some fashion to the previously discussed net analyte signal vector. Although there is some ongoing discussion as to the proper expressions of these figures of merit, most of the definitions to be used throughout the course of this work are those that are generally accepted, and summarized in a 1994 article by Booksh and Kowalski [15].

The multivariate sensitivity is an extension of the univariate metric, and is the observed change in multivariate signal for a unit change in concentration. Since the NAS is defined to be the multivariate response per unit concentration, the *SEN* for the *i*th component of interest is simply given by

$$SEN_i = \|NAS_i\| \tag{1.58}$$

making it apparent that the *SEN* is a scalar quantity, easily determined from the length of the NAS vector. The multivariate selectivity is a measure of the fraction of the pure-component spectrum which resides in the contravariant domain, and is given by

$$SEL_i = \frac{\|\mathrm{NAS}_i\|}{\|\mathbf{s}_i\|} = \cos(\angle \mathrm{NAS}_i, \mathbf{s}_i) \tag{1.59}$$

As is clear from the illustration in **Figure 1.18**, the *SEL* for component *i* is also given by the cosine of the angle between the pure-component spectrum $\mathbf{s}_i$, and the NAS or contravariant vector. Both the multivariate *SEN* and *SEL* are straightforwardly derived from similar univariate metrics. The extension of a univariate *S/N* to the multivariate realm is much less transparent.

The typical measure of *S/N* in univariate systems is simply the mean response observed over several replicates ($\bar{r}$) divided by the standard deviation of the measurements (σ_{noise}), or

$$S/N_i = \frac{\bar{r}}{\sigma_{noise}} \tag{1.60}$$

Since a host of responses are observed in multivariate measurements, the question arises as to which signal is to be used in **Equation 1.60**. Frequently the maximum signal is used for a signal estimate. While this invariably yields a metric for the *S/N*, it is entirely useless in conveying information about the plausibility of doing accurate quantitation, since any observed signal in the measurement vector may, or may not be attributable to the analyte of interest. This form of the *S/N* is only of practical use in the unusual case in which there are no chemical interferences. Instead, a multivariate signal metric must be used which accounts for the contributions of interferences. One such general measure is given by

$$S/N_i = \frac{SEN_i}{\sigma_{noise}} = \frac{\|NAS_i\|}{\sigma_{noise}} \tag{1.61}$$

In **Section 1.3.1.2**, however, concern was expressed with conventional descriptors of noise which ignore the structure of the measurement errors. In fact, **Equation 1.60** is only useful in the event that the measurement errors are *iid*, in which case the influence of the measurement errors is probabilistically equivalent in all directions in the multivariate calibration space. When the calibration data exhibit non-*iid* noise characteristics, however, the structure of the measurement errors is crucial in determining a meaningful signal-to-noise ratio since, if extensive error variance and covariance happens to exist in the direction of the NAS, then accurate quantitation is severely hampered. Therefore, an extension of **Equation 1.60** is needed for the situations in which measurement errors are non-*iid* (shown below).

$$S/N_i = \frac{SEN_i}{\sqrt{\mathbf{v}_i^T \boldsymbol{\Sigma} \mathbf{v}_i}} \tag{1.62}$$

In **Equation 1.62**, first published by Brown and Wentzell [16], the conventional 'multivariate signal' expression is used in the numerator, however the denominator consists of a more advanced noise estimate than previous equations. $\mathbf{v}_i$ is of course the contravariant vector for the *i*th component, and $\boldsymbol{\Sigma}$ is the now familiar error covariance matrix. Mathematically, the denominator corresponds to a projection of the error covariance matrix onto the contravariant vector, quantifying the amount of error variance that exists in the all-important unidirectional NAS. When the measurement errors are *iid*, $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$, **Equation 1.62** reduces to the anticipated *iid* expression

$$S/N_i = \frac{SEN_i}{\sigma_{noise}\sqrt{\mathbf{v}_i^T \mathbf{v}_i}} = \frac{SEN_i}{\sigma_{noise}} \tag{1.63}$$

because $\mathbf{v}_i$ is of unit length. Three different multivariate *S/N* scenarios are explored in **Figure 1.21**, in which signal and error covariance considerations have significant impacts on the multivariate *S/N*. While additional figures of merit, such as limit of detection (*LOD*), are available in multivariate calibration,

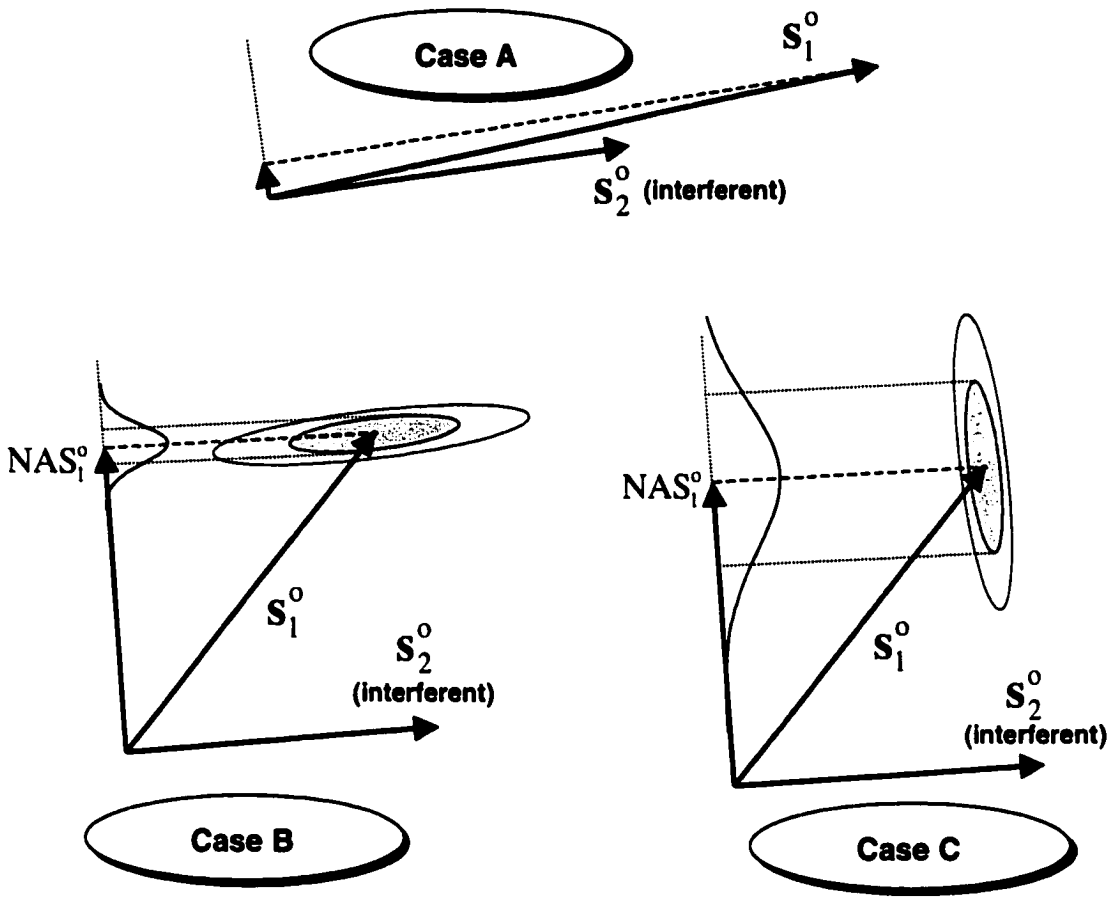


Figure 1.21 An examination of the different issues of importance in multivariate signal-to-noise ratios. **Case A**: low *S/N* - Why? Negligible *SEN*. **Case B**: High *S/N* - Why? Large *SEN*, and low projection of the error covariance onto the contravariant vector. **Case C**: Low *S/N* - Why? Large *SEN*, but error covariance is greatest in the direction of the NAS, and therefore has a very large projection onto the contravariant vector.

they are not of particular interest in this work, and hence, will not be discussed further. The interested reader is referred to the works of Booksh and Kowalski [15], Lorber *et al.* [17], and Faber *et al.* [18] for more involved discussions of multivariate figures of merit in an NAS framework.

2. Digital Filtering and Preprocessing

2.1 Introduction

Preprocessing is usually defined in chemometrics textbooks as 'any transformation of the original data' performed prior to multivariate analysis or calibration. 'Original data' is a rather ambiguous term, but in this work it will be considered to mean the experimental data measured against a *specific* ordinal variable as they are presented to the analyst by the instrument. (Signals measured against heterogeneous ordinal variables are ill-suited for treatment with digital filters, and hence, this restriction is necessary.) For example, the output of a typical Ultraviolet-Visible (UV-Vis) spectrometer is a series of absorbance values measured against an ordinal variable, usually wavelength. Of course the data have already gone through a variety of unseen analog and digital signal processing operations. Built-in signal processing is certainly important; however, since these instrumental operations are typically opaque, and unalterable by the user, they will not be discussed further. Rather, the focus will be on alterations which can be performed by the analyst on these 'original data'. The extensive use of computers in the modern analytical laboratory means that the data will invariably be accessible to the analyst in digital form, and as a consequence, further manipulations of the digital measurements (preprocessing) are most often carried out in software.

The availability of computational platforms and easy-to-use software packages has unquestionably afforded the analyst a far greater degree of 'after-the-fact' flexibility in signal processing options than would have been accessible in the past. A recent encyclopedia article explores many of the numerous signal processing methods currently in popular usage in analytical chemistry [19]. This computational flexibility has encouraged the analyst to take a key role in data

manipulation, trying their own hand at preprocessing. Preprocessing most commonly entails trivial tasks such as scaling and centering the data, however more advanced treatments such as transform methods (*e.g.*, Fourier) and digital filters are now in common use. The impact of the basic preprocessing methods in multivariate analysis has been studied to a reasonable extent elsewhere (see for example reference 20, and references therein) and these simple methods are reasonably well understood. The somewhat more involved techniques, such as digital filtering, are also well understood, but a knowledge of the theoretical influence of these methods on multivariate calibration models is surprisingly absent. This fundamental gap has not, however, impeded the use of digital filters as preprocessing methods. Without question, a more rational use of these methodologies would result if proper theoretical considerations could be made. To investigate the impact of these filtering methods in the calibration model, it is necessary to review some basic concepts in digital filtering [19, 21], and the remainder of this chapter is dedicated to describing the basic implementation and operation of such filters. Since the two chapters that follow deal explicitly with digital smoothing (**Chapter 3**) and digital differentiation (**Chapter 4**), a more general approach will be adopted here, leaving the details of the specific filter types to the more pointed discussions in their respective chapters.

2.2 Digital Filtering

In general, a digital filter can be described as an operation that is carried out on a contiguous subset of a discretely sampled signal vector to produce an estimate of a value in a *filtered* signal vector. This general definition includes a wide variety of digital filtering methods, but of particular interest in this work are the non-recursive filters which generate an estimated signal vector using only the original measurement sequence and a set of filter coefficients, as is shown in **Figure 2.1**. A mathematical expression for the most commonly employed filters in analytical chemistry, the polynomial least-squares filter, is

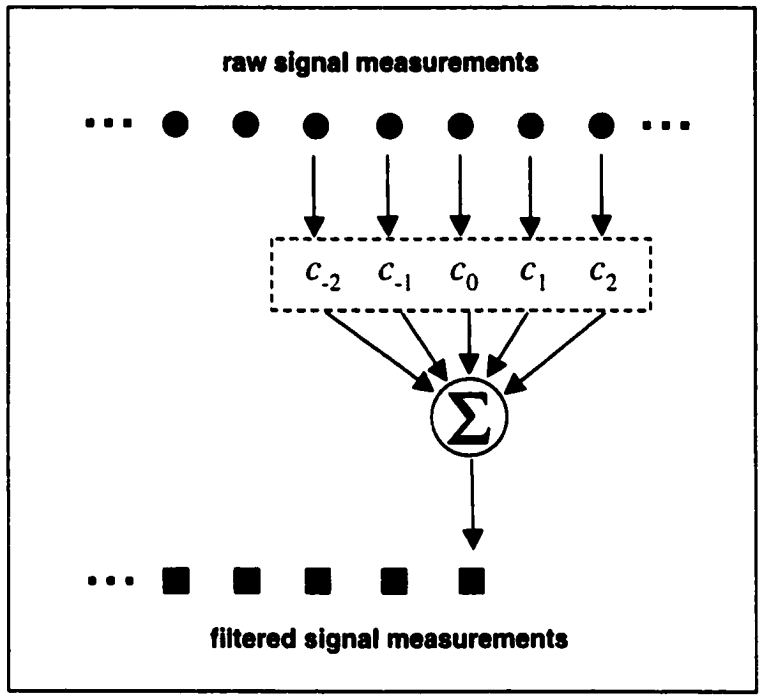


Figure 2.1 Illustration of the convolution of a set of filter coefficients with the raw signal vector to yield a filtered vector of measurements.

$$y_F(n) = \sum_{k=-m}^{m} c_k\, y(n+k) \tag{2.1}$$

where $y_F(n)$ is the filtered estimate of the point $y(n)$, and c_k is the *k*th filter coefficient. (This formula is more general than simply polynomial least-squares filters, and actually applies to all non-recursive filters, but polynomial least-squares filters will be the focus of this discussion.) Polynomial least-squares filters are commonly known to chemists as Savitzky-Golay (SG) filters, so named for their introduction into the chemical literature by Savitzky and Golay in 1964 [22]. The SG filter operates on the premise that a local region of the signal vector can be approximated by a polynomial function, a premise that is entirely reasonable for chemical signals which are most typically continuous and differentiable with respect to the ordinal variable. The central point in the local region (or, the filter 'window') is estimated from the fitted function that is chosen

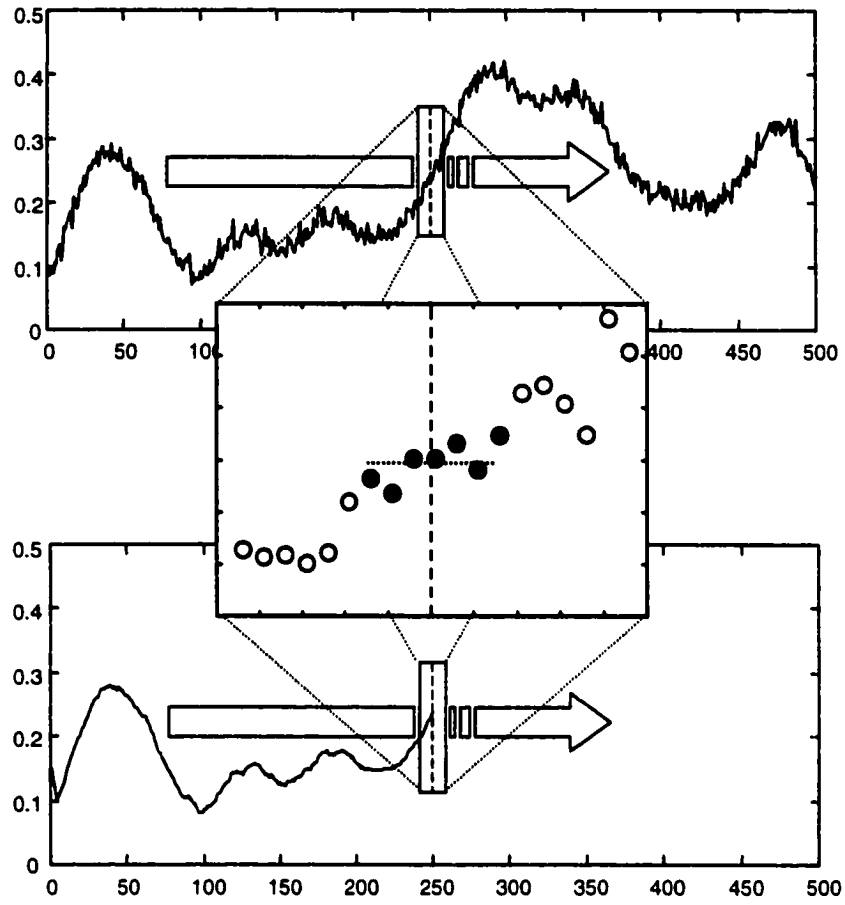


Figure 2.2 Illustration of the application of a seven-point moving-average filter. The points within the filter window are used to estimate a low-ordered polynomial approximation to the data. The estimated centre-point value on this fit is taken as the filtered estimate of the signal vector at the corresponding ordinal variable.

by the analyst. The filter window is then advanced in ordinal position, and a new filtered point is estimated from the original data in the filter window. The process repeats until the entire signal vector has been approximated by locally modeled polynomials. This is illustrated in **Figure 2.2** for a 7-point moving average (zero-order polynomial) filter.

2.2.1 Calculation and Expression of Digital Filter Coefficients

Although several articles and books have tabulated SG filter coefficients for use under a variety of filter conditions, the coefficients can now easily be

calculated with assistance of computers. To illustrate how this is achieved, consider the design of a second-order polynomial smoothing filter using five data points as the filter window. The linear model for a signal (y) measured against an ordinal variable (x) is given by:

$$y = b_0 + b_1 x + b_2 x^2 \tag{2.2}$$

and, in matrix form for the five points in the filter window, this gives

$$\begin{bmatrix} y_{-2} \\ y_{-1} \\ y_0 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & x_{-2} & x_{-2}^2 \\ 1 & x_{-1} & x_{-1}^2 \\ 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \tag{2.3}$$

or,

$$\mathbf{y} = \mathbf{X}\mathbf{b} \tag{2.4}$$

where $\mathbf{X}$ contains only information about the ordinal variable (*e.g.*, time, wavelength). Like any least-squares estimation, the estimated y values are scale invariant with respect to the specific x's, and, if the sampling interval on the ordinal variable is taken to be constant (typically the case), we can arbitrarily set the ordinals to be [-2 -1 0 1 2]. This results in an $\mathbf{X}$ matrix

$$\mathbf{X} = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \tag{2.5}$$

The least-squares estimate of the vector of regression coefficients, $\mathbf{b}$, is given in standard fashion as

$$\mathbf{b} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \tag{2.6}$$

$$\mathbf{b} = \mathbf{A}\mathbf{y} \tag{2.7}$$

The matrix **A** is a 3x5 matrix which can be regarded as being composed of three row vectors, $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{bmatrix} = \begin{bmatrix} \leftarrow \mathbf{a}_1 \rightarrow \\ \leftarrow \mathbf{a}_2 \rightarrow \\ \leftarrow \mathbf{a}_3 \rightarrow \end{bmatrix} \tag{2.8}$$

Note that the intercept coefficient for the fit, b_0, is obtained from the first row of this matrix,

$$b_0 = \mathbf{a}_1 \cdot \mathbf{y} = a_{11}y_1 + a_{12}y_2 + \ldots + a_{15}y_5 \tag{2.9}$$

Because $x = 0$ for the central point in the five point sequence, we have,

$$y_{0,F} = b_0 + b_1(0) + b_2(0)^2 = b_0 = \mathbf{a}_1 \cdot \mathbf{y} \tag{2.10}$$

where $y_{0,F}$ is the filtered estimate of the original point, y_0. Therefore, because of the generality of the ordinal representation, the estimate of the central point in the sequence is obtained simply by multiplying each measurement in the filter window by the corresponding element in $\mathbf{a}_1$. In other words, the five digital filter coefficients are simply the first row of the matrix $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, *i.e.*, $\mathbf{c} = \mathbf{a}_1$. This workup for SG filter coefficient calculation is completely general for filters of any length, and polynomial models of any order, with the appropriate adjustments made in the size of the matrix, **X**.

The above least-squares fitting solution for the SG filter coefficients can also be used to estimate the parameters for derivative filters. The extension for the first-derivative of a 5-point second-order polynomial model is

$$y'_{0,F} = \frac{d}{dx}\left(b_0 + b_1 x + b_2 x^2\right) = b_1 + 2b_2 x = b_1 + 2b_2(0) = b_1 \tag{2.11}$$

when evaluated at the center of the filter window. In a similar manner to **Equation 2.10**, the filtered (derivative) estimate for y'_0 is obtained by

multiplication of the second-row of the **A** matrix from **Equation 2.7**, hence, the five coefficients of this first-derivative filter are given simply as the elements of $\mathbf{a}_2$. Likewise, the filter coefficients for the second-derivative SG filter are given by the third row of **A**.

While the application of SG-type digital filters is commonly thought of as a moving-window approach, as was illustrated in **Figure 2.2**, the convolution of the filter with the signal vector can also be emulated in matrix form. The matrix expression of this operation is:

$$\mathbf{y}_F = \mathbf{F}\mathbf{y} \tag{2.12}$$

In **Equation 2.12** **y** is the raw signal vector (*e.g.*, a spectrum), $\mathbf{y}_F$ is the filtered signal, and **F** is the filter matrix, filled band diagonally with the filter coefficients. A 3-point SG filter matrix would have the form

$$\mathbf{F} = \begin{bmatrix} c_0 & c_1 & 0 & \cdots & \cdots & 0 & c_{-1} \\ c_{-1} & c_0 & c_1 & 0 & \cdots & 0 & 0 \\ 0 & c_{-1} & c_0 & c_1 & & \vdots & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & c_{-1} & c_0 & c_1 \\ c_1 & 0 & \cdots & 0 & 0 & c_{-1} & c_0 \end{bmatrix} \tag{2.13}$$

An additional merit of the matrix expression in **Equation 2.12** is the ease of extension to a matrix of signals:

$$\mathbf{Y}_F = \mathbf{F}\mathbf{Y} \tag{2.14}$$

where, in the convention of previous discussions, the columns of the matrices **Y** and $\mathbf{Y}_F$ contain the unfiltered and filtered signal vectors respectively.

While the notation used thus far is general for any signal vector **y** measured against an ordinal variable, **x**, it does not correspond well with the previous chapter's discussion of multivariate calibration in which **y** commonly represented the concentrations, and **X** the matrix of sample spectra. Therefore, the reader should note that since digital filters are most often applied to the

sample spectra (which occupy the rows of the matrix **X**), the filtering operation in a multivariate calibration sense is often represented as

$$\mathbf{X}_F = \mathbf{X}\mathbf{F}^T \tag{2.15}$$

One operational problem presents itself when these filters are applied to signal vectors of finite length, which are the rule in chemistry. Since the filter uses points in the signal vector both leading and trailing the central point, there will be points at the beginning and end of the vector that cannot be filtered. Several solutions to this problem have been proposed, including simply discarding these points. Alternatively, surrogate data points can be used to temporarily extend the signal vector, allowing the center of the filter to reach the end of the true signal vector. Points from the opposite end of the signal can be used provisionally in this manner, a practice that is commonly referred to as 'wrapping'. This is sound practice if the signal is not radically different on the ends of the vector (*e.g.*, if only baseline exists at the start and end of the vector), or if the signal shows a periodicity as is assumed in Fourier filtering. Other procedures have been proposed to circumvent edge-effects such as initial point or extended sliding window filters [23, 24] which are designed to operate asymmetrically, estimating values other than the central point in the filter window. These ordinal asymmetries can be accommodated, but the distortional effects and noise rejection characteristics of these sorts of filters are quite different than their symmetric counterparts [25]. In this work, central point estimation filters were exclusively used, with the requisite edge-handling procedures (*e.g.*, wrapping) adopted to suit the situation.

2.2.2 Frequency Response of Digital Filters

The responses of digital filters in the frequency domain (often referred to as the filter 'transfer function') can be derived directly from the filter coefficients. The amplitude component of the frequency response at a frequency, f, is given by

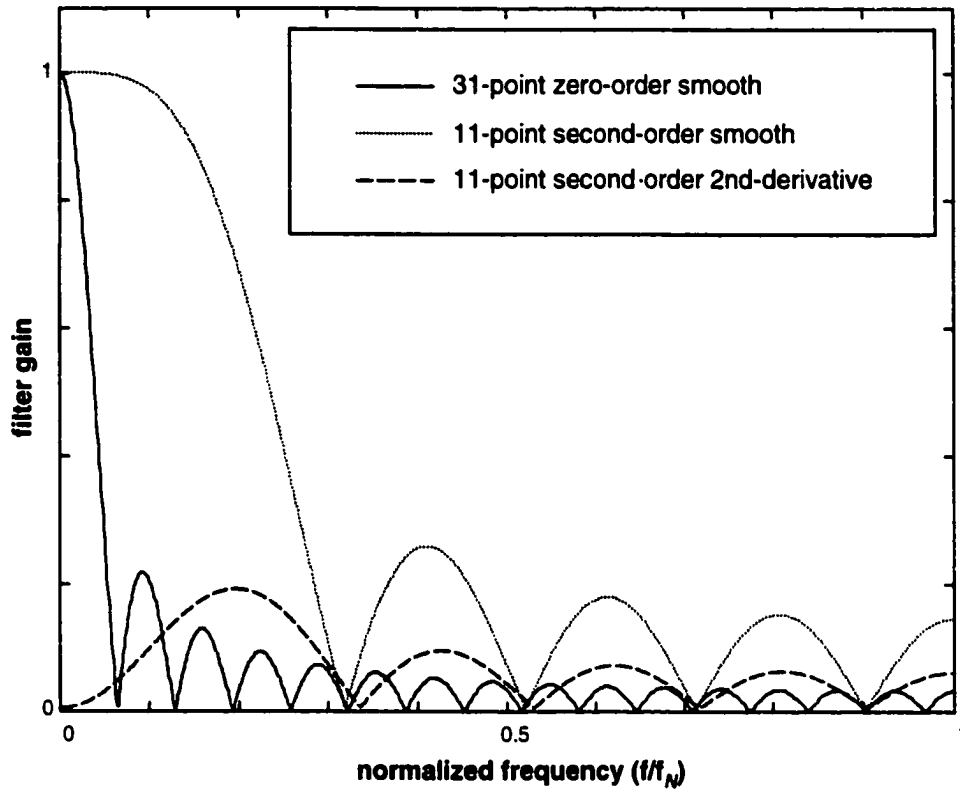


Figure 2.3 Transfer functions for a variety of Savitzky-Golay digital filters including a 31-point moving average, 11-point quadratic smooth and the 11-point quadratic second derivative filter.

$$H(f) = |a \cdot \cos\phi + b \cdot \sin\phi| \tag{2.16}$$

where

$$a = \sum_{k=-m}^{+m} c_k \cos(k\pi f / f_N) \tag{2.17}$$

$$b = \sum_{k=-m}^{+m} c_k \sin(k\pi f / f_N) \tag{2.18}$$

and the phase component, ϕ, of the filter's frequency response is given by

$$\phi = \tan^{-1}\left(\frac{b}{a}\right) \tag{2.19}$$

In **Equations 2.17** and **2.18**, f_N is the Nyquist frequency, which is by definition one half of the sampling frequency. The filter transfer function is an extremely useful descriptor for the influence of digital filtering on chemical signals, as the interaction of the noise with the filter is transparent, and potential distortional effects of the filter can also be anticipated in a qualitative manner. Transfer functions for a variety of polynomial least-squares filters are given in **Figure 2.3**, the characteristics of which will be discussed further in the chapters that follow.

3. Digital Smoothing and Multivariate Calibration

3.1 Introduction

Digital smoothing filters are widely used in analytical applications in chemistry to increase the *S/N* ratio in the data. Their popularity is in part attributable to a certain nostalgia associated with signal averaging in univariate methods, where an enhancement in the signal-to-noise on the order of $\sqrt{m}$ can be achieved by averaging *m* repeat sample measurements. With a progression to multichannel instrumentation this principle still holds – a noise reduction on the order of $\sqrt{m}$ at each channel in the vector measurement can be realized by averaging – however, quite often averaging is applied not over repeated measurements, but over adjacent channels in the vector measurement, a practice that is commonly known as *smoothing*. This 'short-cut' is very often taken in chemistry using polynomial least-squares smoothing filters, or Savitzky-Golay filters [22], as they are commonly known in chemistry circles.

In certain situations, a $\sqrt{m}$ reduction in error variance can be achieved in this manner, but these conditions rarely occur in practice. The averaging theorem more often than not crumbles when stretched in this fashion because the true signal is seldom static over the channels being considered, and as a result less-than optimal noise reduction occurs, and the original characteristics of the signal (and noise) are distorted. Chemical literature on the operational details of SG smoothing is plentiful, although these works tend to concentrate on the specific issue of *S/N* enhancement when a single vector measurement is considered in isolation [26, 27]. As a consequence, *S/N* is used as a univariate concept in a multivariate measurement, and only minor concern is demonstrated for the subtle side-effects of digital smoothing. The effect of signal distortion has been investigated in a limited number of studies [26, 28] resulting in some

general recommendations for choosing smoothing filter parameters, although these are entirely empirical. As a rough guide, it is suggested that the width of the smoothing filter be smaller than the signal features of interest. It is important to note that these empiricisms were developed for univariate procedures on multivariate data, such as peak maxima elucidation, and quantitation from peak area or peak maximum. While these sorts of methods are still in use today in some applications, the same guides are not ensured to prove useful in multivariate methods. The distortional influence of the filters on measurement errors, on the other hand, has received little more than a passing mention in the chemical literature.

While it has long been established that some amelioration in the *univariate S/N* ratio can result from digital smoothing, to our knowledge, the theoretical implications of digital smoothing have never been investigated in the expanded context of multivariate calibration. While the univariate *S/N* metric may be useful in certain cases, as was noted above, it is unrelated to the multivariate *S/N* in the majority of multivariate applications, and therefore more must be considered for multivariate calibration. The signal distortion resulting from digital smoothing can obviously have dire consequences in multivariate analyses, since analyte selectivity is already at a premium in most cases. And one can anticipate that the alteration of the properties of the measurement errors will also have some influence on the mathematical methods employed in multivariate calibration, given that certain measurement error structures are presumed to exist.

In this chapter, the effects of symmetric digital smoothing filters, in particular Savitzky-Golay polynomial filters, on multivariate calibration are explored. Following a theoretical examination of the implications of applying such filters, attention will be turned to the side-effects of smoothing and their effect on the predictive success of the multivariate calibration model. To minimize the number of variable factors in this investigation, the research will focus on calibration systems with well-defined rank, treated using principal

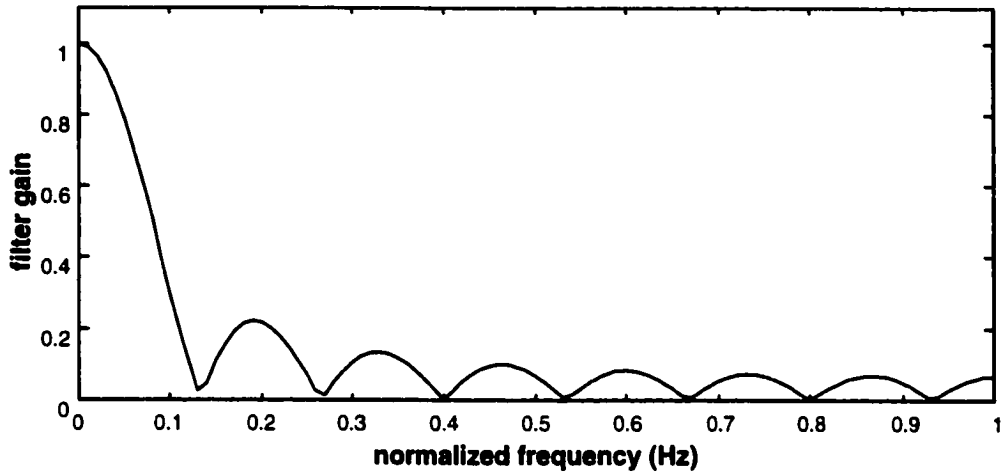


Figure 3.1 Transfer function for a 15-point moving-average (zero-order polynomial) Savitzky-Golay smoothing filter. The frequency axis has been normalized by the Nyquist frequency for generality.

components regression (PCR), and maximum likelihood principal components regression (MLPCR) [10].

3.2 Theoretical Considerations

3.2.1 Characteristics of Digital Smoothing Filters

The Savitzky-Golay smoothing filter has two user-adjustable parameters: the width of the filter, and the order of the polynomial used to approximate the signal. Perhaps the simplest type of SG smoothing filter would be the moving-average or running-average filters, which are commonly implemented by simply averaging the points in the filter windows to yield the 'filtered' signal vector. Although not often thought of as a least-squares polynomial filter, the moving-average is simply a zero-order polynomial fit to the data (fitting a line to the data with no slope). An examination of the transfer function of a typical zero-order SG smoothing filter (**Figure 3.1**) shows that, like all smoothing filters, it is a low-pass operator in the frequency domain, allowing the lower frequencies in the signal to

pass through the filter unabated while severely attenuating the higher frequency components.

3.2.1.1 Filter order

A filter of higher-order can of course be obtained by the measures previously prescribed, and a time domain contrast between the application of a 7-point moving-average filter and a 7-point 4th-order SG smoothing filter is given in **Figure 3.2**. As expected, the noise rejection of the averaging (zero-order) filters appears more appreciable than the higher-order smoothing filters, an observation that can be rationalized by examination of the filter transfer functions (also in **Figure 3.2**). The use of higher-order polynomials to approximate the signal vector allows the filter to accommodate signals of substantially higher frequency than the simple average, and the frequency cutoff in the transfer function is correspondingly moved to a higher frequency for higher-order polynomial filters. Since more of the original signal will be allowed to pass through the filter, more of the noise will be allowed through the filter, and it can be anticipated that the higher-order filters will provide poorer noise reduction than the averaging filters. A quantitative measure of the reduction in noise achieved by a given SG filter can be calculated *a priori* using the filter coefficients:

$$\frac{\sigma^2_{filtered}}{\sigma^2_{unfiltered}} = \sum_k c_k^2 \qquad (3.1)$$

where $\sigma^2_{unfiltered}$ is the variance of the noise before filtering, and $\sigma^2_{filtered}$ is the variance of the filtered noise. **Equation 3.1** is generally applicable to any non-recursive filter, although it is restricted to those cases in which the signal is corrupted with *iid* (white) noise. For 7-point SG smoothing filters, for example, the anticipated noise reductions are:

Zero-order	First-order	Second-order	Third-order
0.1429	0.1429	0.3333	0.3333

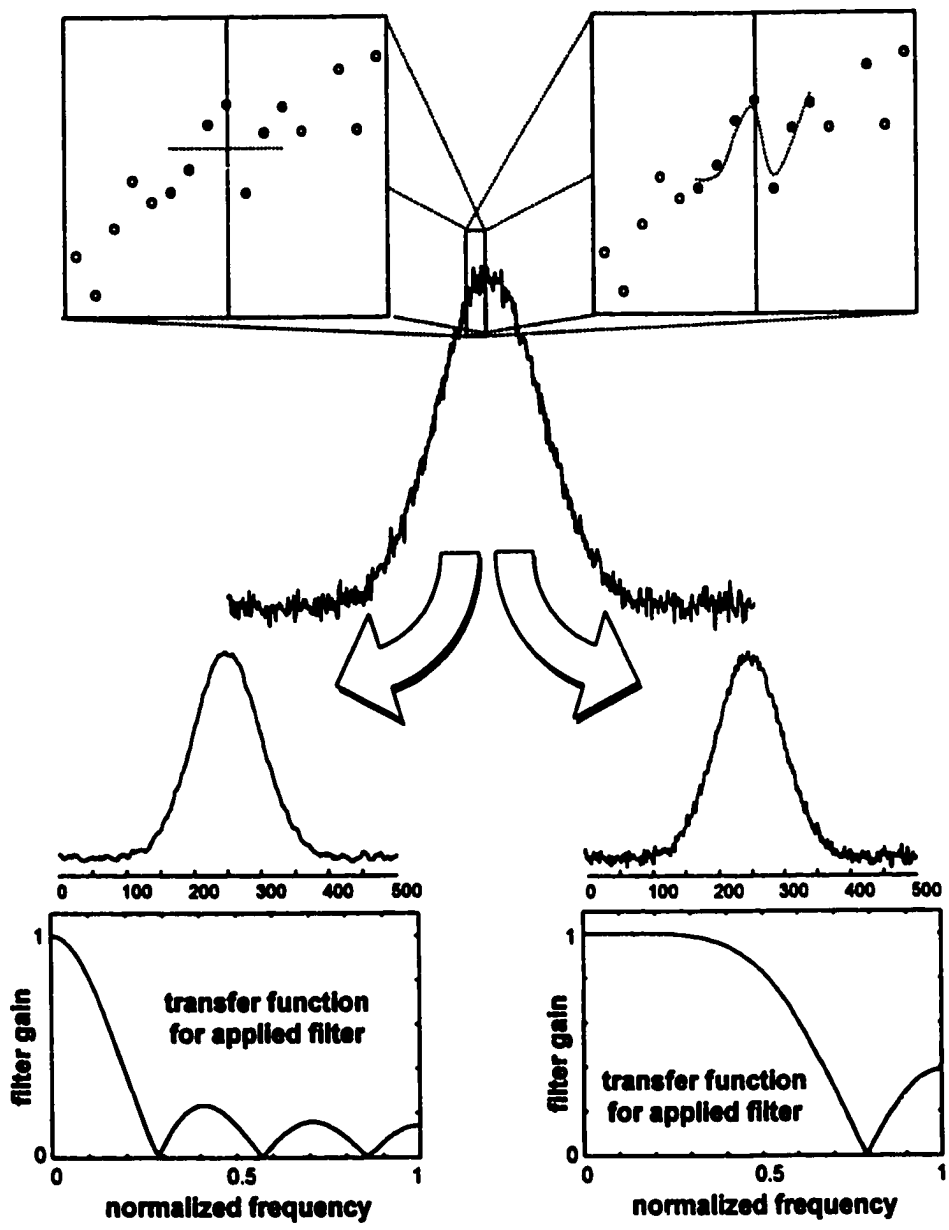


Figure 3.2 A comparison of the use of (right) a 7-point moving-average smooth, and (left) a 7-point quartic polynomial smooth. The moving-average filter achieves more substantial noise reduction due to its more aggressive attenuation of the signal in the frequency domain - only the very low frequencies are unattenuated.

As a consequence of the mathematics, the filter coefficients for zero- and first-order smoothing filters are identical, and as a result their noise rejection properties are identical.

While filters of higher order are not as proficient at reducing the noise variance, it can be said that these filters are better suited to chemical signals of interest than low-ordered filters, since vector measurements in chemistry (*e.g.*, spectra, chromatograms) are usually more reasonably modeled using quadratic or cubic local models than lines with zero slope. This can be reasoned from a frequency perspective as well, since the averaging filters have a lower frequency cutoff than the higher-order smoothing filters, and will therefore be more likely to distort the chemical signals of interest which typically reside at lower frequencies. A close-up illustration of this local modeling ability is given in **Figure 3.3**.

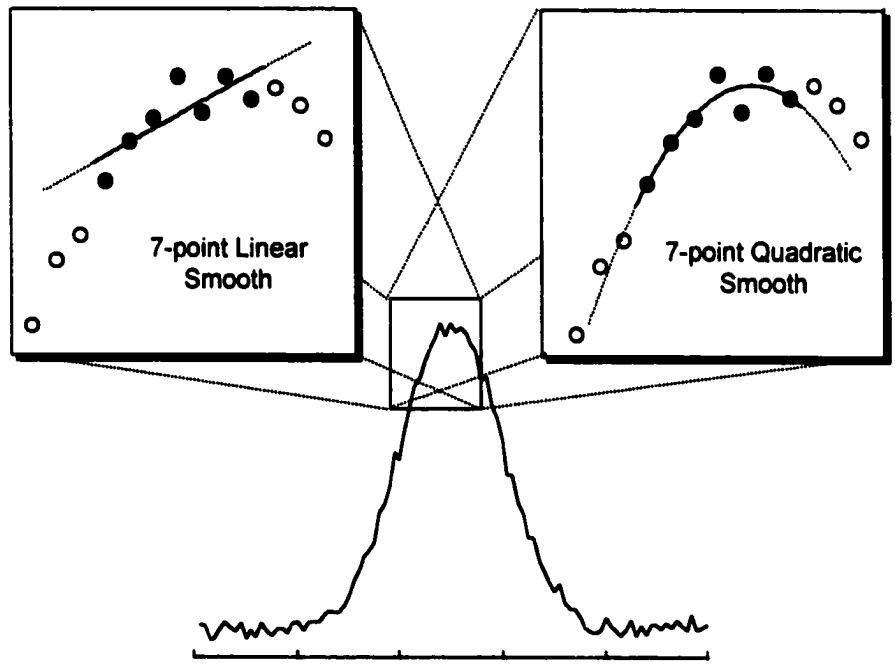


Figure 3.3 Illustration of the importance of the local modeling ability of the polynomial model in minimizing signal distortion. Higher-order polynomial filters will invariably model the true signal better, however lower noise rejection results.

3.2.1.2 Filter Width

In addition to the adjustment of the order of the polynomial function in the SG filter, the width of the filter can be altered. For 31-point smoothing filters, the following improvements in error variance are achieved:

Zero-order	First-order	Second-order	Third-order
0.0323	0.0323	0.0727	0.0727

Since a far greater number of points are used in the estimation, a greater reduction in noise is achieved with the wider filters than was observed for the 7-point filter values above. While the wider smoothing filters are clearly better at reducing the noise variance in the signal, they are much less effective in locally modeling the underlying changes in the signal vector itself. A visual comparison of a noisy signal treated with SG smoothing filters of varying widths is shown in **Figure 3.4**, and substantial distortion is readily observed with the wider filters. The frequency responses of the SG smoothers confirm these observations; several transfer functions for filters of varying widths are presented in **Figure 3.5**. The transfer functions confirm intuition in that the wider filters correspond to heavier filtering, *i.e.*, heavier attenuation in the frequency domain. It can also be seen from the transfer functions in **Figure 3.5** that the wider smoothing filters should have a greater distortional impact on the chemical signals of interest, since the frequency cutoff is pushed to lower frequencies as the width of the filter is increased. This effect of smoothing filter application is extremely difficult to quantify in a general way, although it can be said that in general, features in the signal vectors are broadened and flattened, which geometrically corresponds to a reduction in the length of the signal vectors. If several different mixture spectra are considered to be the signal vectors, it can be anticipated that the broadening and smudging of the spectral features will also decrease the angle between the vectors. The extent of this distortion will depend not only on the type of filter but

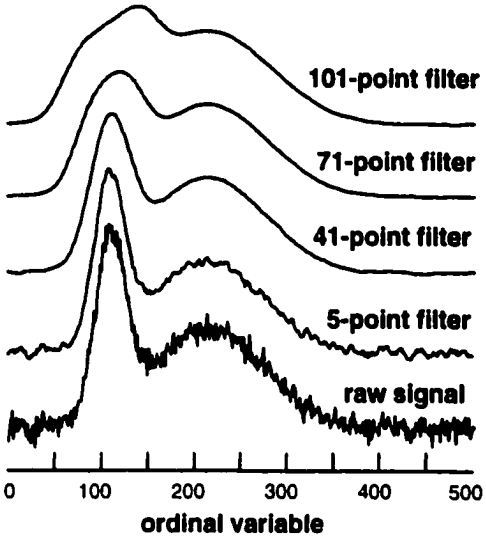


Figure 3.4 Example of the increased distortion observed in the signal features with increased filter widths.

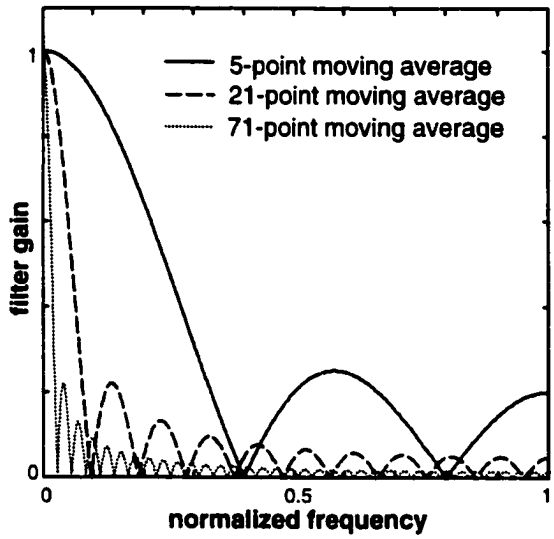


Figure 3.5 Transfer functions of zero-order smoothing filters with a variety of widths.

also on the characteristics (frequency content) of the noise-free signal, so the effects of filter distortion in multivariate calibration are not immediately obvious.

The previous paragraphs make clear that a trade-off exists with the two adjustable parameters of the polynomial smoothing filters. Larger window sizes and smaller polynomial orders correspond to heavier filtering (higher noise reduction, and a sharper, and lower frequency cut-off), and thus more effective noise reduction; however these desirable results come at the expense of increasing signal degradation. Previous studies in the literature have reached similar conclusions through a variety of empirical means. As noted in this chapter's introduction, however, the third implication of digital smoothing – the correlation of measurement errors – has yet to be addressed in any completeness.

3.2.1.3 Correlation of Noise

The use of a digital smoothing filter not only distorts the frequency characteristics of the true signal of interest, but also the frequency characteristics of the measurement errors. It is often presumed that the measurement errors in

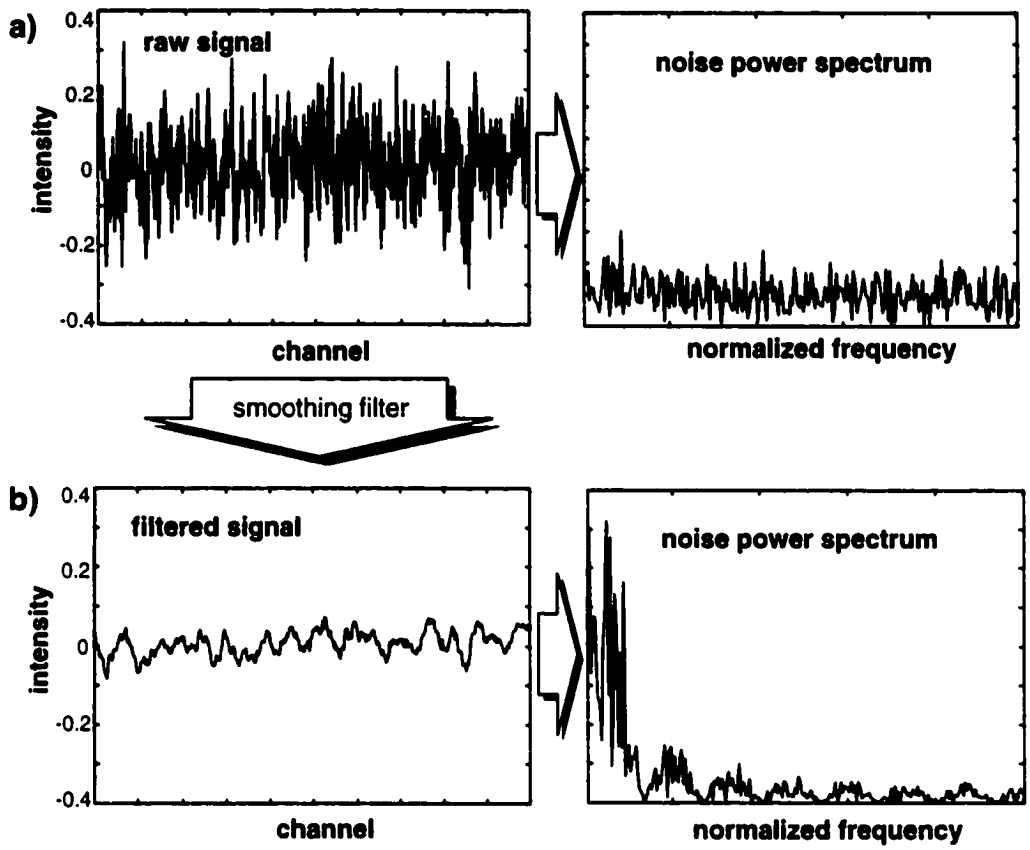


Figure 3.6 **a)** A vector of uncorrelated measurement errors, and the calculated noise power spectrum for the given noise sequence (white noise). **b)** The vector of smoothed noise values from a) using a 15-point moving-average filter. The noise power spectrum for these smoothed values is also shown (coloured noise).

the signal vector are *iid* (as was required for **Equation 3.1** to be valid). **Figure 3.6** illustrates the distortional effects of a 15-point moving-average filter on white noise in the time, and frequency domain. Because the filters heavily attenuate the higher frequency components of the signal, they also heavily attenuate the higher frequency components of the measurement errors, thereby introducing a low-frequency dominance in the noise power spectrum. As discussed in **Section 1.3.1.1**, this low-frequency dominance is characteristic of correlated measurement errors. Therefore, by the rather innocent use of digital smoothing filters to suppress the contribution of measurement errors to the multivariate data, the analyst is *introducing* non-*iid* noise into the system. Like the previous

implications of smoothing filters, the extent of this noise distortion is directly dependent on the width and order of the smoothing filter used. Since wider filters have lower frequency cutoffs and more drastic attenuation of high-frequency components, they tend to introduce far greater discrepancies between the power of the lower frequency noise and the higher frequency noise, and thus a greater degree of 'colour' in the NPS. When considered from a time domain perspective, the wider filters correlate the measurement errors from many channels because many points are used to estimate the filter central point of the filter window. To derive a quantitative expression for the effect of the filter on the measurement error structure of the data, recall (**Equation 1.41**) that the error covariance matrix for the measurement errors in a signal vector is given in the expectation by

$$\boldsymbol{\Sigma} = E\left(\mathbf{e}\mathbf{e}^{\mathrm{T}}\right) \tag{3.2}$$

The impact of the application of a filter matrix, $\mathbf{F}$, can be incorporated in this expression as

$$\boldsymbol{\Sigma}_F = E\left(\mathbf{F}\mathbf{e}\mathbf{e}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\right) \tag{3.3}$$

where $\boldsymbol{\Sigma}_F$ is meant to indicate the resulting error covariance structure of the filtered data. Removing the filter matrices from the expectation expression, **Equation 3.3** yields

$$\boldsymbol{\Sigma}_F = \mathbf{F}E\left(\mathbf{e}\mathbf{e}^{\mathrm{T}}\right)\mathbf{F}^{\mathrm{T}} \tag{3.4}$$

$$\begin{aligned}\boldsymbol{\Sigma}_F &= \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}^{\mathrm{T}} \\ &= \mathbf{F}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{F}\end{aligned} \tag{3.5}$$

If the measurement errors in the raw signal are *iid*, then **Equation 3.5** simplifies to

$$\boldsymbol{\Sigma}_F = \mathbf{F}^{\mathrm{T}}\mathbf{F}\sigma^2_{unfiltered} \tag{3.6}$$

Therefore, the impact of a particular smoothing filter on the measurement error structure can be determined in advance provided the error structure of the raw data is known or has been estimated.

If anything is *clear* from the above discussions of noise variance reduction, signal distortion and error correlation, it is that the overall result of digital smoothing is *unclear*. Smoothing filters achieve some reduction in the noise variance corrupting the signals, which is obviously beneficial if the data are to be used in multivariate calibration. This benefit is marred, however, by the two deleterious side-effects of the smoothing operation. Signal distortion is assured, which will adversely affect subsequent calibration proceedings due to the resulting decrease in *SEL* and *SEN*. The other side-effect, error correlation, can be expected to be particularly problematic with standard calibration algorithms such as PCR and PLS due to their implicit error structure assumptions. In an effort to alleviate this haze of qualitative uncertainty, the next section will involve a theoretical examination of the influence of symmetric digital smoothing filters in multivariate calibration.

3.2.2 Smoothing Filters and Calibration Theory

Any theoretical examination of the influence of smoothing filters in multivariate calibration must go beyond the individual effects of digital smoothing as outlined above, and answer the larger question: will digital smoothing enhance the predictive power of a multivariate calibration method? Unfortunately, dozens of different numerical methods exist for performing multivariate calibration and therefore any theoretical answer to the question might seem to rely on the particular calibration method employed on a given occasion. To side-step this undesirable complication, an approach was devised from the perspective of the net analyte signal, a concept discussed in **Section 1.5**. For the purposes of this discussion, we will assume that the determination of the NAS for the analyte is *the* goal of any calibration process. This assumption removes the necessity of dealing with particular calibration methods

(PCR vs. partial least-squares vs. continuum regression, *etc.*), since all of these regression methods are just slightly different methods of estimating the NAS.

A further complication is that the predictive power of the calibration model can be limited if the NAS has been poorly estimated. To eliminate this interfering aspect from the theoretical examination, we will therefore assume that the calibration conditions are perfect, *i.e.*, we have at our disposal the noise-free pure-component spectra of all active chemical constituents, or, a perfect calibration method which allows the NAS to be perfectly estimated by regression methods. Using the pure-component spectra, the true NAS vector can be calculated for the *i*th component using **Equation 1.53**, which has been reproduced here for convenience.

$$\mathrm{NAS}_i = \left(\mathbf{I}_n - \mathbf{S}_{-i}\mathbf{S}_{-i}^{+}\right)\mathbf{s}_i \tag{3.7}$$

In order to determine whether a digital smoothing filter will enhance the predictive power of the calibration procedure, we must use a theoretical metric that will correlate highly with the prediction error. The univariate signal-to-noise ratio of the data would be convenient since classically it is a direct expression of the precision of the measurement, and hence, in direct proportion to the precision of subsequent predictions; however, the univariate measure is wholly unsatisfactory in multivariate applications. Therefore, the multivariate *S/N* is preferred, previously defined (**Section 1.5.2**) [15] as

$$S/N_i = \frac{SEN_i}{N_i} = \frac{\left\|\left(\mathbf{I}_n - \mathbf{S}_{-i}\mathbf{S}_{-i}^{+}\right)\mathbf{s}_i\right\|}{N_i} \tag{3.8}$$

where N_i is the noise level of the data for component *i* in a multivariate sense. If it is assumed that the raw calibration data are corrupted with *iid* measurement errors, then the *S/N* for the unfiltered data is given by

$$\frac{SEN_i}{N_i} = \frac{\left\|\left(\mathbf{I}_n - \mathbf{S}_{-i}\mathbf{S}_{-i}^{+}\right)\mathbf{s}_i\right\|}{\sigma_{noise}} \tag{3.9}$$

as discussed in **Section 1.5.2**. With the measurement error distribution being equal in all spatial directions the noise level will be independent of the analyte.

The application of a filter matrix to the multivariate spectra data alters both the noise level and the spectral vectors defining the calibration space. Therefore we can consider a new NAS determined from the *filtered* pure-component spectral vectors, referred to below as NAS_F. This follows from the consideration that, in linear mixture theory, a mixture spectrum is considered to arise from the simple model

$$\mathbf{x} = \mathbf{Sc} \tag{3.10}$$

and the application of a filter matrix to a mixture spectrum is equivalent to applying the filter matrix to the pure-component spectra, since

$$\mathbf{x}_F = \mathbf{F}\mathbf{x} = \mathbf{F}(\mathbf{Sc}) \tag{3.11}$$

$$\mathbf{x}_F = (\mathbf{FS})\mathbf{c} \tag{3.12}$$

This convolution can be incorporated into the expression for the NAS in **Equation 3.7**, resulting in a filtered, NAS_F, displaced somewhat from the NAS determined from the unfiltered spectral data.

$$\mathrm{NAS}_{i,F} = \left(\mathbf{I}_n - (\mathbf{FS}_{-i})(\mathbf{FS}_{-i})^+\right)\cdot(\mathbf{Fs}_i) \tag{3.13}$$

Thus, the multivariate sensitivity for the filtered data is given by

$$SEN_F = \left\| \left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^+\right)\cdot(\mathbf{Fs}) \right\| \tag{3.14}$$

The reader will note that most of the descriptive subscripting has been dropped to reduce the clutter in this, and subsequent equations.

The application of the smoothing filter has an appreciable effect on the measurement errors, in addition to the signal effects accounted for by **Equation 3.13**. This can be incorporated if one recalls that the filter alters the measurement error structure via the error covariance matrix (**Equation 3.5**).

Since the noise corrupting the unfiltered calibration data is assumed to be *iid*, the error covariance matrix will be $\Sigma = \sigma^2 \mathbf{I}_n$, and the resulting expression for the noise in **Equation 3.8** becomes

$$\begin{aligned} N_F &= \sqrt{\mathbf{v}_F^T \Sigma_F \mathbf{v}_F} = \sigma \sqrt{\mathbf{v}_F^T \left(\mathbf{F}^T \mathbf{I}_n \mathbf{F}\right) \mathbf{v}_F} \\ &= \sigma \sqrt{\mathbf{v}_F^T \left(\mathbf{F}^T \mathbf{F}\right) \mathbf{v}_F} \end{aligned} \tag{3.15}$$

which can be expressed as

$$N_F = \sigma \cdot \left\| \mathbf{F} \mathbf{v}_F \right\| \tag{3.16}$$

if the $\mathbf{a}^T \mathbf{a} = \|\mathbf{a}\|^2$ identity is used. Since the contravariant vector, $\mathbf{v}_F$, is simply the NAS vector normalized to unit length, the following expression can be substituted for the filtered contravariant vector:

$$\mathbf{v}_F = \frac{\mathrm{NAS}_F}{\left\| \mathrm{NAS}_F \right\|} \tag{3.17}$$

Using **Equation 3.13** for the filtered NAS, and the definition of the contravariant vector above, **Equation 3.15** becomes

$$N_F = \sigma \cdot \left\| \mathbf{F} \cdot \frac{\mathrm{NAS}_F}{\left\| \mathrm{NAS}_F \right\|} \right\| = \sigma \cdot \left\| \mathbf{F} \cdot \frac{\left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^+\right) \cdot (\mathbf{Fs})}{\left\| \left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^+\right) \cdot (\mathbf{Fs}) \right\|} \right\| \tag{3.18}$$

for the noise level of the multivariate data after symmetric smoothing filter application.

Similar to the unfiltered case, the multivariate *S/N* ratio for the filtered data can be derived by combining **Equation 3.14** and **Equation 3.18** to give

$$S/N_F = \frac{SEN_F}{N_F} = \frac{\left\| \left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^+\right) \cdot (\mathbf{Fs}) \right\|}{\sigma \cdot \left\| \mathbf{F} \cdot \frac{\left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^+\right) \cdot (\mathbf{Fs})}{\left\| \left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^+\right) \cdot (\mathbf{Fs}) \right\|} \right\|} \tag{3.19}$$

which can be simplified to

$$S/N_F = \frac{\left\|\left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|^2}{\sigma\cdot\left\|\mathbf{F}\left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|} \tag{3.20}$$

Since $\|\mathbf{a}\|^2 = \mathbf{a}^T\mathbf{a} = \|\mathbf{a}^T\mathbf{a}\|$, the numerator of **Equation 3.20** can be expressed as the inner product:

$$S/N_F = \frac{\left\|\left[\left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right]^T\cdot\left[\left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right]\right\|}{\sigma\cdot\left\|\mathbf{F}\left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|} \tag{3.21}$$

Canceling the cross terms, and further expanding the product results in

$$S/N_F = \frac{\left\|\mathbf{s}^T\mathbf{F}^T\left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|}{\sigma\cdot\left\|\mathbf{F}\left(\mathbf{I} - (\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|} \tag{3.22}$$

With **Equation 3.9** expressing the *S/N* of the calibration data before filtering, and **Equation 3.22** expressing the *S/N* of the filtered data it is apparent that we have a standard metric for both the filtered and unfiltered case that should indicate whether calibration model performance is enhanced or degraded by the application of a symmetric digital smoothing filter.

To evaluate the relative performances of the filtered and unfiltered data in terms of multivariate calibration, we can further define a *theoretical performance ratio* (PR_{theo}) as

$$PR_{theo} = \frac{S/N_F}{S/N} \tag{3.23}$$

If this performance ratio exceeds unity then the *S/N* has been enhanced by filtering, and we would expect to see an improvement in the calibration performance as a result of the smoothing procedure. In other words, smoothing must result in an enhancement in the signal-to-noise ratio to be beneficial in multivariate calibration.

It can be shown that the numerator of **Equation 3.22** can be factored as

$$\left(\mathbf{s}^{T}\mathbf{F}^{T}\left(\mathbf{I}-(\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right)=\left[\mathbf{F}\left(\mathbf{I}-(\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right]^{T}\cdot\left[\left(\mathbf{I}-\mathbf{SS}^{+}\right)\mathbf{s}\right] \tag{3.24}$$

provided the filter matrix is symmetric ($\mathbf{F}^{T}=\mathbf{F}$). Using a property of vector norms, namely that $\|\mathbf{a}^{T}\mathbf{b}\|\leq\|\mathbf{a}\|\cdot\|\mathbf{b}\|$, we can see that **Equation 3.24** implies that

$$\left\|\mathbf{s}^{T}\mathbf{F}^{T}\left(\mathbf{I}-(\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|\leq\left\|\mathbf{F}\left(\mathbf{I}-(\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|\cdot\left\|\left(\mathbf{I}-\mathbf{SS}^{+}\right)\mathbf{s}\right\| \tag{3.25}$$

or, rearranging,

$$\frac{\left\|\mathbf{s}^{T}\mathbf{F}^{T}\left(\mathbf{I}-(\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|}{\left\|\mathbf{F}\left(\mathbf{I}-(\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|}\leq\left\|\left(\mathbf{I}-\mathbf{SS}^{+}\right)\mathbf{s}\right\| \tag{3.26}$$

Dividing both sides of **Equation 3.26** by σ, we get

$$\frac{\left\|\mathbf{s}^{T}\mathbf{F}^{T}\left(\mathbf{I}-(\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|}{\sigma\cdot\left\|\mathbf{F}\left(\mathbf{I}-(\mathbf{FS})(\mathbf{FS})^{+}\right)\cdot(\mathbf{Fs})\right\|}\leq\frac{\left\|\left(\mathbf{I}-\mathbf{SS}^{+}\right)\mathbf{s}\right\|}{\sigma} \tag{3.27}$$

or,

$$\begin{aligned}\frac{SEN_{F}}{N_{F}}&\leq\frac{SEN}{N}\\ S/N_{F}&\leq S/N\end{aligned} \tag{3.28}$$

Expressing this result in the performance ratio implies that

$$PR_{theo}=\frac{S/N_{F}}{S/N}\leq 1 \tag{3.29}$$

and, in the context of the stated assumptions, no gains can be made in multivariate calibration performance by preprocessing with a symmetric smoothing filter, because the multivariate *S/N* of the calibration data is degraded by filter application. Of course prediction errors can also originate from a poor estimate of the calibration model, but a general equation describing the effect of filtering on this process would be far more involved. In **Section 3.4** it is shown that smoothing can bring about improvements in calibration performance if a large proportion of the prediction error variance is attributable to inadequate

subspace estimation. Conditions under which this may be anticipated are investigated using simulated and experimental data.

3.3 Experimental

3.3.1 Simulated Data Sets

In order to systematically evaluate the performance of multivariate calibration procedures under the influence of symmetric smoothing filters, simulation studies were conducted. Simulations are instrumental in achieving this goal since so many factors can be of importance in calibration. The use of calibration systems with well-known and adjustable properties allows one to systematically alter the factors of interest while holding other factors constant, a task that would be near impossible with laboratory generated data.

The behavior of two multivariate calibration techniques, principal components regression and maximum likelihood principal components regression, was examined with respect to the degree of filtering under tightly controlled conditions. All simulations conducted in the course of this work mimicked three-component systems of well defined rank whose pure-component spectra were unimodal gaussian peaks. To standardize the 'instrument' sensitivity to each of the three components, the pure-component spectra were normalized to unit height. For convenience and clarity it was decided to define a set of "standard simulation conditions". Simulation studies of *particular* factors in calibration performance involve *deviations* from these conditions were noted, while all other controllable factors will be fixed at the values indicated as "standard".

For standard simulation conditions, the calibration set was generated as 20 mixtures in which concentrations of the 3 mixture components were drawn randomly from a uniform distribution between zero and one. The prediction sets consisted of 100 mixtures (concentrations drawn from the same distribution as in the calibration set).

An important consideration in these simulations was the initial *S/N* imparted on the data with *iid* noise. This was set at 565 according to the following formula

$$S/N = \frac{\|\bar{\mathbf{x}}\|}{\sigma} \tag{3.30}$$

where $\|\bar{\mathbf{x}}\|$ is the length of the mean spectral vector in the calibration data, and σ is the standard deviation of the noise (*ca.* σ = 0.01 at *S/N* = 565 for standard conditions). **Equation 3.30** can be derived from a projection of the noise (diagonal error covariance matrix) onto the normalized mean signal vector ($\bar{\mathbf{x}}/\|\bar{\mathbf{x}}\|$), or by simple error propagation formulae. To control the degree of spectral overlap (and hence the *SEL* and *SEN*), the pure-component spectra were generated such that the angle between the spectral vector of component two (the centre gaussian) and the other mixture components was 45 degrees in a 200 channel spectrum (*i.e.*, R = 0.707). This implies that the spectral angle between, component one and three is larger than 45 degrees, and so component two was used consistently as the analyte of interest for most of this work. The simulation results for components one and three were essentially identical to those for component two, and so only the behavior of the second component is shown in most figures. The gaussian peaks providing the spectral characteristics were given a peak width (σ_{peak}) of ten channels. A large baseline region was generated on either side of the peaks to alleviate any concerns with data distortion in the filtering process arising from edge effects. A typical set of calibration spectra under these standard conditions is shown in **Figure 3.7**, and the inset shows the pure-component spectra for the 3 analytes. To examine the effect of specific factors on calibration model performance with smoothing, parameters such as the *S/N* (**Equation 3.30**), spectral angle, and peak width were varied from standard conditions. A host of different polynomial smoothing filters could have been used, but highly-similar results were obtained for filters of varying orders; therefore, it was decided that a simple moving-average filter

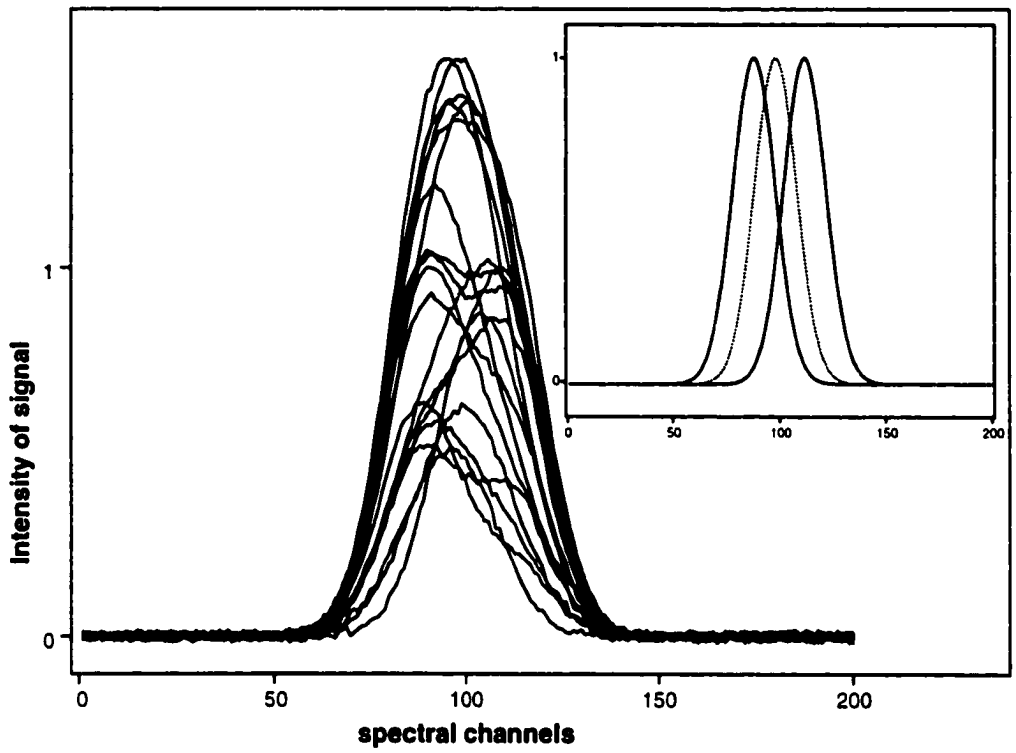


Figure 3.7 An example of the calibration spectra employed for the simulation studies under standard conditions. Also shown in the inset are the 3 pure-component spectra, with the middle (dotted) gaussian band corresponding to the analyte of interest, component 2.

offered representative behavior, and consequently, said filters were used in the simulation results presented in **Section 3.4.**

3.3.2 Experimental Data Sets

The effects of smoothing in practical applications were examined using a data set consisting of spectra of 128 metal ion mixtures (Cr(III), Co(II), and Ni(II)) employed previously by Wentzell *et al.* [10] These mixture spectra were truncated to a 418-588 nm wavelength range so as to not include portions of the spectra in which the data showed non-uniform noise characteristics. Sixty-four mixtures were randomly chosen as calibration samples, and the remaining sixty-

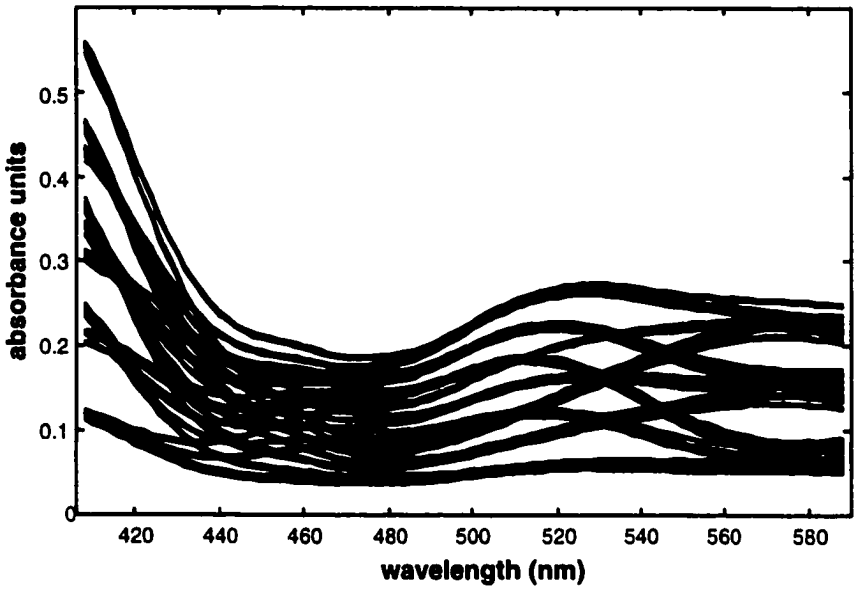


Figure 3.8 Experimental data used to validate the simulation studies. The data shown consist of 128 UV-Vis mixture spectra for metal ion mixtures in nitric acid.

four samples were taken to constitute the prediction set. The 128 calibration and prediction spectra used are shown in **Figure 3.8**.

3.3.3 Computational Aspects

All computations performed in the course of the work presented in this chapter were carried out on a Sun Microsystems Sparc Server 1000 with 4 parallel 50 MHz processors and 230 MB of memory. All simulation scripts were written in MATLAB v. 5.1 for the Unix platform (The Mathworks, Natick, MA).

3.4 Results and Discussion

In **Section 3.2** it was demonstrated theoretically that when the calibration model (*i.e.*, the NAS vector) is reasonably well determined, the application of a symmetric smoothing filter to multivariate calibration data with *iid*-normal errors impairs the predictive success of the system. In order to confirm this result, simulations were conducted comparing the theoretical result to observed

behavior. A useful diagnostic for these post-prediction comparisons was an *observed* performance ratio (PR_{obs}), defined as

$$PR_{obs} = \frac{RMSEP}{RMSEP_F} \tag{3.31}$$

where *RMSEP* and $RMSEP_F$ are the root mean-squared errors of prediction for the unfiltered and filtered data, respectively. The *RMSEP* is essentially the standard deviation of the prediction errors, and is calculated according the formula

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}{m}} \tag{3.32}$$

where y_i is the *known* concentration for the ith prediction sample, and $\hat{y}_i$ is the predicted concentration for the ith sample. The value of $\hat{y}_i$ will depend whether the filtered data, or unfiltered data is being used (given $RMSEP_F$, or *RMSEP*). According to this definition, a PR_{obs} value greater than 1 would indicate that the filtered data gives a better calibration result (in terms of predictive ability) for the filtered data. By direct analogy to univariate calibration, PR_{obs} should be equal to PR_{theo}, which is based on multivariate *S/N* values, and we can anticipate a high degree of correlation between the theoretical and observed *PR* if the theoretical results hold in practice.

Thirty replicate simulations were carried out under the standard simulation conditions given in **Section 3.3.1**. For each set of data, the PR_{theo} was calculated as the size of the moving-average filter was increased from one channel (no filtering) to 61 channels (extremely heavy filtering). The PR_{obs} resulting from PCR predictions was also calculated. The results (averaged over the 30 replicates) are shown in **Figure 3.9** for components 1 and 2 (component 3 is similar to component 1 by symmetry). **Figure 3.10** gives the results of an identical simulation using a second-order smoothing filter. In both cases, the theoretically calculated ratios are in very good agreement with the observed

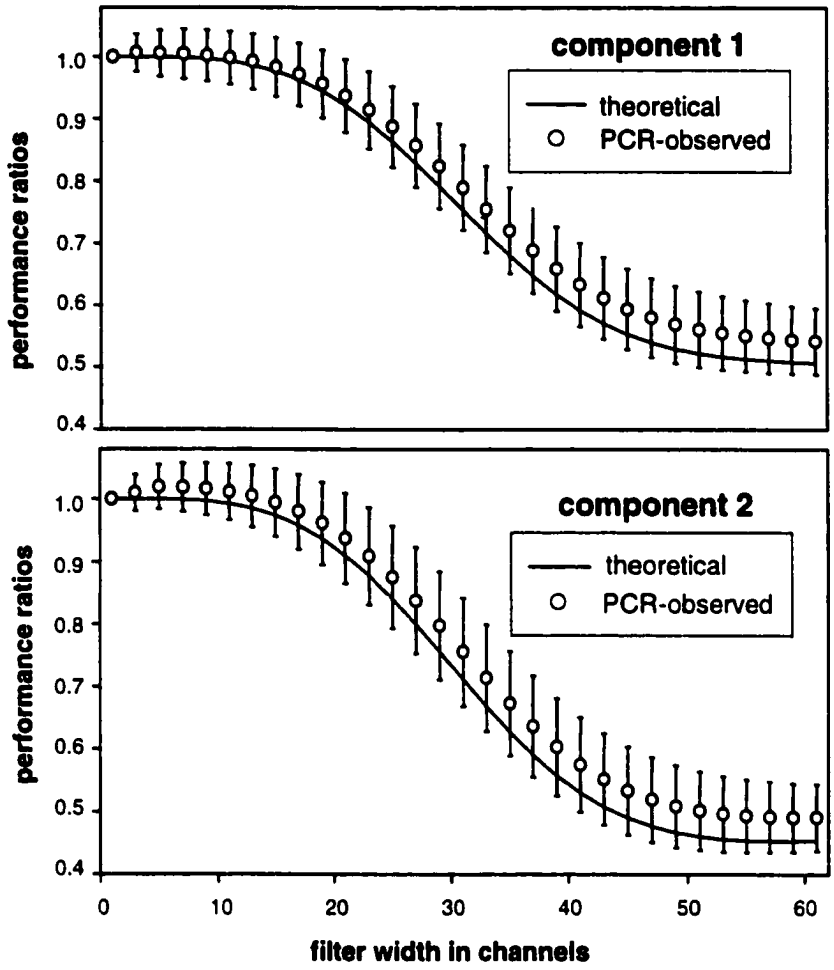


Figure 3.9 Theoretical and PCR-observed performance ratios for multivariate calibration and prediction under the influence of a moving-average (zero-order) Savitzky-Golay smoothing filter. The observed values are the resulting average from 30 replicate trials (error bars represent ± 1s)

performance ratios, although the observed *PR* is always slightly higher than the theoretical value and marginal gains from filtering can be observed at very low filter sizes. This yields an apparent contradiction to the theoretical predictions. The possible reasons for this deviation from theory were subsequently investigated.

The modest gains in predictive performance that result from filtering in **Figures 3.9** and **3.10** can be justified as follows. In **Section 3.2**, it was assumed that the calibration space is known exactly, so that errors in prediction are

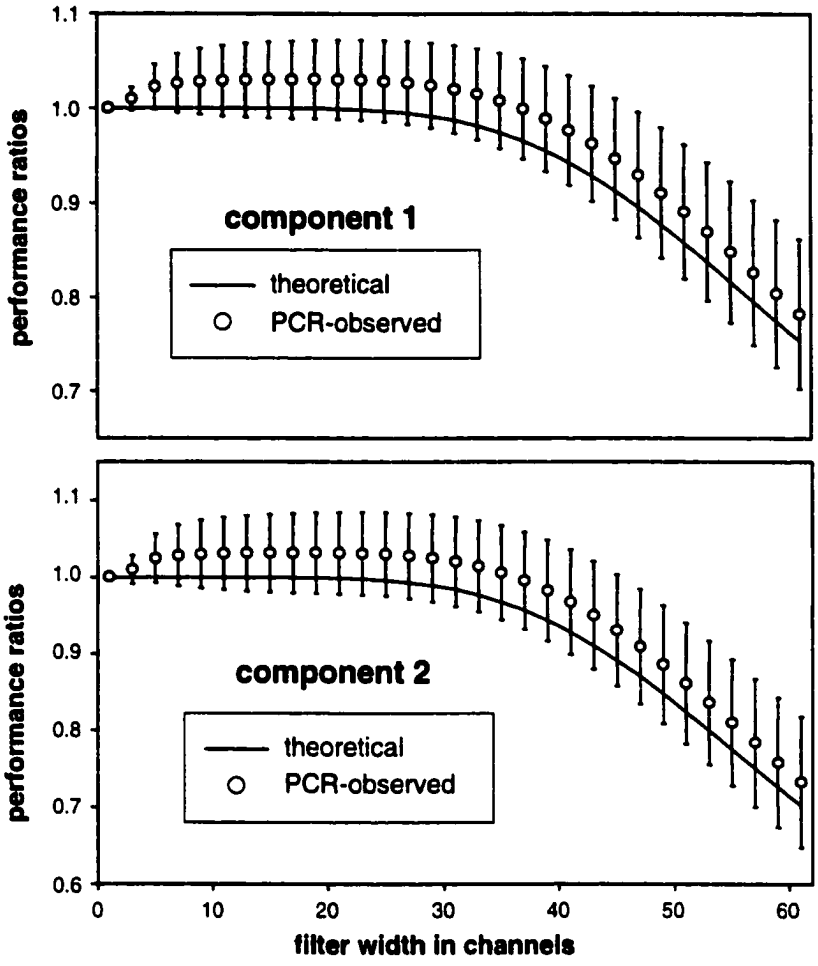


Figure 3.10 Theoretical, and PCR-observed performance ratios for multivariate calibration and prediction under the influence of a quadratic (second-order) Savitzky-Golay smoothing filter.

entirely associated with the prediction step (no error can be attributed to the accuracy of the estimated calibration parameters). Of course, in practice, there is prediction error stemming from inaccurate estimations of the calibration model itself, since measurement error will also affect the calibration step. The errors in the calibration will of course depend on factors such as the number of calibration samples, the design of the experiment, and the calibration method.

With methods such as PCR, errors in the calibration model can be attributed to two main sources. The first is the estimation of the PCA subspace

of the mixture spectra, which in this work will be referred to as subspace estimation error. Since the NAS vectors for individual components are assumed (in the calibration step) to lie in this subspace, an error in the estimation of this subspace will result in a reduction in the length of the NAS vectors, since they will be projected into a space outside that of the pure-component spectral vectors. This will result in a loss of sensitivity, and hence a reduced *S/N*. The second type of calibration error, which will be referred to as regression error, is introduced in the calculation of the regression vector from the latent variables and reference values. Errors at this stage depend largely on the design of the calibration set and errors in both sets of measurements, and can ultimately be classified as bias for a given calibration model, since they will lead to inaccuracy in the model.

In the application of digital smoothing filters to multivariate calibration data, it is sometimes possible that the smoothed data will result in an improvement in the determination of the calibration model that overcomes the degradation in the performance at the prediction step. In this work, it has been found that this is particularly true when there is a large uncertainty in the primary estimation of the spectral subspace. The extent of improvement is difficult to predict, however, since it depends on factors such as noise level, spectral shape, and calibration design. Some of these factors were further investigated using simulated data.

3.4.1 Noise Level

Without question the level of noise prior to filtering plays a large role in the success of the calibration. Indeed, it is this obvious property that leads the analyst to smooth data in the first place. A reduction in the noise level corrupting the data decreases the uncertainty in both the prediction and calibration steps. However, if applying a smoothing filter brings about the noise reduction, a decrease in the true *S/N* will result as outlined in **Section 3.2**. Simulation studies were conducted to examine the degree of improvement one sees from filtering as a function of the noise level of the data. To illustrate the dependence of calibration performance on filter characteristics, 25 replicate calibration sets were

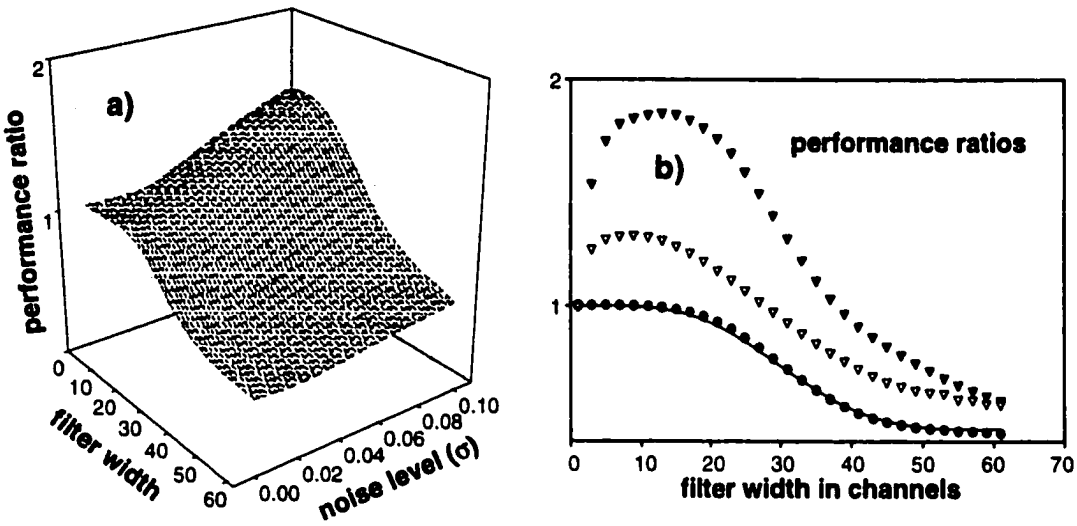


Figure 3.11 **a)** Performance ratios observed for PCR over a variety of filter sizes as the noise level of the data is systematically changed from 0.001 to 0.1 (σ) for standard simulation conditions. **b)** An example of the individual performance ratios observed with specific noise levels of 0.001 and 0.1 (σ). PR_{theo} (—), observed *PR* values for: σ=0.001 and 20 calibration samples (●), σ=0.1 and 20 calibration samples (▼), and σ=0.1 and 40 calibration samples (▽).

constructed under the standard conditions described previously, however the noise level of these data sets was varied from a standard deviation of 0.001 to 0.100 (the maximum signal in the calibration set was around 1.5). This roughly corresponds to *S/N* values of 5650, and 56 as defined by **Equation 3.30**. **Figure 3.11a** shows the change in the observed performance ratio as a function of noise level and filter width for a simple moving-average filter. **Figure 3.11b** shows a comparison of the observed and theoretical *PR*s for noise levels of 0.001 and 0.1. From the figures, it is apparent that there is a relative improvement in the predictive ability for the filtered data for small filters at higher noise levels, although the *RMSEP* improvements over the unfiltered systems are still quite negligible (less than a factor of 2). Qualitatively, it could be said that the gains in *RMSEP* achieved by filtering increased linearly with the standard deviation of the measurement noise corrupting the unfiltered data. It should also be noted that the absolute performance of the model under these circumstances is quite poor,

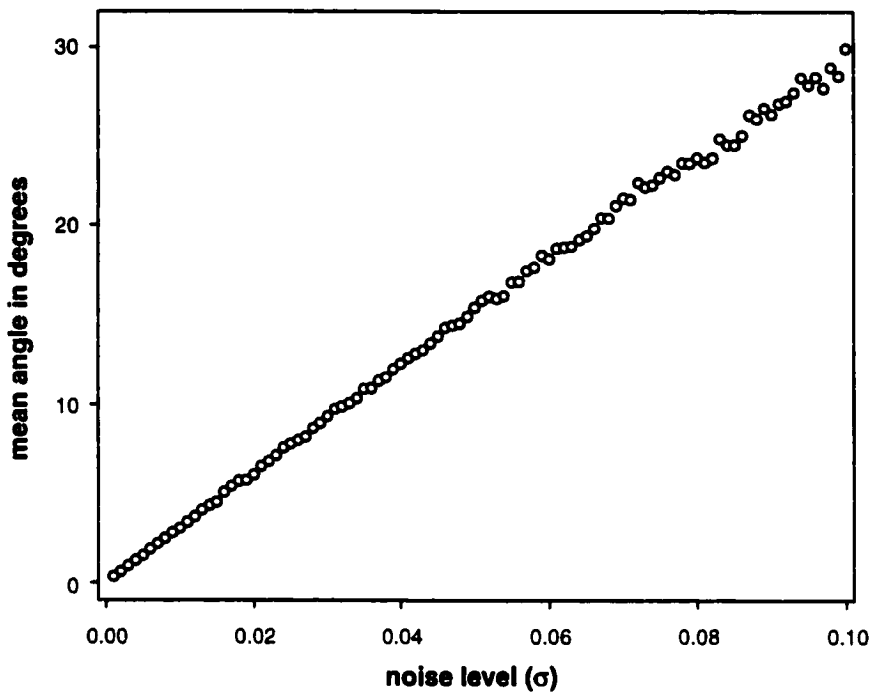


Figure 3.12 Plot of the mean angle between the true and PCA-estimated pure-component subspace (25 repeat measurements) as a function of the level of the error corrupting the spectral vectors.

in the worst case giving a relative prediction error ($RMSEP/\bar{c}$ where $\bar{c}$ is the mean concentration of the component in the prediction data) of 25% with no filtering. So although there is a relative improvement, the results are poor to begin with.

As suggested earlier, the small enhancements brought about by filtering under high noise conditions are largely due to the improvement in the estimate of the calibration model. To demonstrate this is the case, **Figure 3.11b** also shows a PR_{obs} curve generated under conditions of high noise (σ = 0.1) and 40 calibration samples. The larger number of calibration samples improves the quality of the calibration model, but not the error in the prediction step. This is reflected by the fact that the PR_{obs} curve moves closer to the theoretical model, suggesting that the enhancements due to filtering do indeed arise in the calibration step.

One of the reasons for improvement of the calibration model after filtering under high noise conditions is a superior estimation of the spectral subspace by PCA. PCA in essence gives the best *p*-dimensional estimate of the subspace spanned by the pure spectral vectors. With very high noise levels, the latent structure of the spectral data is much harder to extract from the data, and the subspace estimation is greatly hampered. **Figure 3.12** shows how the angle between the PCA-estimated subspace and the true spectral subspace deviates as a function of measurement noise. When this deviation becomes large, the sensitivity of the calibration must be reduced since it relies on a projection of the NAS vector into a different subspace. This is illustrated in **Figure 3.13**. **Figure 3.13a** shows the PR_{obs} curves for the three components for one simulation under high noise conditions. The maximum enhancement is about a factor of 2. **Figure 3.13b** shows the theoretical *SEN/N* ratio for the filtered data which suggests no such enhancement. **Figure 3.13c**, however, shows the theoretical *SEN/N* for the filtered data based on the PCA estimated subspace; *i.e.*, it is the norm of the NAS vector for each component projected into the suboptimal subspace. This represents the *best SEN/N* that could be realized using this rather inaccurate subspace estimate as determined by PCA. From the figure, it is clear that there is an improvement in the *SEN/N* resulting from a more reliable decomposition of the filtered subspace, although the projected *SEN/N* is always below the theoretical *SEN/N*.

A few other features of **Figure 3.13** should also be noted. At first glance, the curves in **Figure 3.13c** do not appear to match up with those in **3.13a**, but it is the relative change in the *SEN/N* that is important. It is also clear from **Figure 3.13** that the percentage changes in *SEN/N* in **Figure 3.13c** which are about 10 – 20 %, do not fully account for the improvements in the performance ratio in **Figure 3.13a**, which are 60 – 80%. This is because the reduction in maximum sensitivity from subspace estimation error is only one factor leading to a reduction in the predictive ability for the unfiltered case. Errors in determining the regression vector in the calibration step (regression error) will also contribute in

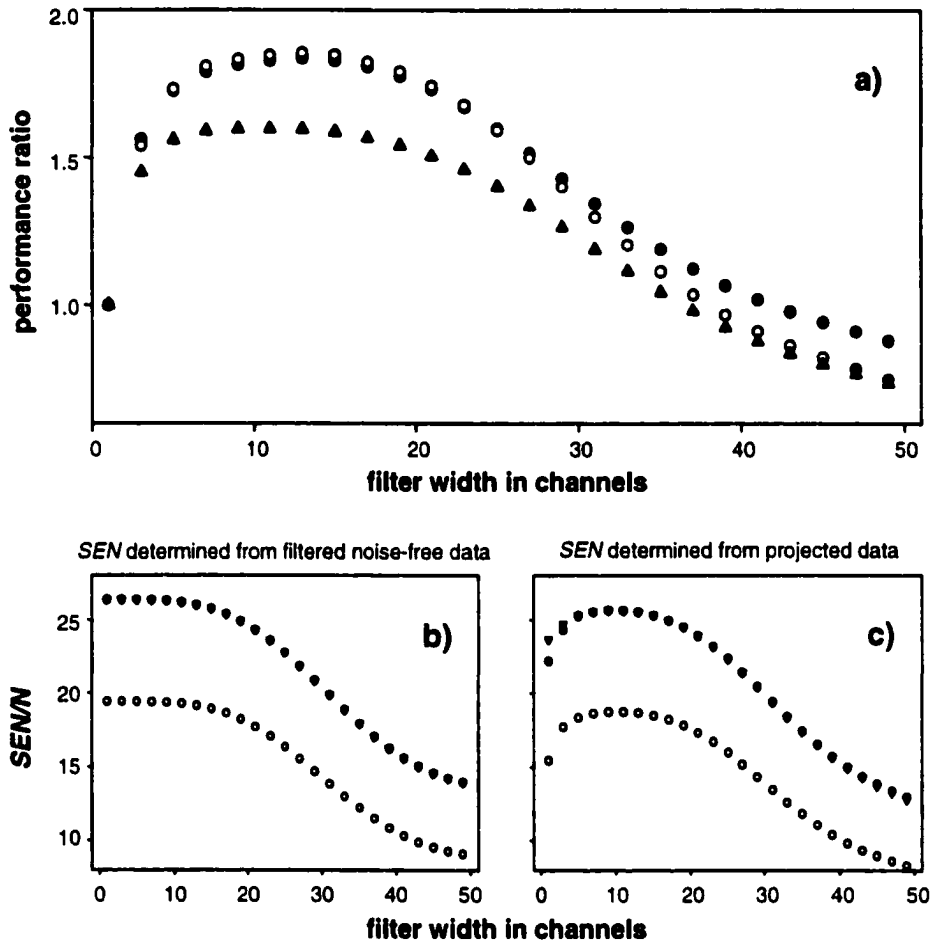


Figure 3.13 **a)** observed performance ratios for components 1 (•), 2 (○) and 3 (▲) at a noise level of 0.1 (σ_{noise}). **b)** Theoretical *S/N* curves as calculated for all 3 components. **c)** Theoretical *S/N* curves determined for the filtered data based on the subspace as estimated by PCA.

similar fashion. In fact, error propagation formulae [29] show that the contribution of calibration variance to the prediction error is directly related to the square of the norm of the regression vector (*i.e.*, $1/SEN^2$) as well as to the variance in the measurements. Therefore a reduction in the sensitivity also increases this regression error.

The results outlined above indicate that some enhancement in predictive ability can occur due to improvements in the subspace estimate. However, in

these simulations such improvements were quite small and only occurred when the spectral data were severely corrupted with noise.

3.4.2 Spectral Correlation

Another major factor affecting the accuracy of the prediction after filtering is the degree of spectral correlation, which in this work was measured as the angle between the pure-component spectral vectors. In addition to reducing the noise level of the data, the applied smoothing filter distorts the spectral vectors. This distortion manifests itself geometrically as a reduction in the angle between the pure-component spectra. When pure-component spectra are highly overlapped, filtering should increase the degree of correlation more quickly, leading to a negative impact on calibration due to a reduction in multivariate sensitivity. This supposition was investigated by simulations involving 25 replicate calibration sets for each spectral angle selected from 10 to 80 degrees. **Figure 3.14a** shows the observed performance ratios as a function of spectral

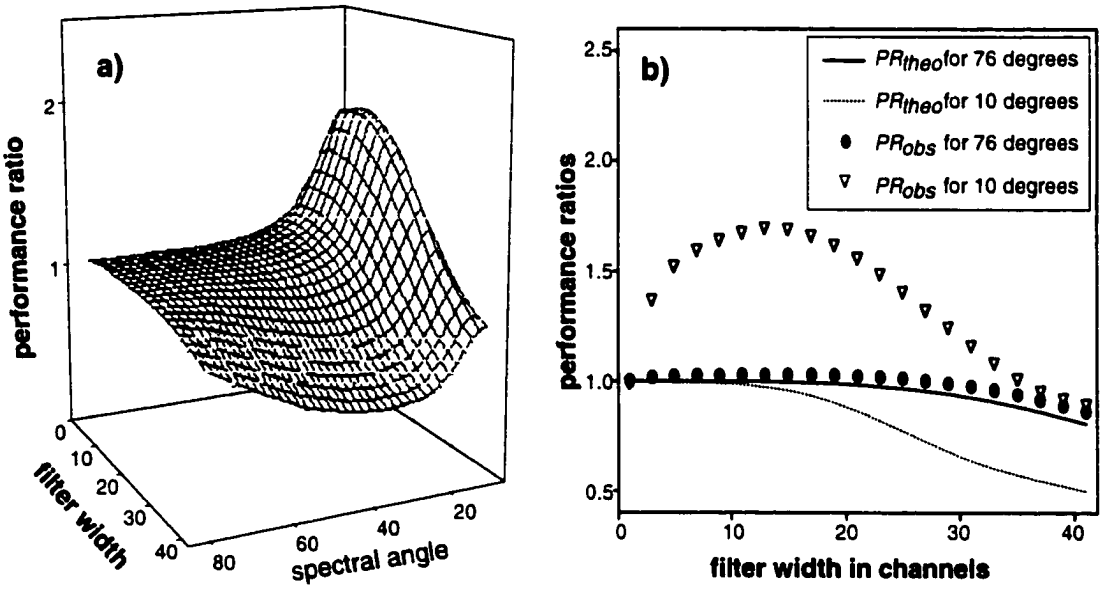


Figure 3.14 a) Performance ratios observed over all filter widths for spectral angles varying between 10° and 85° for standard simulation conditions. **b)** Observed performance ratios for calibration systems comprised of pure-component spectral overlap of 10° and 76°.

angle and filter width for a moving-average filter. **Figure 3.14b** compares the observed and theoretical *PR* values for spectral angles of 10 and 76 degrees. The experimental values obtained at high spectral angles (*i.e.*, close to 90 degrees) are in reasonable agreement with the calculated values of PR_{theo}. At low spectral angles, some improvement in the prediction error can be observed with filter application, but as in **Section 3.4.1**, only about a two-fold enhancement is observed at best.

The effect of spectral correlation in smoothing filter performance can be explained in much the same way as the discussions regarding the noise in **Section 3.4.1**. The accuracy of a PCA estimated calibration space is dependent upon the recognizable latent structure exhibited by the mixture spectra. This structure can be difficult to extract when the noise severely obfuscates the true characteristics of the mixture space (as noted in **Section 3.4.1**). And, like any sort of signal-to-noise consideration, one can consider either the signals being 'too small' or the noise being 'too large'. If the important structure in the subspace is very subtle, *i.e.*, small in magnitude relative to the noise level, then estimates of the true subspace will tend to be quite imprecise. Conversely, if the structure is very prominent, PCA subspace estimates will tend to be far more accurate. A simplified 2-dimensional illustration of this noise-spectral angle duality is given in **Figure 3.15**, using the simple analogy of putting a line through two observation points, **a** and **b**. In the first scenario (**Figure 3.15a**) the experiment happens to yield two sets values which produce model estimates that represent significant departures from the *true* model (defined by $\mathbf{a}^{\circ}$ and $\mathbf{b}^{\circ}$). With a large separation between the two points (large spectral angle), however, these deviant measurements still result in a reasonably estimated model. In measurement scenario two (**Figure 3.15b**), the spectral angle is very small, and therefore small deviations from the true values of $\mathbf{a}^{\circ}$ and $\mathbf{b}^{\circ}$ yield model estimates that can be almost orthogonal to the true model space.

This principle has an equivalent realization in real-life. If one wishes to carry a lengthy item in their arms, such as a wooden plank, they are much more

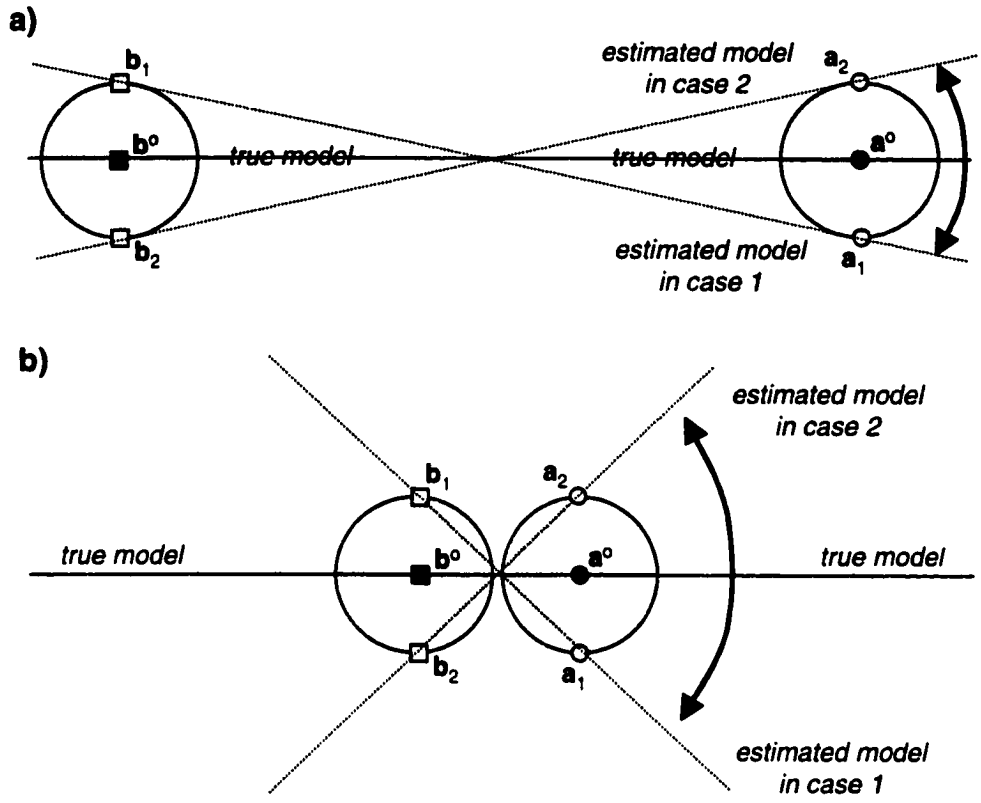


Figure 3.15 Illustration of the dual effects of noise level (shaded area indicates a 1σ level) and spectral angle on the precision of the calibration subspace estimation. **a)** A representation of a large subspace angle, with a noise level *N*. The orientation of the calibration subspace is uncertain, but the uncertainty in the estimate is relatively small. **b)** A representation of a similar system corrupted by a noise level, N, but with very small subspace angles. Because the pure-components are very close together, the noise contributes a much greater uncertainty in the subspace orientation.

successful at balancing the item if their arms are widely spaced apart. With their arms essentially together, it is very difficult to balance the item. In this analogy, the angle between the outstretched right and left arms is like the subspace angle between the pure-component spectra, and the noise would be akin to 'natural instabilities' and other shakes and wobbles.

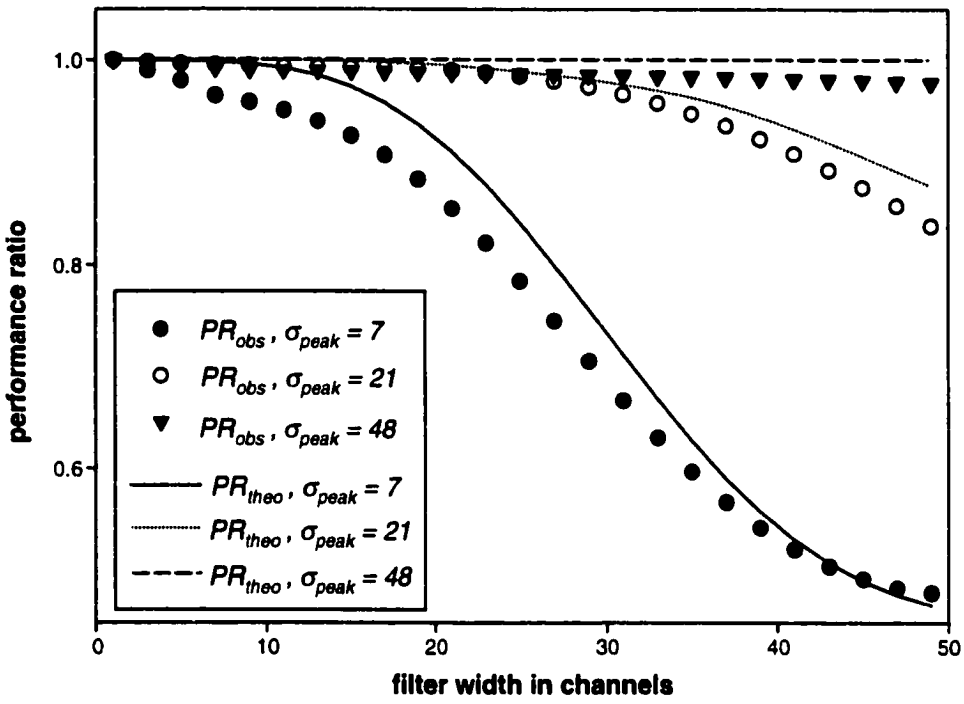


Figure 3.16 Comparison of the PR's (both theoretical, and observed) for spectral data with varying frequency content. Small σ_{peak} values correspond to higher frequency spectral features (more readily distorted by smoothing filters).

3.4.3 Spectral Bandwidth

Smoothing filters operate by removing the high-frequency components of a noisy signal, and leaving the low-frequency components (usually characteristic of the information of interest) relatively untouched. Unfortunately this separation of high and low frequencies is difficult to achieve without some distortion of the lower-frequency components. Smoothing filters unfailingly corrupt the signals of interest and, as is implied from the derivation in **Section 3.2.2**, reduce the angles between the pure-component spectra, as well as their effective magnitudes (hence the shorter NAS_F vector). Although some distortion is inevitable, the effect is minimal when the filter window is small compared to the signal bandwidth since less low-frequency attenuation occurs [15, 19, 25, 30]. In all of the simulations conducted here in which the filter size was altered, very little degradation in the performance ratios were observed at small filter sizes,

suggesting that, when the filter size was small relative to the signal bandwidth, little distortion occurs. To more thoroughly examine the effect of this signal distortion on the filtered multivariate calibration procedure, systematic alterations in the width (*i.e.*, the sampling rate) of the gaussian peaks in the pure-component spectra were made while keeping the multivariate signal-to-noise (as defined in **Equation 3.30**) constant. The spectral vectors were composed of 500 channel spectra to alleviate any difficulties arising from the largest peak widths in the simulation. **Figure 3.16** gives an example of the results from for this experiment, with all other calibration characteristics at the standard conditions. The reader may note that the observed performance ratios do not exceed one in this case, and are generally observed to fall slightly below the theoretical values as filtering increases. While this is somewhat atypical, it is not unexpected, as the example given here is a single set of calibration and prediction data. Replication would undoubtedly demonstrate similar behavior to that exhibited in **Figures 3.9** and **3.10**. The wider peaks appear less susceptible to serious signal distortion which can lead to larger prediction errors, although any enhancement is consistently negligible as the peak width is increased. These observations are in accordance with generally accepted rules of thumb for filtering which suggest that the filter width in channels should not exceed half of the peak width of the features in the spectra. Indeed, if this suggestion is followed, impairment of the calibration due to signal distortion is apt to be minimal, but there is also little apparent advantage to applying the filter.

3.4.4 Correlated Errors

The introduction of correlated error into the filtered spectra is a difficult effect to examine using conventional PCR due to its assumptions of uncorrelated error. However, MLPCR [10], is capable of accounting for row-correlated measurement errors [11]. Since calculation of the filtered error covariance matrix is made possible using **Equation 3.6**, the covariance information known to be corrupting the calibration data can be utilized when constructing the calibration model and when making predictions from that model. This affords a reference

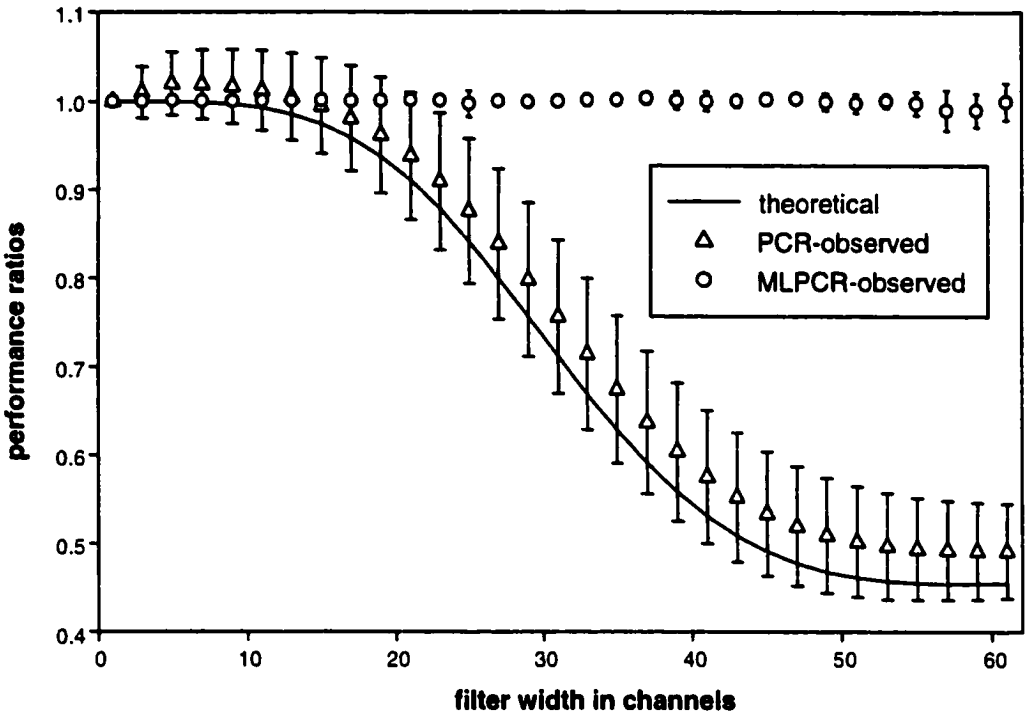


Figure 3.17 MLPCR observed performance ratio compared to the PR_{obs} for PCR with increasing smoothing filter width. The depicted values are averages of 20 replicate trials (error bars indicate ± 1s). Also shown are the theoretical performance ratios.

method which should account for the presence of correlated measurement errors. In **Figure 3.17** the resulting performance ratios for the theoretical, PCR and MLPCR calibration models are displayed under standard conditions. Note that virtually no change is observed in the MLPCR ratios with increasing filter width, while the PCR and theoretical performance ratios fall quickly below unity. The fundamental difference between the behavior of MLPCR and PCR under these conditions is not the orientations of the estimated subspaces, which are very nearly identical with row-correlated measurement errors, but the estimates of the scores for the sample spectra on the subspace. As outlined in **Section 1.4.1**, MLPCR establishes much the same spectral subspace estimate, but the projections of the calibration spectra onto that space are quite different from the PCR projections since the latter uses a simple orthogonal projection, while the

former uses the maximum likelihood projection. This effect is additionally present in the projections of the prediction spectra as well.

In some respects, MLPCR can be regarded as an optimal method because it takes the error structure of the measurements into account. Indeed, regardless of the degree of filtering, the quality of the MLPCR results does not appreciably diminish. Philosophically, this is satisfying when we consider that filtering does not really remove noise, but only changes its structure, substituting error covariance for error variance. Therefore, a technique that adapts to the change in the noise structure should produce consistent results. On the other hand, it will be noted that traditional PCR, which does not take error covariance into account, performs better than MLPCR in some cases, notably at very narrow filter widths and when the calibration space is very poorly estimated. PCA is biased in its estimation of the spectral subspace and sample scores, whereas MLPCA is not. The bias-variance trade-off results in the former giving a smaller prediction variance at the expense of accuracy in the limit.

3.4.5 Experimental Data

The observations made for the simulation studies were further investigated using the experimental data set described in **Section 3.2**. The observed performance ratio for each analyte (with PCR as the calibration method) was determined as the amount of smoothing was systematically altered (**Figure 3.18a**). As is apparent from the figures, the performance ratios for Ni and Co behave in a manner predicted by theory as the width of the moving-average filter is increased, *i.e.*, no improvement in the prediction error is observed. The PR_{obs} for Cr can be seen to slightly exceed unity at almost all filter sizes, although the magnitude of this ratio indicates only nominal improvements over the unfiltered case. **Figure 3.18b** shows the actual *RMSEP*'s for the analytes.

Although these data did not entirely satisfy the initial assumptions of the theoretical development in that some degree of error covariance was present in the raw data, the results still buttress the principal conclusion of the simulation

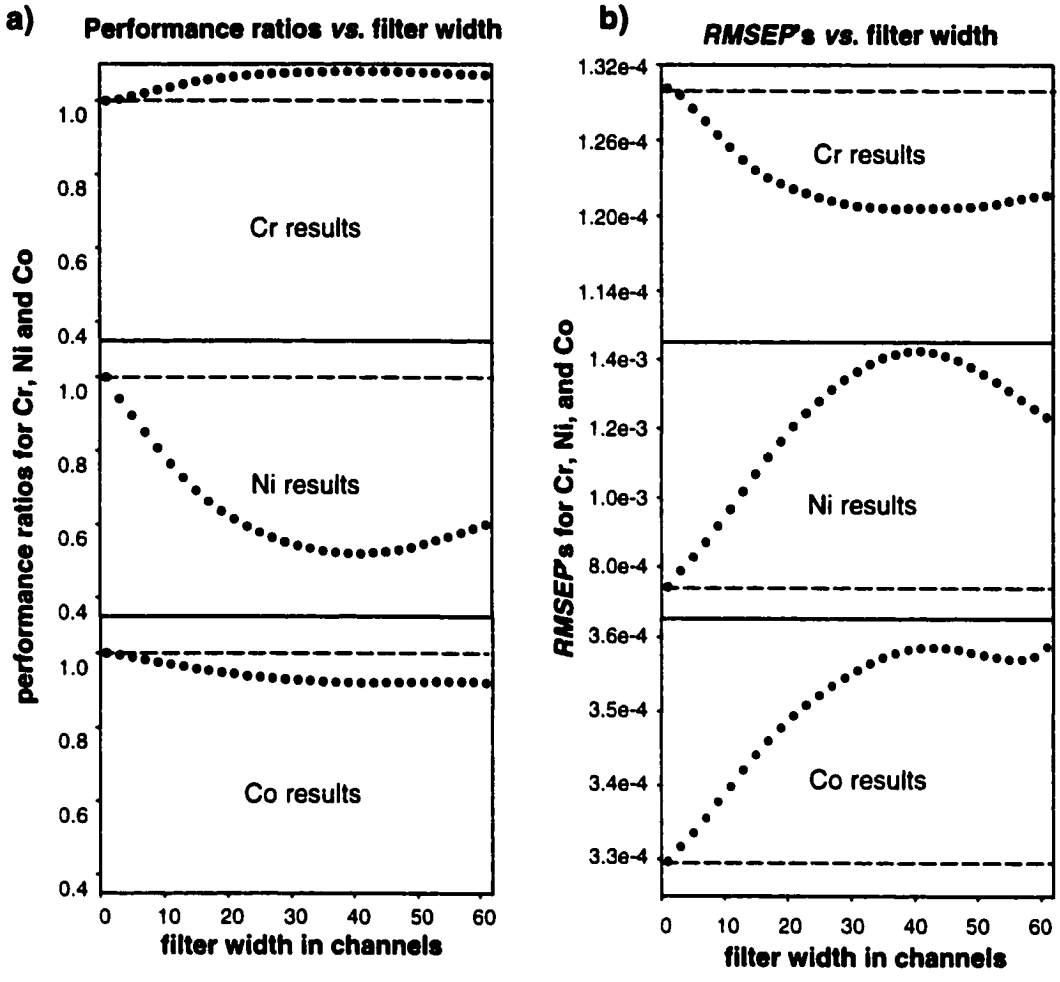


Figure 3.18 **a)** Performance ratios observed after applying smoothing filters to the experimental data set discussed in **Section 3.2**. The dashed line indicates the PR_{obs} = 1 (no change) mark. **b)** Actual *RMSEP*'s for each of the three analytes of interest as a function of applied smoothing filter width.

studies – that little or no improvement in performance can be anticipated by preprocessing with a symmetric smoothing filter.

3.4.6 A Dissection of Prediction Error

The most complicating issue in examining the net result of digital smoothing filters in multivariate calibration is that one cannot easily ascertain how each of the individual effects contributes to the observed prediction error, or

performance ratio. Using carefully planned simulation studies with pseudo-filtering operations, however, it is possible to isolate each effect.

The ideal smoothing filter, if it were realizable, would reduce the level of the noise while leaving the chemical signals and white noise structure undistorted. The reduction in variance as given in **Equation 3.1**, would be achieved without the injurious introduction of error covariance. The behavior of an ideal smoothing filter can therefore be mimicked by simply reducing the noise level of the data without introducing correlation, and without distorting the spectra. To examine the behavior of this optimal smoothing filter, then, reference calibration and prediction data sets were constructed at standard conditions (the 'unfiltered calibration/prediction data'). The application of ideal smoothers was feigned by calculating the theoretical reduction in noise achieved by a given smoothing filter via **Equation 3.1**, and subsequently re-generating the exact same calibration and prediction data using the reduced noise level. **Figure 3.19** shows the ideal filter behavior resulting from these conditions (solid squares). As is to be anticipated, the reduction in the noise level brings about a steadily increasing improvement in the calibration performance, as the noise is reduced to lower and lower levels. For this ideal filter the calibration model performance is always enhanced by smoothing, and hence, the performance ratio always exceeds one. As noted earlier in the chapter, noise variance reduction (when considered in isolation) will always result in an improvement in predictive power.

Simulating the isolated consequence of spectral distortion introduced by the smoothing operation involves smoothing the noise-free spectral data with a smoothing filter. If the properties of the measurement errors are unaltered (the variance is not reduced, and no error covariance is introduced), then the resulting performance ratios should be an indicator of the isolated contribution of spectral distortion in digital smoothing. The performance ratios under these conditions, also shown in **Figure 3.19** (unfilled triangles), are always below one. Substantial deviations from PR=1 begin to occur when the filter size surpasses the half-width of the spectral peaks, an effect that was observed indirectly in **Section 3.4.3**. The

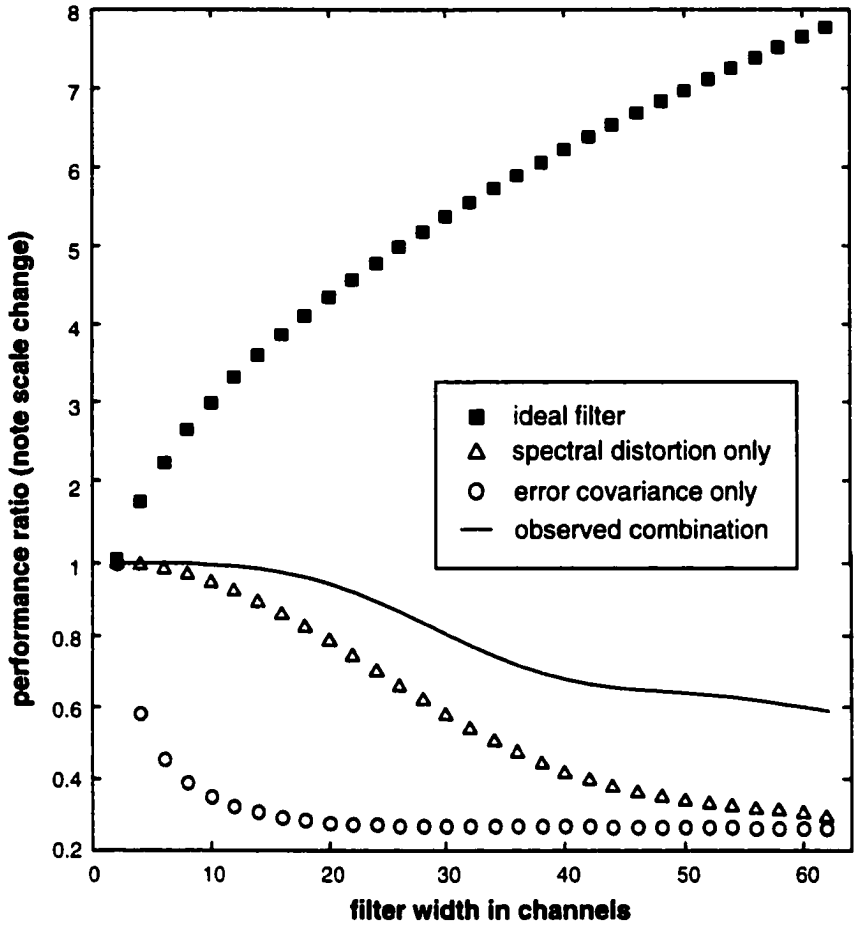


Figure 3.19 Results of the simulation studies examining the effect of each of the three effects of filtering in isolation. The performance ratios are displayed for an "ideal filter" (one which achieves noise variance reduction only), a filter which introduces error covariance effects, and a filter which introduces signal distortion. The observed performance ratio is also shown for the application of a real smoothing filter (all effects present).

spectral distortion from digital smoothing always decreases the *SEN* relative to the unfiltered case, and since the properties of the measurement errors are fixed in this scenario, we can anticipate that the reduction in *SEN* will result in a degradation in the *RMSEP*. The signal distortion will therefore always detract from the positive aspects of noise variance reduction, however the effect is evidently minimal at narrow filter widths. This distortion becomes much more significant when the filter width exceeds the half-width of the signals since the

smoothing filter bandpass limit begins to encroach on the low-frequency chemical information.

The effect of correlated errors was introduced by filtering the noise prior to applying it to the calibration and prediction data (which were not treated by the filter). The filtered noise was subsequently scaled to maintain the same noise variance at all filter sizes. Therefore, the change in the performance of the calibration model with filtering should be the sole result of the introduction of correlated measurement errors. The performance ratios in this case are observed to rapidly worsen even with very narrow filter widths (**Figure 3.19**, unfilled circles), and a steady decline in the performance of the calibration is observed as the filter width is continually increased. This is somewhat surprising, especially for small filters that introduce very limited correlation, since the only difference between this case and the unfiltered calibration is the dependence of the noise at a few neighboring channels. The observed performance ratio for all of the combined effects (*i.e.*, standard filter application) is also shown in **Figure 3.19**. The correlation of measurement errors is evidently an extremely important factor at small filter widths since spectral distortion has little effect at that point. The error covariance effect alone accounts for the marked departure from ideal filter behavior that is observed in practice.

When the effects of signal distortion and correlated noise are isolated and applied individually, a very clear picture of observed performance ratio develops. The reduction in error variance will always contribute positively to calibration performance, working to improve the *RMSEP* of the filtered calibration system. The introduction of error covariance, on the other hand, will always degrade the performance of the model relative to the unfiltered case. This effect is most influential at small filter widths, when spectral distortion is at a minimum, and is typically harmful enough to offset any gains made by noise variance reduction. At larger filter widths, the error covariance effects persist, but the relative contribution of the spectral distortion increases rapidly, pulling the performance

ratio well below the 'no-change' mark of unity. In short, it is two against one, and the error covariance and spectral distortion are most often the winners.

3.5 Conclusions

The objective of the work in this chapter has been to examine the utility of symmetric smoothing filters as a preprocessing method for multivariate calibration, and the implications of applying these filters to multivariate data. In terms of calibration performance, it has been demonstrated theoretically that, when prediction errors alone are considered, the application of a symmetric smoothing filter will, at best, yield no improvement in predictive ability and will actually degrade the quality of prediction as the filter size is increased. This net result is a composite of the effects of noise reduction, signal distortion, and error correlation. The assumptions in the theoretical development were: 1) the measurement noise in the raw calibration data is *iid*-normal, 2) calibration errors are negligible (*i.e.*, the NAS is exactly known, or the model space is very accurately estimated) and, 3) the system is linear with a well-defined rank. In cases where these assumptions do not hold, the development of a general model for the performance ratio would be difficult to obtain, but is expected that the main conclusion will remain valid as long as perturbations from the assumptions are not extensive. Indeed, simulations demonstrated excellent agreement between theoretical and observed performance ratios at modest noise levels, and filtering of an experimental calibration-prediction set showed the expected degradation in predictive ability. At small filter widths, this degradation is principally due to the introduction of error covariance. When the filter width becomes large enough to cause substantial distortions in the spectral information, the degradation becomes more extensive. Therefore, although smoothing is cosmetically pleasing to the analyst, the signal distortion and correlation of noise which result from smoothing filters have been demonstrated to often corrupt multivariate calibration data.

In certain cases for both simulated and experimental data, marginal gains in predictive ability were observed to occur with moderate filtering. In all cases examined here, the improvement in the *RMSEP* was less than a factor of two and may not justify application of the filter, especially given the degradation in performance that can potentially occur. The increase in performance ratios was traced to improvements in the calibration step, particularly with regard to the estimation of the spectral subspace. This study was necessarily limited in the number of spectral shapes simulated and so it cannot be stated absolutely that other circumstances will not yield larger gains after filtering, but, based on the large number of unreported simulations carried out during the course of this investigation, it is believed that marginal gains are the rule rather than the exception.

The scenarios in which smoothing has the greatest chances of being beneficial tend to be systems in which the estimation of the calibration model is likely to be poor due to relatively large uncertainty in estimating the spectral subspace. These situations are typically characterized by high measurement noise or large spectral correlation. The subspace estimate is bound to be highly uncertain when high noise levels corrupt the calibration data, or the pure-component spectral correlations are very high. In these circumstances, the reduced uncertainty in the subspace estimate attained via noise reduction by filtering sometimes offsets losses due to spectral distortion and error covariance introduction. Systems composed of almost uncorrelated pure-component spectra with low levels of measurement noise generally show no enhancement and only degradation after filtering. Signal bandwidth is also a factor which affects the quality of the calibration after filtering. It is recommended that in order to avoid serious signal corruption, the filter width should be keep below the bandwidth of the most important signals of interest.

It was also demonstrated that MLPCR, which accounts for correlated errors, can be applied to calibration data after filtering because the covariance information is readily available from the filter matrix. This technique essentially

deconvolves the filter from the filtered data, and because of this, no changes in the *RMSEP* are observed with increased filtering.

Smoothing with a symmetric filter matrix is often used 'black-box style' as a preprocessing tool for multivariate calibration. It is hoped that the theoretical results, the simulated experimental observations, and the demonstrated behavior of real experimental data will prompt more caution in multivariate calibration with respect to smoothing, and lend insight into the specific conditions under which filtering may prove somewhat beneficial. Although there are cases where filtering can enhance calibration performance, the gains in observed in *RMSEP*'s were consistently nominal. When weighed against the potential for reduction in predictive performance from symmetric smoothing, it is the recommendation of this work that smoothing filters be avoided as a preprocessing tool in multivariate calibration.

4. Drift Correction in Multivariate Calibration

4.1 Introduction

Digital smoothing filters, like the ones discussed in the previous chapter, are widely used in both qualitative and quantitative chemical analyses to reduce the variance of the measurement errors. As noted, this is achieved at the unfortunate expense of introducing spectral distortion and error covariance. But preprocessing extends well beyond the realm of variance reduction, and arguably is of greater importance today in handling other artifacts in the experimental data such as offsets, drift and scatter.

Baseline drift, or drift noise, is inherently imbedded in the information generated by most types of analytical instrumentation. The source of these drift effects are highly dependent on the nature of the experiment, the type of instrument used, and the types of transformations and processing used on the raw data. The magnitude of the drift also varies considerably between different experiments and instrumental methods. Instrumental contributions to drift include source intensity instability (flicker), detector response variations, temperature induced changes in critical instrumental components, and spatial correlations in the detection sensors. Non-instrumental sources of drift noise are also ubiquitous, and can be tied to everything from physical and chemical properties of the samples, to ambient temperature and pressure. Post-experiment transformations of the data can also introduce substantial levels of drift noise. **Chapter 3** discussed the aspects of drift noise introduced by symmetric smoothing filters in some detail. Other transformations of the data performed either within the instrument (*e.g.*, apodization functions), or outside of the instrument (*e.g.*, other digital filters, wavelet transforms) can also introduce drift noise that was not inherently present in the original measurement sequence.

While drift noise arises from manifold origins, it will be broadly characterized in this work as any undesirable fluctuation in instrument response which results in correlated measurement errors. Under this definition, offset noise, and multiplicative scattering effects (which are both non-*iid* due to error correlations) are simply special cases of drift noise. As discussed in **Section 1.3.1**, correlated measurement errors manifest themselves in the frequency domain as 'coloured' noise, typically exhibiting low-frequency dominance in the noise power spectrum. If the magnitude of baseline offsets is substantial enough to be observed in the time domain, it is most likely very easily observed as a dominant dc (direct current) contribution in the NPS. It can also be shown that offset noise results in an non-*iid* error covariance matrix, since offset noise is typically thought of as arising from the model

$$\mathbf{x} = \mathbf{x}^{\circ} + \mathbf{e} \tag{4.1}$$

where the measurement errors can be further broken down as

$$\mathbf{e} = a \cdot \mathbf{1} + \boldsymbol{\varepsilon} \tag{4.2}$$

In **Equation 4.2**, a is a scalar offset term (assumed to vary normally about a zero mean with a standard deviation σ_a), $\mathbf{1}$ is a one-vector, and $\boldsymbol{\varepsilon}$ is the iid component of the typical measurement errors. Since the error covariance matrix for the measurement errors in $\mathbf{e}$ is given by

$$\boldsymbol{\Sigma} = E\left(\mathbf{e}\mathbf{e}^{\mathrm{T}}\right) \tag{4.3}$$

upon expanding the product from **Equation 4.3** we find that the error covariance matrix for a signal corrupted by noise having both offset and *iid* components is

$$\boldsymbol{\Sigma} = \sigma_a^2 \cdot \mathbf{1}\mathbf{1}^{\mathrm{T}} + \boldsymbol{\Sigma}_{\varepsilon} \tag{4.4}$$

The error covariance matrix arising from offset corrupted data, then, is simply a matrix of offsets (σ_a^2) in addition to the *iid* contribution, $\boldsymbol{\Sigma}_{\varepsilon}$, making the observed error covariance structure of the data non-*iid*. Optical effects, such as sample scatter and pathlength effects can also be shown to result in correlated measurement errors by similar treatment.

As outlined in **Chapter 1**, most multivariate calibration methods in common use in chemistry assume the independence of noise at different spectral channels (*i.e.*, the noise is uncorrelated). Drift noise can certainly be a detriment to these methods when the *observed* properties of the data represent a significant departure from the model's *assumed* properties of the data. Multivariate models built from data with significant levels of drift often require a greater degree of parameterization (more factors in PCR; more variables in multiple linear regression (MLR)) to satisfactorily model the property of interest than models built from drift-free data. Several researchers have probed the issue of parsimony with respect to heteroscedastic and drift-corrupted data [31, 32], and commented on the characteristics and disadvantages of these models [33, 34]. With drift noise having such a negative impact on the success of a calibration model, many analysts attempt to minimize its contribution to the observed data variance by spectral preprocessing.

In recent years the number of preprocessing methods available to combat drift noise has proliferated. This is in part attributable to the now widespread use of instrumental methods which are slightly more susceptible to drift, such as Fourier transform spectrometers, and also due to the increasingly complex physical and chemical properties of sample matrices. Paradoxically, general improvements in analytical instrumentation have also contributed to this increased concern with drift, since white detector noise in newer instruments is often at a sufficiently low level to make drift noise arising from non-fundamental sources analysis-limiting. It has even been remarked that current spectroscopic instrumentation is inherently far more precise than most laboratory glassware [35].

A variety of methods are now at the disposal of the analyst to reduce the contributions of drift noise to the imprecision of analytical methods. The majority of these preprocessing techniques are empirical normalizations in a variety of guises which are used to reduce the contributions of offset and scatter noise to spectral data. These include wavelength ratioing [36] and differencing methods

in which the spectral measurements at every channel are normalized against some chosen reference wavelength. Numerous other standardizations are also available, such as multiplicative scatter correction (MSC) [37] (and associated variants [38]), Murray and Hall's multiplicative correction [39], standard normal variate (SNV) transformations [40], detrending, and simple background fitting, to name but a few. The performance of MSC will be discussed briefly in **Chapter 5**.

The selection of the best drift correction method or combination of methods appears, therefore, a rather formidable task. Additional complications associated with many of the above recommended methods are that only minor differences exist between them, and many are overly empirical – why should the mean spectrum represent the 'shape' of scatter? As a result, many of these methods have met with only lukewarm appeal in routine multivariate analysis. Another method, signal differentiation, predominates as the method of choice, and will therefore be the focus of this chapter.

Derivative preprocessing is among the oldest and most frequently used baseline correction techniques. The mass appeal of derivatives possibly stems from a simple historical familiarity with the technique, but it is also undoubtedly attributable to the conceptual simplicity of the method. In principle, first-derivative spectra should be free of baseline offset effects, since the first-derivative of any function will eliminate constant factors. Second- (and higher) derivative spectra should reduce baseline effects which can be modeled as polynomial functions of the ordinal variable (*e.g.*, variations proportional to the ordinal variable will be eliminated with second derivatives). Therefore, if a set of calibration spectra are collected which are corrupted with offset noise of the sort described in **Equations 4.2** and **4.4**, the first-derivative spectra will be free of offset contributions.

While derivative preprocessing for multivariate analysis is most certainly a routine practice, the role of derivative preprocessing from the calibration perspective has, surprisingly, yet to be explored with any degree of rigor. The earliest implementations of spectral differentiation were for the purposes of

feature extraction rather than drift-noise reduction in calibration [41]. Since these early qualitative applications, several researchers have undertaken studies of the effect of differentiation on peak positions, intensities, and areas [42], as well as attempted to ascertain the optimal parameters for derivative calculation when resolution enhancement was the goal [43]. Others have explored the calibration effects of differentiation in specific circumstances based on such things as the multivariate sensitivity, selectivity [32], and other figures of merit [44]; however, these investigations have almost entirely been focused on post-calibration observation of derivative performance. The actual role of derivative preprocessing, and drift correction in general, has not been explored from a theoretical calibration perspective.

In this chapter, the properties of derivative filters and their effects on chemical signals will be discussed, followed by an examination of the theoretical mechanism by which differentiation alleviates drift-noise. From this perspective, it will become clear that, in addition to other noted drawbacks, derivative filters are suboptimal in correcting for baseline drift. Based on theoretical considerations of the structure of the measurement errors, an optimal filter is derived for the correction of baseline drift in spectral calibration and prediction data. This optimal drift correction filter can be determined from the structure of the noise in a straightforward manner, and this approach to eliminating correlated measurement errors is subsequently shown to be a special case of maximum likelihood PCA [9]. MLPCA will be contrasted to derivative preprocessing, and the efficiency of maximum likelihood PCR [10] in correcting for baseline drift will be explored. The majority of the following has been published by Brown and Wentzell [45], and elements of this chapter which have been garnered from other sources are duly referenced.

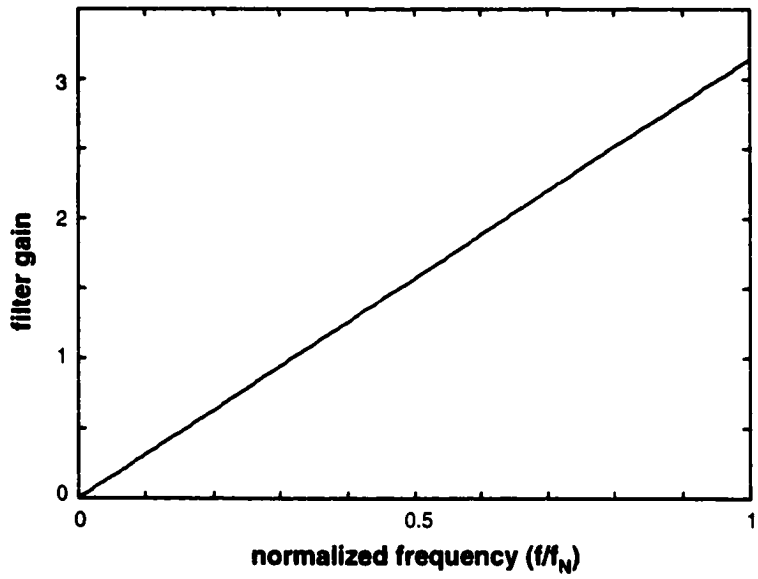


Figure 4.1 Transfer function for a 'true' derivative filter.

4.2 Theory

4.2.1 Derivative Filters

The approximation of signal derivatives can be achieved by a variety of means. Analytic derivatives could, in theory, be obtained if the true signal vector were available to the analyst in the form of an differentiable function, however this is rarely the case in practice. Historically, signal derivatives were approximated using analogue or digital hardware, but in today's laboratory differentiation can be readily achieved in software, which affords greater flexibility. When thought of as a filter, true signal differentiation corresponds to a filter with gain directly proportional to frequency, as shown in **Figure 4.1** [19]. This can be easily derived by realizing that any signal in the time domain can be represented as the sum of a series of sines and cosines, and analytic differentiation of such a signal yields

$$\frac{d(\sin(\omega t))}{dt} = \omega\cos(\omega t) \tag{4.5}$$

where $\omega(=2\pi f)$ is the angular frequency. Therefore, true differentiation, in approximating the rate-of-change of the signal vector, selectively amplifies the higher frequency components of an observed signal. Since chemical signals seldom reside at even mid- to high-frequency ranges, this true derivative largely acts to only amplify high-frequency noise, making calculation of the true derivative for practical measurements useless.

In chemical applications, differentiation of spectral signals is most often accomplished using finite differencing or polynomial least-squares filters, the simpler method of the two being differencing. In this method the rate of change of the signal vector is approximated by finding the difference between the signals at adjacent channels in the spectral vector. Higher derivatives can be obtained by reapplying the differencing to the first-derivative spectrum. This simple filter, although a reasonably accurate representation of the observed signal derivative, is typically undesirable due to its extreme sensitivity to high-frequency noise. The transfer function for a difference filter is shown in **Figure 4.2a**, and an example of its application to a spectrum is shown in **Figure 4.3**. Examining the transfer function, we can see that there is substantial attenuation of low-frequency signals, and that the difference filter actually amplifies mid-, and high-frequency regions. As a result of the objectionable response of these filters at higher frequencies, differencing methods are of limited utility unless the spectral data exhibit very classical high signal-to-noise ratios (as is often said to be the case with NIR measurements). Consequently, differencing is often used in conjunction with smoothing procedures to achieve some low-pass filtering. These two operations can, however, be achieved simultaneously with polynomial least-squares filters [22, 46].

Polynomial least-squares filters yield a least-squares estimate of the derivative over a window of points in the spectrum (see **Section 2.2.1**). The least-squares properties of these functions achieve a degree of low-pass filtering, while the derivative properties provide some relief from the low-frequency drift effects; the result is a form of band-pass filter. Transfer functions for 3- and 13-

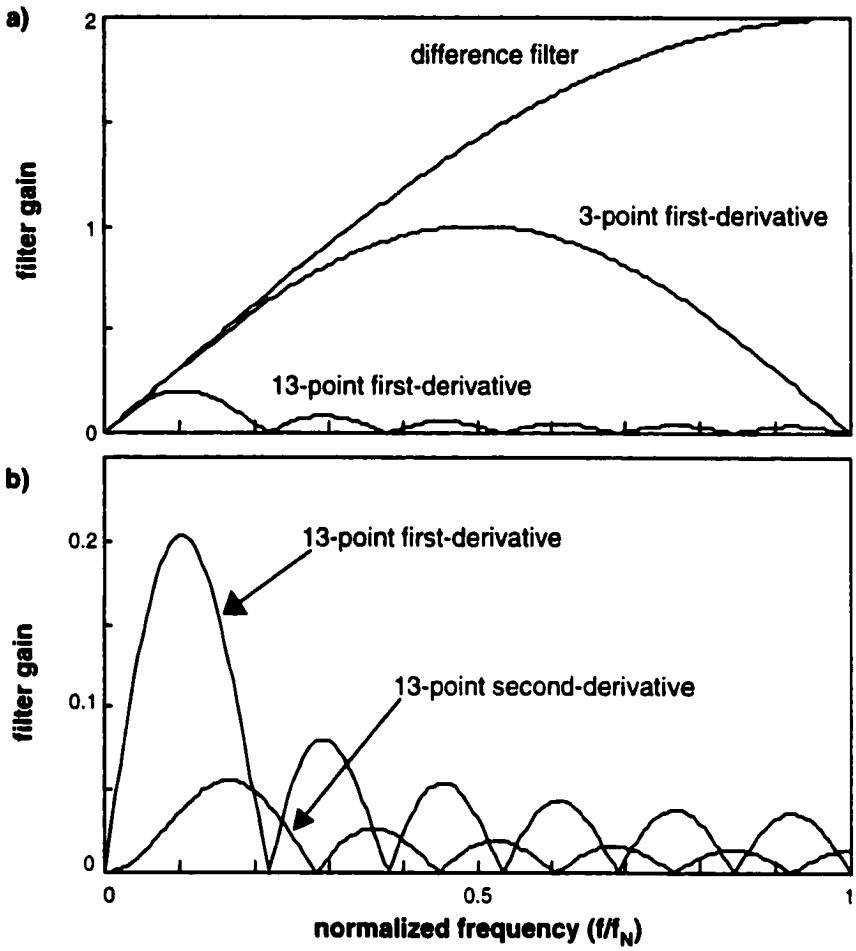


Figure 4.2 **a)** Transfer functions for a variety of realistic derivative filters including a difference filter, as well as 3- and 13-point Savitzky-Golay derivative filters (all first-order polynomials). In **b)** the transfer functions for 13-point linear first- and quadratic second-derivative filters are compared, showing the change in bandpass associated with higher derivatives.

point linear first-derivative least-squares filters are shown in **Figure 4.2a** as examples. In **Figure 4.2b**, 13-point linear first-, and quadratic second-derivative transfer functions are compared, showing the greater low-frequency attenuation achieved by the higher derivative. **Figure 4.3** is also included as a depiction of the characteristics of derivative spectra obtained by the application of these filters. In general, increasing the order of the derivative will result in a greater degree of low-frequency attenuation ('smoothness') since the low-frequency

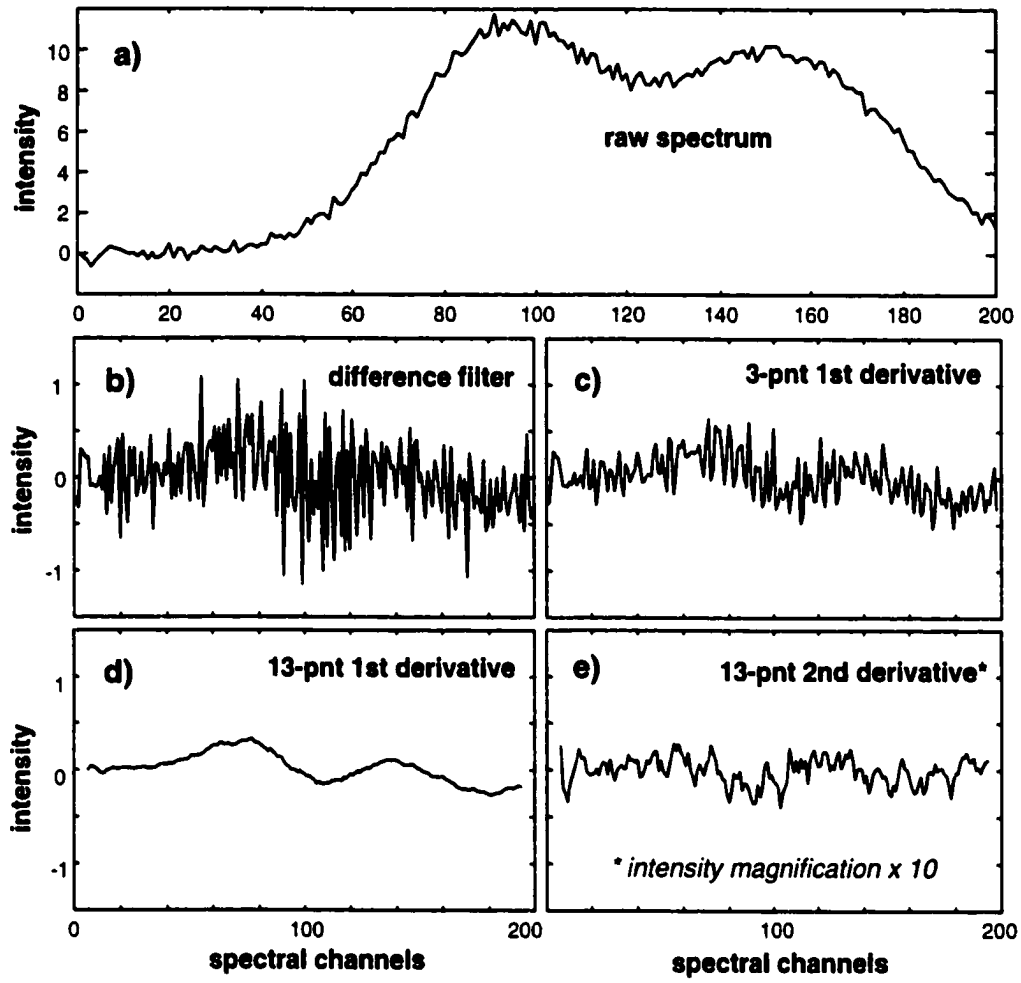


Figure 4.3 **a)** A simulated spectrum corrupted by drift noise, and the resulting derivative spectra from applying **b)** difference, **c)** 3-point linear first-derivative, **d)** 13-point linear first-derivative, and **e)** 13-point quadratic second-derivative filters (magnified by a factor of 10 for clarity) to the original spectrum.

cutoff for higher derivatives is essentially moved to higher frequencies. Wider filters will achieve greater reductions in the higher frequencies due to greater suppression of high-frequency components and a shift in the high-frequency cutoff to lower regions. The order of the polynomial function fitted to the data window is inversely related to the extent of high-frequency attenuation (*i.e.*, higher-order fits retain more high-frequency components).

From **Section 3.2.1** the reader will recall that the coefficients of a digital filter can be used to determine the relative variance of the measurement errors in

the filtered and unfiltered signal. (**Equation 3.1** has been reproduced here for convenience.)

$$\frac{\sigma^2_{filtered}}{\sigma^2_{unfiltered}} = \sum_i c_i^2 \tag{4.6}$$

Under the requisite noise conditions – measurement errors corrupting the signal vector are *iid* – difference filters will increase the noise variance by a factor of 2. In contrast, most Savitzky-Golay derivative filters actually *reduce* the level of noise due to their band-pass properties. This is an important characteristic to note, since it is often said that the differentiation of spectra *increases* the noise. While statements such as these may convey the correct connotation (*i.e.*, the univariate signal-to-noise ratio is reduced by differentiation), the statement is erroneous as written. The band-pass nature of Savitzky-Golay derivative filters and their transfer functions reveal that in the majority of scenarios the noise variance is substantially reduced; however, the low-frequency attenuation of these filters also substantially reduces the slowly varying chemical signals of interest. These combined effects very often result in a reduced univariate *S/N* as defined in **Equation 1.57**. It should also be apparent that symmetric Savitzky-Golay derivative filters (which arise from even derivatives) are subject to the same proof as given in **Chapter 3**, implying that the multivariate *S/N* can never be enhanced by derivative filters either (unless the exceptional conditions outlined in the **Section 3.4** are met). It is necessary to reiterate, however, that this proof, and **Equation 4.6** are only valid when the original signal is corrupted *exclusively* by homoscedastic white noise, which is clearly not the case in spectra exhibiting baseline drift.

The Savitzky-Golay implementation of derivative filtering (particularly second-derivatives) would appear to be the most popular in the chemical literature to date. The favorable repute most likely results from the desirable band-pass properties of these filters, as well as their simplicity and ease of use. For these reasons, the Savitzky-Golay method of differentiation was used in the discussions and experiments that follow.

4.2.2 Sensitivity and Selectivity Considerations

A source of considerable literature regarding the use of Savitzky-Golay filters is their rather unpredictable effect on the signal quality. In **Chapter 3**, this effect was noted to be problematic when SG smoothing is done prior to multivariate calibration [15], and similar perils exist with SG derivative filtering.

The use of derivative filters as preprocessing tools for multivariate calibration requires theoretical consideration of the calibration procedure itself. As discussed in previous chapters, it would be erroneous to use such factors as the univariate *S/N* ratio in examining the effect of derivative preprocessing, since this univariate measure is rarely a valid indicator of the predictive success achievable. It is therefore necessary to consider the effect of derivative preprocessing on multivariate figures of merit [13,14,17]; in particular, the sensitivities, selectivities, and *S/N* ratios.

The multivariate *SEN*, as discussed in **Section 1.5.2**, is a scalar measure describing the magnitude of the signal specifically attributable to the analyte of interest in the calibration system. Two factors contribute to this sensitivity metric: the magnitude of the values in the spectra themselves, and the similarity (or correlation) of the pure-component spectrum for the analyte of interest with all other interfering analyte spectra. The magnitude of spectral values can be changed simply by changing the scale of the *y*-axis or, in the case of derivative spectra, the ordinal variable. This property makes the *SEN* in the absence of context an arbitrary figure of merit, and thus one of little use in examining the effect of derivative filtering. The other principal factor contributing to the *SEN* for an analyte of interest is the correlation among the pure-component spectra, which is better characterized by the multivariate selectivity.

The multivariate *SEL* (also discussed in **Section 1.5.2**) contributes to the *SEN*, but is a unit-less measure of the extent to which interfering components obscure the signal of the analyte of interest. The *SEL* can take on a value between zero and one – zero corresponding to complete obfuscation of the analyte by interferences, and one corresponding to total selectivity for the analyte

(no interferences). While a selectivity of one rarely occurs in practice, **Equation 1.56** suggests that, for the *SEL* to increase upon differentiation, the correlation of the pure-component spectrum of the analyte, s_i, with the subspace defined by the interfering pure-component spectra must decrease. Given that the form that the derivative spectrum takes is specific to the frequency content and location of the features contained in the original spectrum, it is difficult to know *a priori* whether the derivative spectra will be more or less correlated than the original data.

It is often said in the literature that differentiation of spectra 'enhances the subtle differences in the spectra'. This would appear to be another case of semantic inaccuracy, as differentiating filters are traditionally far more responsive to large changes in the signal than small ones. What is typically implied is that differences between different spectra are enhanced, and it is assumed that this bears direct rewards in calibration. While this may be true in certain circumstances, differentiation by Savitzky-Golay methods operates by *suppressing* the low-frequency character of a spectrum (and typically the very high-frequency signals as well). This has the effect of not only suppressing drift noise, but also reducing the low-frequency character of the chemical responses in the spectrum. Therefore, if low-frequency content is largely responsible for overlap between the spectra of difference mixtures, it would be anticipated that differentiation would make them look 'more different', and *SEL* enhancement *may* result. If, however, the lower frequencies are heavily attenuated and they are important in the success of the calibration procedure, then derivative filtering will obviously be detrimental to the calibration procedure.

The previously discussed problem with the multivariate sensitivity is that, without context, its value is rather arbitrary. The multivariate signal-to-noise ratio, however, is the ratio of the multivariate signal attributable exclusively to the analyte of interest to the level of the noise corrupting the calibration data [13,14,17]. Brown and Wentzell's definition for multivariate *S/N*, as discussed in **Section 1.5.2**, allows the incorporation of heteroscedastic and correlated measurement errors [15] in the metric, and hence is particularly useful for

estimation of *S/N* in the presence of drift. As stated, this definition of *S/N*, reproduced here for convenience, depends on both the *SEN* (numerator) and a measure of the noise level (denominator) for the analyte of interest.

$$S/N_i = \frac{SEN_i}{\sqrt{\mathbf{v}_i^T \boldsymbol{\Sigma} \mathbf{v}_i}} \tag{4.7}$$

Since changes in this multivariate *S/N* are expected to be proportional to changes in calibration model performance, it is feasible that derivative filtering can alter the predictive ability of the calibration procedure by either enhancing the *SEN*, reducing the noise level, or improving the ratio of the two. It can be anticipated that derivative filter application will significantly reduce the *SEN* since this metric is almost always reduced by differentiation. (Derivative spectra are much lower in magnitude than the original spectra assuming a unitary change in the ordinal variable – the only exception is spectra consisting of delta signals.) The filter application may also suppress the noise to some extent (**Section 4.21**). It is, therefore, not immediately clear whether the overall *S/N* will benefit as a result of derivative filtering.

4.2.3 Derivative Filters and Baseline Drift

As discussed in the introduction, the low-frequency nature of drift noise is indicative of errors being correlated among channels. The greater the low-frequency components of the drift the more extensive the correlation, with offset noise being the extreme case (entirely dc in nature). How the derivative filter explicitly interacts with this noise is of particular interest, since derivative filters are typically applied with this task in mind.

The reader will recall from **Chapter 2** that a Savitzky-Golay filter of any sort can be applied to an *m* x *n* matrix of sample spectra via

$$\mathbf{X}_F = \mathbf{X}\mathbf{F}^T \tag{4.8}$$

where the rows of the matrices $\mathbf{X}$ and $\mathbf{X}_F$ contain the spectra (measured over *n* channels) for each of *m* different samples for the unfiltered, and filtered cases respectively. Given that the observed data matrix, $\mathbf{X}$, can be considered the

sum of the true data, $\mathbf{X}^{\circ}$, and a matrix of measurement errors, $\mathbf{E}$, **Equation 4.8** can be expressed as

$$\begin{aligned}\mathbf{X}_F &= (\mathbf{X}^{\circ} + \mathbf{E})\mathbf{F}^T \\ &= \mathbf{X}^{\circ}\mathbf{F}^T + \mathbf{E}\mathbf{F}^T\end{aligned} \tag{4.9}$$

Under normal circumstances (*iid* noise) it would be assumed that the elements of $\mathbf{E}$ are normally distributed, and that the error covariance matrix associated with any given row of $\mathbf{E}$ is diagonal ($\Sigma = E(\mathbf{e}\mathbf{e}^T) = \sigma^2\mathbf{I}_n$). With correlated errors, however, we are assured this is not the case, and $\Sigma = E(\mathbf{e}\mathbf{e}^T) \neq \sigma^2\mathbf{I}_n$ (Σ cannot be expressed as a multiple of the identity matrix). With the application of a filter matrix to the data, and thus to the individual error vectors, the error covariance matrix *after* filtering (Σ_F) can be expressed, as in **Equation 3.5**, as

$$\Sigma_F = \mathbf{F}\Sigma\mathbf{F}^T \tag{4.10}$$

Therefore, the efficiency of the applied derivative filter in eliminating baseline drift can be appraised by examining the structure of Σ_F, and considering how closely it approximates *iid* conditions. If the derivative filter is completely successful in removing error covariance (and thus drift noise), then Σ_F will be a diagonal matrix. If the filter also removes heteroscedasticity, then Σ_F will be a multiple of the identity matrix.

4.2.4 Optimal Corrections for Baseline Drift

In data that are cured of drift noise by derivative filtering, the filtered error covariance matrix, Σ_F, will be diagonal. If the filtered noise is also desired to be homoscedastic, then $\Sigma_F = \sigma^2\mathbf{I}_n$. Dropping the proportionality constant (which can be viewed simply as a scaling factor), and substituting this relation into **Equation 4.10** leads to

$$\begin{aligned}\mathbf{I}_n &= \mathbf{F}_o\Sigma\mathbf{F}_o^T \\ \mathbf{F}_o^{-1}(\mathbf{F}_o^T)^{-1} &= \Sigma\end{aligned} \tag{4.11}$$

$$\left(\mathbf{F}_o^{-1}(\mathbf{F}_o^T)^{-1}\right)^{-1} = \Sigma^{-1} \tag{4.12}$$

Rearranging, this equation leads to

$$\mathbf{F}_o^T \mathbf{F}_o = \boldsymbol{\Sigma}^{-1} \tag{4.13}$$

$\mathbf{F}_o$ indicates that the filter is now *optimally* suited to remove baseline drift. Therefore, for derivative filters to perform optimally in removing drift noise, the condition described in **Equation 4.13** must be met. Clearly, this could occur only in rare circumstances due to the necessary confines on the structure SG derivative filter matrices, and the fact that derivative preprocessing traditionally does not make use of error covariance information.

Equation 4.13 also suggests that the optimal removal of drift noise requires knowledge of the error covariance structure of the data. This knowledge of the error covariance structure of the data is, of course, also required by MLPCA and MLPCR to achieve maximum likelihood representations of the data in lower dimensioned spaces. The application of derivative filters is commonly achieved by applying the same filter to all rows in the data matrix, $\mathbf{X}$, which in essence assumes that the same filter is adequate for alleviating the drift effects in all samples. If the same assumptions are made in MLPCA and MLPCR, it can be shown that the optimal drift reduction filter when applied in conjunction with PCA is coincident with the maximum likelihood solution to the principal component space.

Returning to **Equation 4.13**, and recalling that the error covariance matrix, $\boldsymbol{\Sigma}$, is by definition symmetric, it is evident that a singular value decomposition of $\boldsymbol{\Sigma}^{-1}$ can be expressed as

$$\boldsymbol{\Sigma}^{-1} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{4.14}$$

If we let $\mathbf{S} = \boldsymbol{\Lambda} \cdot \boldsymbol{\Lambda}$, and make use of the equivalence of $\mathbf{U}$ and $\mathbf{V}$ in this case (due to the symmetry of $\boldsymbol{\Sigma}^{-1}$) then **Equation 4.15** can be expressed as

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &= \mathbf{U}\boldsymbol{\Lambda}\boldsymbol{\Lambda}\mathbf{U}^T \\ &= (\mathbf{U}\boldsymbol{\Lambda})(\mathbf{U}\boldsymbol{\Lambda})^T \end{aligned} \tag{4.15}$$

Substitution for $\boldsymbol{\Sigma}^{-1}$ from **Equation 4.13** into **Equation 4.15** yields

$$\mathbf{F}_o^T \mathbf{F}_o = (\mathbf{U}\boldsymbol{\Lambda})(\mathbf{U}\boldsymbol{\Lambda})^T \tag{4.16}$$

and thus,

$$\begin{aligned} \mathbf{F}_o &= (\mathbf{U}\Lambda)^T \\ \mathbf{F}_o^T &= \mathbf{U}\Lambda \end{aligned} \tag{4.17}$$

Therefore, the optimally designed filter matrix, $\mathbf{F}_o$, is easily determined provided the error covariance matrix is available. It should be noted that this calculated optimal filter matrix will not be of the typical form of the least-squares polynomial filter matrices (band diagonal and symmetric / antisymmetric), and cannot be implemented through a convolution operation with spectra. Although the matrix cannot be considered a filter in the traditional sense, the term "filter matrix" will still be used for convenience. With $\mathbf{F}_o$ determined from **Equation 4.17**, the optimal drift-noise filter can be applied to the spectral data in the standard fashion:

$$\begin{aligned} \mathbf{Z}_F &= \mathbf{X}\mathbf{F}_o^T \\ &= \mathbf{X}(\mathbf{U}\Lambda) \end{aligned} \tag{4.18}$$

where $\mathbf{Z}_F$ indicates that the spectral data has been operated on by the filter matrix. Conceptually, this filtering process can be thought of as removing correlations in the measurement errors present in the spectral data by rotating the spectral vectors into directions in which the error is uncorrelated, and stretching the vectors such that the measurement errors are homoscedastic. This concept is rendered in **Figure 4.4**. Once in this orientation, standard PCA (as described in **Section 1.2.4.1**) for a chosen rank, *p*, can be employed on $\mathbf{Z}_F$, resulting in $\tilde{\mathbf{Z}}_F$.

$$\mathbf{Z}_F \xrightarrow[PCA,p]{} \tilde{\mathbf{U}}_Z \tilde{\mathbf{S}}_Z \tilde{\mathbf{V}}_Z^T = \tilde{\mathbf{Z}}_F \tag{4.19}$$

The subscript '*Z*'s are meant to provide a distinction between the **U**, **S**, and **V** matrices that result from a singular value decomposition of **Z**, and those that result from such a decomposition of the error covariance matrix. With rank reduction achieved by PCA, the spectral vectors can be returned to their approximate original positions by the inverse operation

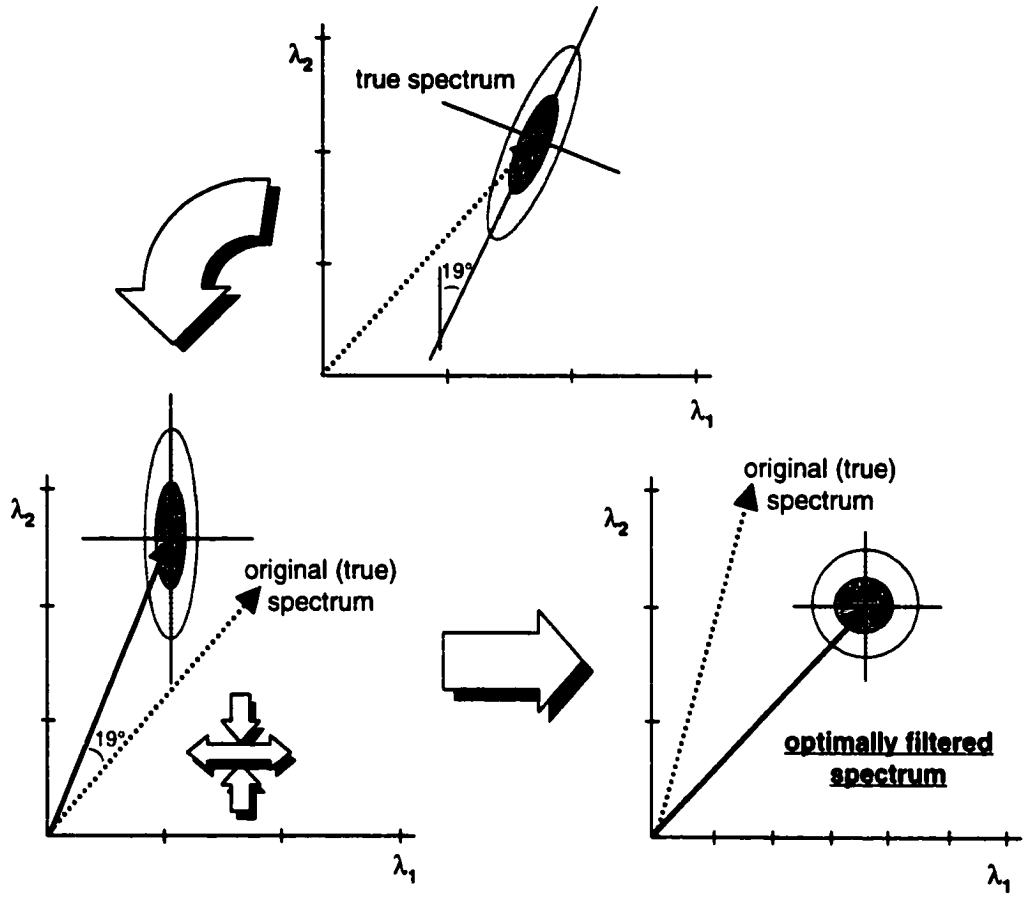


Figure 4.4 An illustration the concept of optimal filter application. **a)** original spectrum exhibiting heteroscedasticity and error covariance characteristics, **b)** rotation of the spectrum by some angle to eliminate error covariance, and **c)** scaling of the scores on each axis to eliminate heteroscedasticity.

$$\begin{aligned}\tilde{\mathbf{X}}_F &= \tilde{\mathbf{Z}}_F\left(\mathbf{F}_o^T\right)^{-1} \\ &= \tilde{\mathbf{Z}}_F\left(\Lambda^{-1}\mathbf{U}^T\right)\end{aligned} \tag{4.20}$$

where $\tilde{\mathbf{X}}$ is the rank p maximum likelihood solution to the PCA of the original data. Additional numerical concerns, namely the stable inversion of the error covariance matrix for **Equation 4.14**, can be addressed by obtaining the optimal filter matrix from the non-inverted error covariance matrix. The required adaptation is simply that

$$\left(\boldsymbol{\Sigma}^{-1}\right)^{-1} = \left(\mathbf{U}\Lambda^2\mathbf{U}^T\right)^{-1} \tag{4.21}$$

$$\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}^{-2}\mathbf{U}^{T} \tag{4.22}$$

and thus, the optimal filter matrix, when **U** and **S** are calculated from the *non-inverted* error covariance matrix, is given by

$$\mathbf{F}_o^{T} = \mathbf{U}\mathbf{\Lambda}^{-1} \tag{4.23}$$

This alteration in no way changes the rotation itself, since the rotation of a subspace, S_{Σ}, will produce an identical rotation of a subspace necessarily orthogonal to it, $S_{\Sigma^{-1}}$, and it can be easily shown by algebraic manipulation that

$$(\mathbf{U}\mathbf{\Lambda})\mathbf{\Sigma}^{-1}(\mathbf{U}\mathbf{\Lambda})^{T} = (\mathbf{U}\mathbf{\Lambda}^{-1})\mathbf{\Sigma}(\mathbf{U}\mathbf{\Lambda}^{-1})^{T} \tag{4.24}$$

Thus, for situations in which there is equal row error covariance, as is usually assumed when digital filters are prescribed, this simple method of performing MLPCA avoids the inversion of the error covariance matrix while achieving the optimal baseline drift correction, and rank *p* - PCA simultaneously.

While MLPCA can be considered a preprocessing step and PCA combined, it, like PCA, can also be used directly in multivariate calibration as MLPCR [10]. It is therefore proposed that MLPCA is an optimal drift-noise preprocessing method, and it's regression counterpart, MLPCR, is an optimal regression method for use in calibration and prediction systems corrupted by drift-noise. These methods are optimal in the statistical sense, in that they generate the *most likely* (maximum likelihood) principal component subspaces based on (1) the spectral data at hand, and (2) the knowledge the analyst has of the error structure of the data (via replication or otherwise). As discussed in **Chapter 1**, the projections of spectra onto the MLPCA subspace do not necessarily occur orthogonally – the obliqueness of the projection is determined by the magnitudes and directions of error variance-covariance corrupting the calibration space (see **Figure 1.15** in **Chapter 1**). Since the optimal filter matrix derived above contorts the spectral vectors into an orientation which can be considered *iid*, an orthogonal projection *can* be made in this 'filtered' domain. The procedure for the prediction step, then, can be given as

$$\mathbf{Z}_{pred,F} = \mathbf{X}_{pred}\mathbf{F}_o^{T} \tag{4.25}$$

Calibration

1. $\Sigma \xrightarrow{\text{svd}} \mathbf{USV}^{\mathrm{T}} = (\mathbf{U\Lambda})(\mathbf{U\Lambda})^{\mathrm{T}}$
2. let $\mathbf{F}_{\mathrm{o}}^{\mathrm{T}} = \mathbf{U\Lambda}$
3. $\mathbf{Z}_{cal,\mathrm{F}} = \mathbf{X}_{cal}\mathbf{F}_{\mathrm{o}}^{\mathrm{T}}$
4. $\mathbf{Z}_{cal,\mathrm{F}} \xrightarrow[PCA,p]{} \tilde{\mathbf{U}}_Z\tilde{\mathbf{S}}_Z\tilde{\mathbf{V}}_Z^{\mathrm{T}} = \tilde{\mathbf{Z}}_{cal,\mathrm{F}}$
5. $\tilde{\mathbf{X}}_{cal,\mathrm{F}} = \tilde{\mathbf{Z}}_{cal,\mathrm{F}}\left(\mathbf{F}_{\mathrm{o}}^{\mathrm{T}}\right)^{-1}$
6. $\hat{\mathbf{b}} = \left(\tilde{\mathbf{X}}_{cal,\mathrm{F}}\right)^{+}\mathbf{y}_{cal}$

Prediction

1. $\mathbf{Z}_{pred,\mathrm{F}} = \mathbf{X}_{pred}\mathbf{F}_{\mathrm{o}}^{\mathrm{T}}$
2. $\tilde{\mathbf{Z}}_{pred,\mathrm{F}} = \mathbf{Z}_{pred,\mathrm{F}}\tilde{\mathbf{V}}_Z\tilde{\mathbf{V}}_Z^{\mathrm{T}}$
3. $\tilde{\mathbf{X}}_{pred,\mathrm{F}} = \tilde{\mathbf{Z}}_{pred,\mathrm{F}}\left(\mathbf{F}_{\mathrm{o}}^{\mathrm{T}}\right)^{-1}$
4. $\hat{\mathbf{y}}_{pred} = \tilde{\mathbf{X}}_{pred,\mathrm{F}}\hat{\mathbf{b}}$

Figure 4.5 An algorithmic summary of the procedures required to perform MLPCR when derived as an optimal filtering method. This derivation implicitly assumes that the error covariance structure of the prediction data is not significantly different than the error covariance structure of the calibration data.

$$\tilde{\mathbf{Z}}_{pred,F} = \mathbf{Z}_{pred,F}\tilde{\mathbf{V}}_Z\tilde{\mathbf{V}}_Z^{\mathrm{T}} \tag{4.26}$$

which is the orthogonal projection of $\mathbf{Z}_{pred,F}$ onto the *p*-dimensional subspace defined by $\tilde{\mathbf{Z}}_F$, and hence, $\tilde{\mathbf{V}}_Z$ (obtained from **Equation 4.19**). The rotation and distortion of the optimal filter can now be undone by multiplication by the inverse

$$\tilde{\mathbf{X}}_{pred,F} = \tilde{\mathbf{Z}}_{pred,F}\left(\mathbf{F}_{\mathrm{o}}^{\mathrm{T}}\right)^{-1} \tag{4.27}$$

leaving the prediction spectra in the relevant chemical space, and thus

$$\hat{\mathbf{c}}_{pred} = \tilde{\mathbf{X}}_{pred,\mathrm{F}}\left(\tilde{\mathbf{X}}_{\mathrm{F}}\right)^{+}\mathbf{c}_{cal} \tag{4.28}$$

The calibration and prediction aspects of MLPCR when derived in this fashion are summarized for convenience in **Figure 4.5**. The reader should note that the previous theoretical treatment differs slightly from the account given in Brown and Wentzell's article on the subject [45]. In reference 45 the filter matrix was defined such that it could be applied to the spectral data as $\mathbf{X}_F = \mathbf{XF}$; therefore, these two theoretical treatments differ only in the sense that $\mathbf{F}^{\mathrm{T}}$ here is equivalent to $\mathbf{F}$ in the article.

4.3 Experimental

4.3.1 Simulations

To study the process of drift noise reduction, simulation studies were carried out using three calibration methods: PCR, derivative preprocessed spectra with PCR (derivative PCR), and MLPCR. The differentiation of spectra was achieved using the method of Savitzky and Golay [22] with a variety of widths and orders for the filter function. All of the simulated calibration systems in this work involved three spectrally active chemical components, whose pure-component spectral features were generated either according to 'controlled criteria' or 'randomly'. Regardless of the type of spectral vectors used, the pure component spectra were normalized to unit length to standardize the simulated instrumental responses. Calibration sets consisted of 20 mixture spectra in which the concentrations of each of the three components were drawn randomly from a uniform distribution between zero and unit concentration. The prediction sets consisted of 100 mixture spectra with concentrations also drawn from a uniform distribution between 0 and 1.

4.3.1.1 Controlled Spectral Data

In studies in which it was desirable to fix the correlation of the pure-component spectra (spectral angle), each pure-component spectrum was generated from a single gaussian peak (σ_{peak} = 10 spectral channels) placed in a 200 channel spectrum such that the correlation between spectrum 2 and all other interfering spectra was 45°. These simulated calibration and prediction sets were consequently of similar constitution as those used in **Chapter 3**. A set of noise-free calibration mixture spectra generated under these conditions is shown in **Fig. 4.6**, with the inset showing the pure-component spectra. The baseline regions on the ends of the spectra were useful to minimize edge-effects from the filtering process.

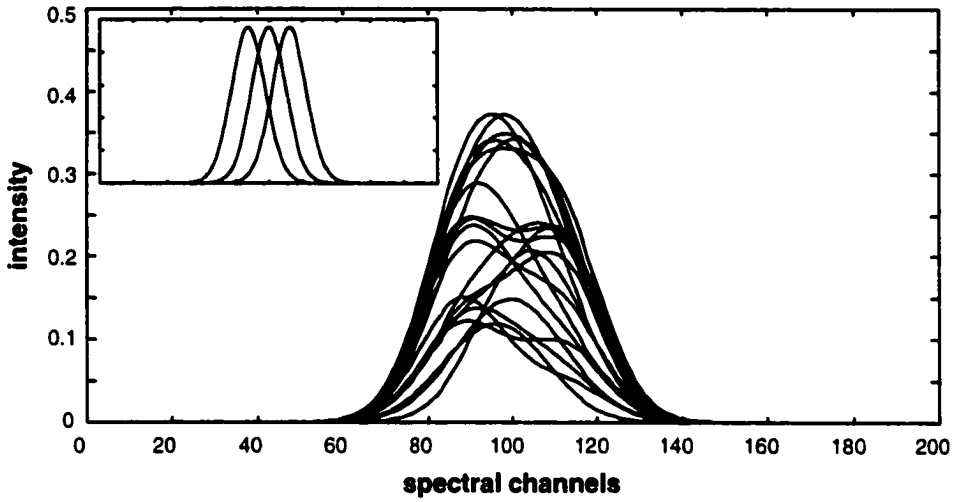


Figure 4.6 An example of 20 noise-free calibration spectra generated using controlled criteria for the spectral correlations and frequency composition. The inset shows the 3 pure-component spectra generating this observed set of mixtures.

4.3.1.2 Randomly Generated Spectral Data

In the examinations of the multivariate figures of merit for differentiated spectra, it was necessary to minimize the effect of the shape of the spectra on the studies. This was achieved, at the expense of lost control over the angle between the pure-component spectra, by generating each pure-component spectrum from 4 additive Gaussian bands whose locations in the spectrum were centered at randomly chosen channels. If very broad Gaussian bands (*e.g.*, σ_{peak} = 25 channels) are used in generating the spectra, then the spectra tend to be comprised of mostly very low frequency signals, be broad and generally featureless, and thus highly overlapped with other pure-component spectra. In contrast, spectra generated from narrow Gaussian bands (*e.g.*, σ_{peak} = 2 channels) tend to have higher frequency signals, and thus sharper spectral features. As a consequence, sets of pure-component spectra generated from narrow Gaussian bands tend to be less correlated than their broad-featured counterparts. Two pure-component spectra of this sort are shown in **Figure 4.7**.

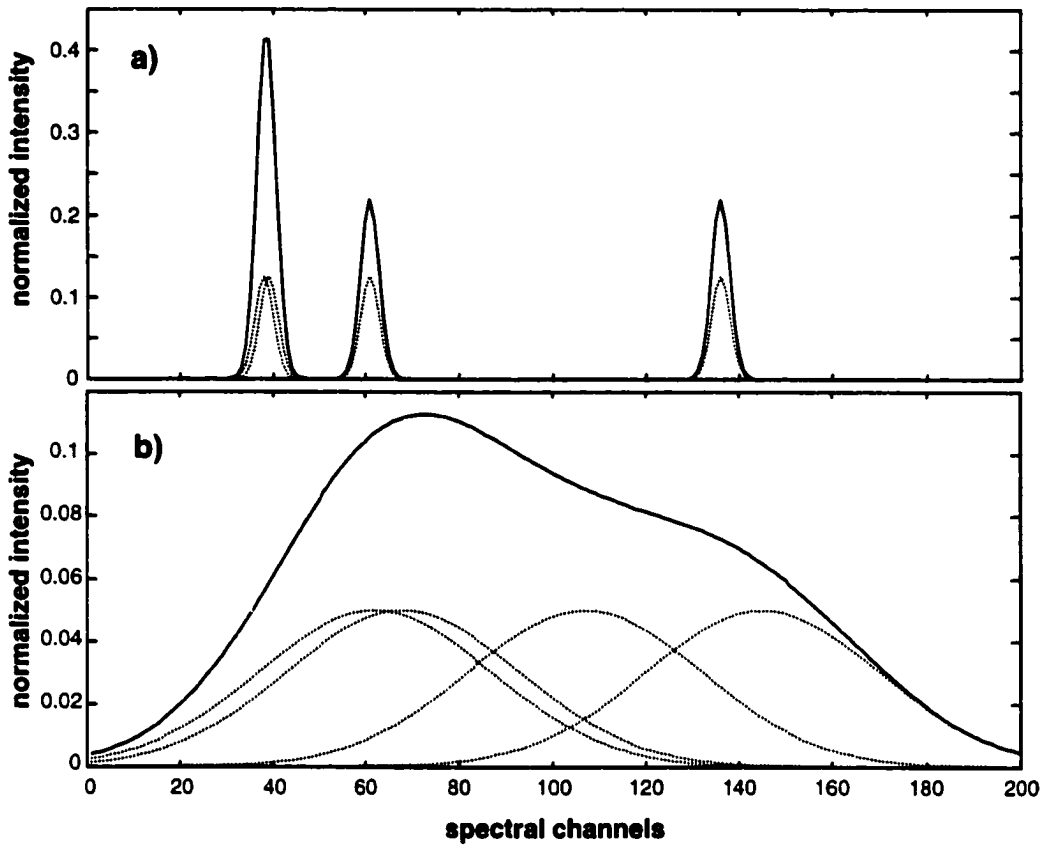


Figure 4.7 Two examples of pure-component spectra generated with random features in the spectral domain (—), and the Gaussian bands summed to generate them (·····). **a)** Pure-component spectrum generated using 4 randomly located Gaussian bands with width σ_{peak}=2 channels, and **b)** a pure-component spectrum generated using 4 randomly located gaussian bands of width σ_{peak} =25 channels.

Figure 4.7a illustrates a pure-component spectrum generated with random features using narrow Gaussian bands (σ_{peak}=2 channels), while **Figure 4.7b** shows a pure-component spectrum generated with broad features (σ_{peak}=25 channels). **Figure 4.8** shows a set of noise-free calibration spectra generated from pure-component spectra that were randomly generated by this method, using a Gaussian band width (σ_{peak}) of 25 channels. The inset shows the pure-component spectra in this instance.

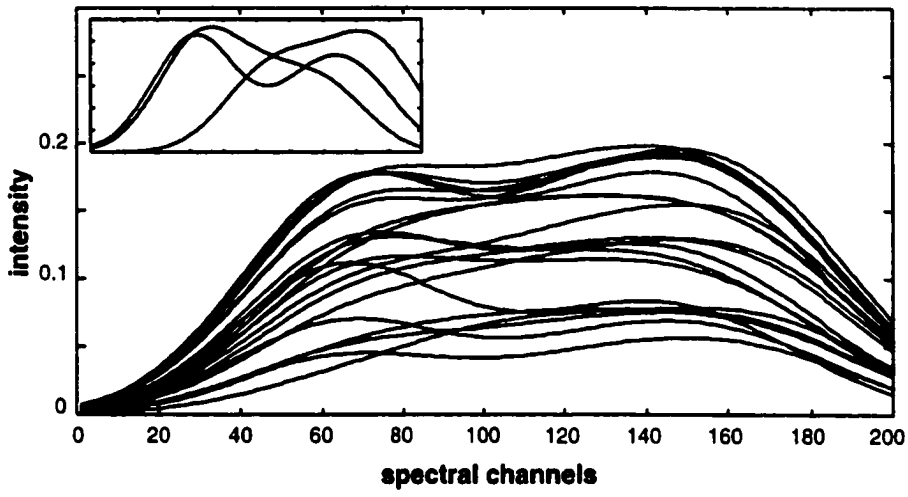


Figure 4.8 An example of 20 noise-free calibration spectra generated using random criteria for the spectral correlations, and gaussian bands of width σ_{peak} =25 channels. The inset shows the 3 pure-component spectra generating these mixture spectra.

4.3.1.3 Introduction of Correlated Measurement Errors

Typically, *iid* noise is desired in simulation studies, allowing the measurement errors to be generated from a standard Gaussian random number generator in any of a number of commercial software packages. Since the focus of this research is on the properties and behavior of drift corrupted calibration data, however, simulation of correlated measurement errors is required. Provided one pre-selects the structure of the correlated errors and embodies these characteristics in an error covariance matrix, a simple rotation can be found which correlates errors which were originally independent [47]. Pell and Kowalski have recommended using a Cholesky factorization to find the rotation matrix, although the result of this procedure can be rather unpredictable given the properties of error covariance matrices. As an alternative, we propose finding this rotation matrix in the manner used in **Section 4.2.4** to effectively 'decorrelate' the measurement errors in MLPCA. While this procedure can be easily accomplished, it requires the user to specify an error covariance matrix.

The actual error covariance structure associated with particular analytical applications and instruments has never been explored in a comprehensive manner, and consequently, there are no accepted 'pictures' of error structure associated with a particular method. Even if some particular methods were to be investigated in this regard, it is unlikely that the findings would be easily generalizable to other experimental procedures. Pell and Kowalski [47] elected to use a flat error covariance matrix, indicative of offset noise, based on some experimental evidence available for their particular application (replicate infrared transmission measurements on polyurethane films). While offset noise could have been used in this work, it is a rather simple case for derivatives, and can also be handled in a rather straightforward manner by several other methods, such as ratioing. In the simulations performed in this work, a 'ridged' error covariance matrix was used, in which the error covariance tends to be greatest between neighboring channels in the spectrum, and long range correlations are minimal, or negligible. This structure will not only provide more of a challenge for the derivative drift correction, but it is also representative of what one would expect if instrument characteristics were prominently reflected in the error covariance structure, such as cross-talk between spatially proximate detector channels, or discrete events occurring in the time domain in Fourier Transform instruments. A further benefit is that these ridged error covariance structures can be easily generated and systematically altered using simple smoothing filters.

The measurement errors for the calibration and prediction simulations in this work were generated in two steps. First, spectral noise was generated randomly from a normal distribution (mean of 0, standard deviation of 1). The ridged error covariance structures discussed above were subsequently generated by applying simple Savitzky-Golay moving-average smoothing filters of preselected widths to the original *iid* errors. More extensive error covariance structure can be introduced using wider smoothing filters (e.g., offset noise results if the smoothing filters are 200 channels wide), while filters with a width of one leave the error structure *iid*. The reader will recall (**Chapter 3**) that **Equation**

3.6 (reproduced below) conveniently allows the calculation of the specific error covariance matrix resulting from this filtering operation.

$$\mathbf{\Sigma} = \mathbf{F}\mathbf{F}^{T}\sigma^{2}_{noise} \tag{4.29}$$

With proper error covariance structure established, the error matrices were scaled to the desired magnitude. In simulation studies in which it was necessary to regulate the magnitude of the noise variance (such as comparison studies of PCR, derivative PCR, and MLPCR under different levels of correlated error), the noise variance was kept constant by scaling the errors such that the standard deviation of the errors at a given channel in the spectra, regardless of the smoothing filter used, was 0.005. Throughout this work when the level of correlated error is discussed it will be indicated by the width of the moving-average smoothing filter used to generate it. **Figure 4.9** shows a set of calibration mixtures generated from controlled spectral data, and corrupted with correlated measurement errors (σ = 0.005) using a smoothing filter width of 111 channels.

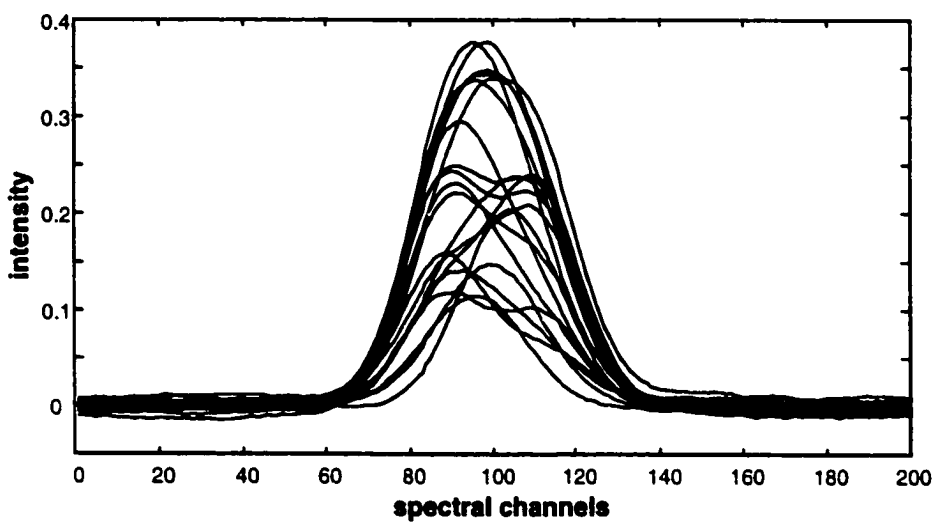


Figure 4.9 A typical example the controlled calibration data when corrupted by drift noise introduced with a 111-point moving average filter.

4.3.2 Experimental Data

Experimental data, consisting of diffuse reflectance measurements on 16 acrylonitrile-butadiene-styrene (ABS) formulated resin samples, were supplied by Dow Chemical Company [48]. Measurements were obtained using a Bomem MB155S spectrometer, outfitted with a DiffusIR™ attachment, designed to allow large area (13 cm^2) reflectance sampling on coarse materials. The indium-arsenide detector was thermoelectrically cooled to minimize temperature fluctuation effects. Petri dishes were filled to a depth of approximately 1 cm with the ABS samples, and spectra were acquired through the bottom of the containers using the spectrum of a spectralon disk as a reference. Five repeat analyses were performed for each sample (each in a different dish) in the region from 10005-3695 cm^{-1} using 16 cm^{-1} resolution, and 128 scans. The resulting data matrix, therefore, consisted of 80 spectra (5 repeats for 16 samples).

Initial data exploration found several spectra showing unusually high leverages and concentration residuals. These 7 spectral vectors (spectrum numbers 51-55, 74, and 75) were excluded from building subsequent calibration models, leaving a reduced calibration set of 73 spectra. As is expected, calibration performance was significantly enhanced by these deletions. It is possible that some wavelength regions in the spectra would prove more useful than others for calibration and prediction purposes. Wavelength selection procedures (methods for identifying the wavelengths which appear to have greater utility in prediction) may have reduced the absolute prediction error of the resulting calibration models, however the intended use of the experimental data was for the *comparison* of various preprocessing methods for multivariate calibration. Since only the relative calibration performances are of interest, wavelength selection was deemed an unnecessary procedure. The diffuse reflectance spectra for the 73 calibration samples are shown in **Figure 4.10**.

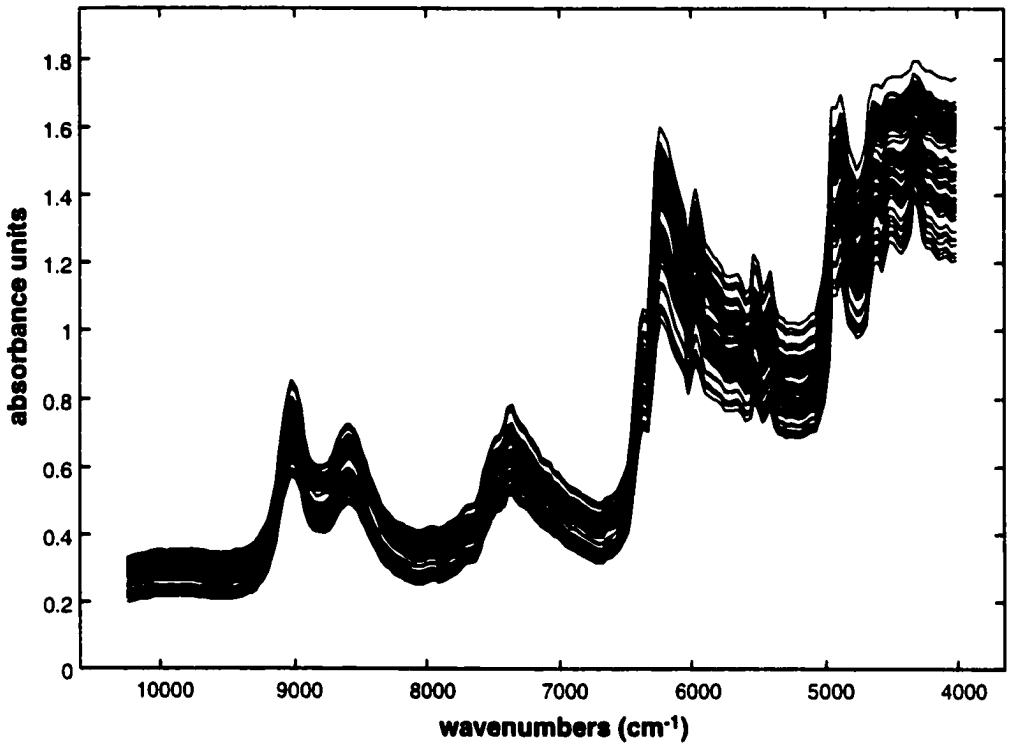


Figure 4.10 73 NIR diffuse reflectance spectra constituting the experimental data set.

4.3.3 Computational Details

All computations performed in the course of this work were carried out on a Sun Microsystems Ultra 60 with 4–300 MHz CPU's and 1 GB of RAM. All scripts were written in house, and executed in MATLAB v.5.2 and 5.3 (The Mathworks, Natick, MA) for the Unix platform.

4.4 Results and Discussion

4.4.1 Derivative Filtering and Signals

In order to explore the effect of derivative filtering on chemical signals and multivariate calibration figures of merit, 500 sets of 3 pure-component spectral vectors were generated with random features as described in the **Section**

4.3.1.2. To simulate spectra with relatively broad features, the Gaussian band width was chosen to be 75 channels. An example of one of the sets of pure-component spectra generated under these conditions is shown in **Figure 4.11a**. Access to the noise-free pure-component spectra allowed calculation of the *true* figures of merit for each system according to the formulae in **Chapter 1**. The true NAS vectors were calculated (via **Equation 1.53**) for each component before and after the spectral data were treated with derivative filters. Using these true NAS vectors and the pure-component spectra, the exact *SEL*'s could be determined for each component before and after filtering. **Figure 4.11b** shows the results obtained from these simulations for the first analyte of the three using a 11-point quadratic second derivative filter. (The results were highly similar for all three components, so only the results for component 1 are shown.) The *SEL*'s shown in **Figure 4.11b** all improved to some degree after derivative treatment when compared to the results obtained for the untreated spectral data. When the simulation was repeated with spectra exhibiting higher frequency components (Gaussian band width at 10 channels), the beneficial effects of derivative filtering were much less pronounced. Typical pure-component spectra, and the selectivity simulation results are shown in **Figures 4.11d** and **e** respectively. Multivariate *S/N* ratios were also calculated in these studies and are summarized for the two scenarios (broad and narrow featured spectra) in **Figures 4.11c** and **f**. The observable results of the *S/N* studies seem to contradict the *SEL* studies (derivatives enhance the *SEL*'s in the broad spectra, but degrade the *S/N*), and are highly dependent on the characteristics of the signals. Distinctly different trends are observed in the two different cases studied here.

Further simulation studies were carried out involving the calculation of the true multivariate *S/N*, and some typical results are shown in **Figure 4.12**. In these simulations the spectral features were varied from narrow to very broad (σ_{peak}: 5, 25, 55 channels), and the derivative filter width was altered from narrow

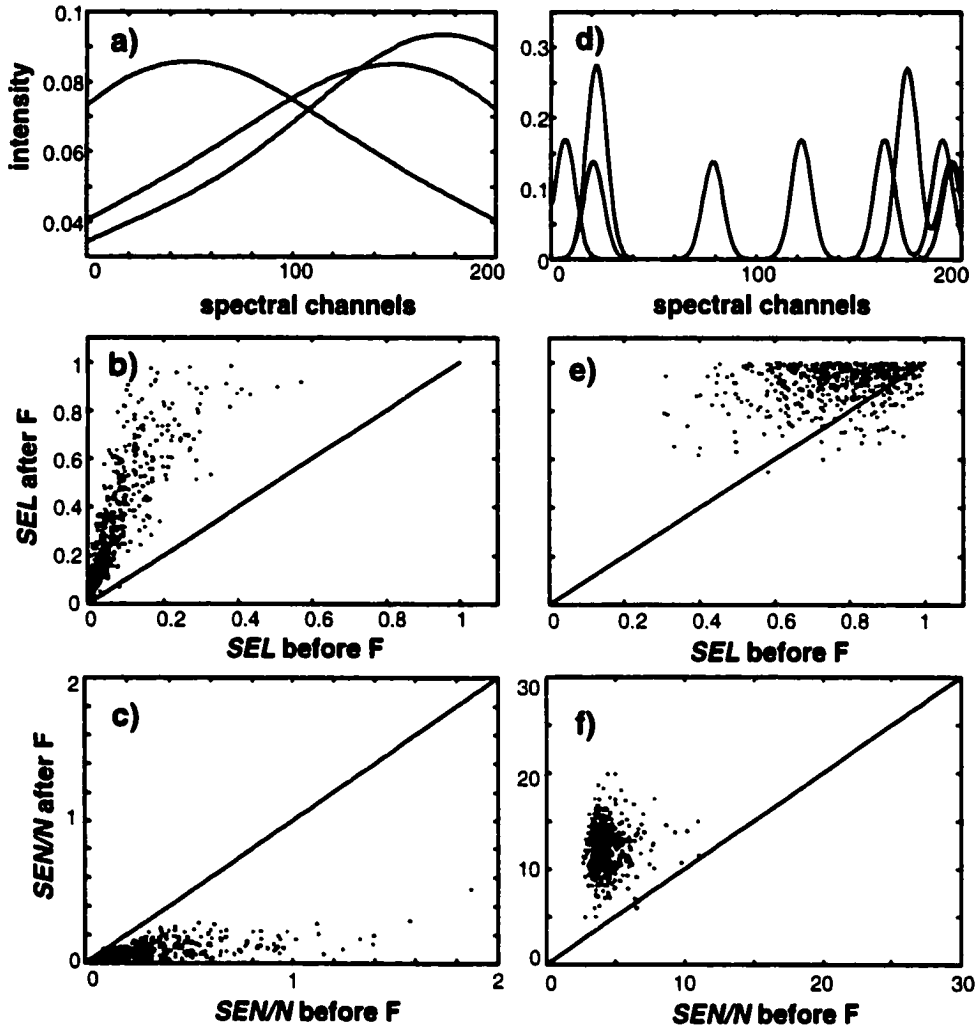


Figure 4.11 Figures of merit studies on sets of pure-component spectra. **a)** An example of the pure-component spectra generated randomly with very broad features. **b)** and **c)** depict the results of 500 replicate *SEL* and *S/N* measurements on data with these characteristics both before, and after filtering with a 11-point quadratic second-derivative filter. **d)** Sample pure-component spectra used (narrow features) in an identical study of **e)** *SEL* and **f)** *S/N* changes as a result of derivative filtering.

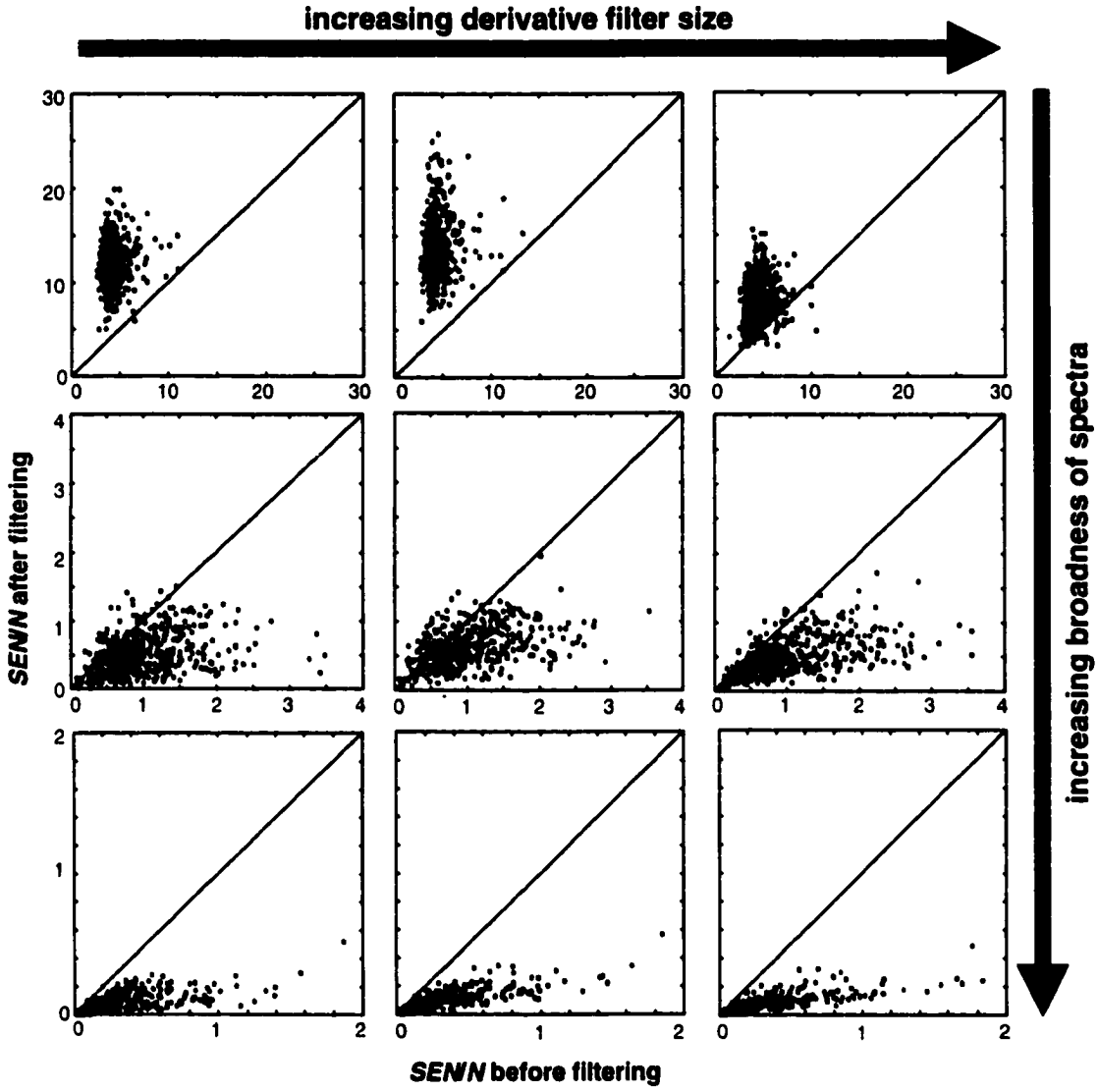


Figure 4.12 Multivariate *S/N* studies with varying spectral characteristics and derivative filter treatments both before, and after derivative preprocessing. The grid-layout of the 9 plots is as follows: travel down the vertical axis corresponds to increasing broadness of the pure-component spectra (σ_{peak} values were 5, 25, and 55 channels), travel from left to right on the horizontal corresponds to increasing the second-derivative filter width (filter widths were also 5, 25, and 55 channels).

to very broad (filter width: 5, 25, 55 channels). The results clearly indicate that no singular conclusion can be drawn, *i.e.*, we cannot say whether derivative filters will improve, or degrade the multivariate figures of merit since the results are entirely dependent on the pure-component spectra involved, and the error covariance structure of the data. Some trends are apparent in **Figure 4.12**, such as a general degradation in filter performance with spectra exhibiting broader features. Interestingly, this observation contradicts popular knowledge on derivative filter performance which suggests that broad spectra stand to benefit most from derivative preprocessing. Extrapolating these results to the general case, however, is tenuous since the results will invariably depend on the situation at hand.

Figure 4.13 demonstrates two synthetic contradictory situations. In "Case One" (**Figure 4.13a**), the spectra exhibit both high and low frequency characteristics, but the low-frequency content of the pure-component spectra essentially only contributes to overlap, and is therefore of little model utility in prediction. All of the distinction between the three components comes in the high-frequency signals located at approximately 150 channels. Indeed when the net analyte signal vectors are calculated for each component, it is evident that the low-frequency overlap has been ignored for all three components. Since derivative filtering will heavily attenuate low-frequencies, it can be anticipated that the selectivities will be dramatically improved in this case, because the overlapping low-frequencies will be attenuated. This is confirmed in **Figure 4.13b**, where the derivative pure-component spectra have been calculated using a 13-point quadratic second-derivative filter. Their corresponding NAS vectors are given as well.

In Case Two (**Figure 4.13c**), the pure-component spectra again are composed of both high and low frequencies, however in this case the low-frequency character of the signal is of use in resolving the components, while the high frequencies only contribute to overlap. Derivative filtering in this case

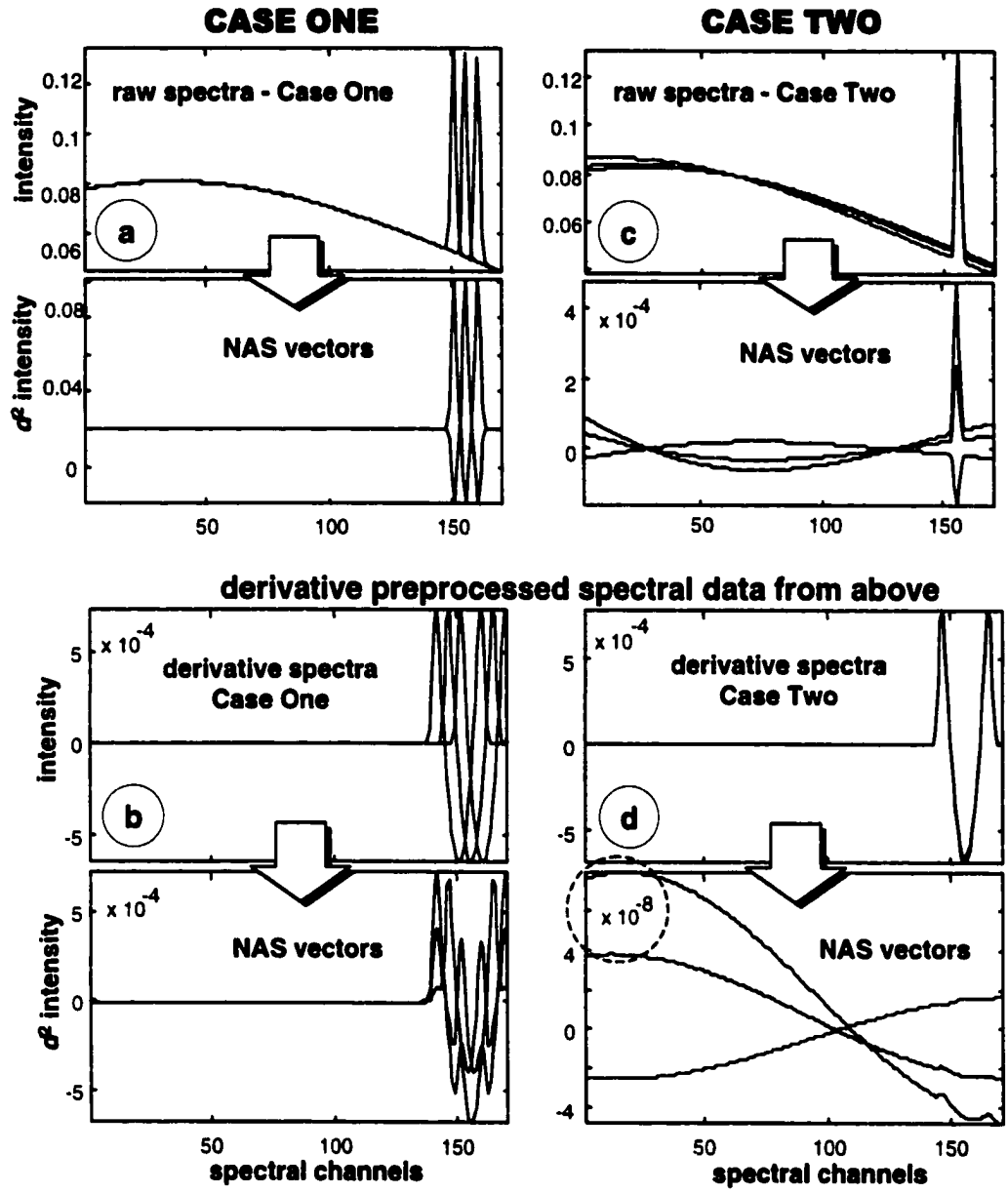


Figure 4.13 Simulation of two distinct cases for derivative filtering in which the filter performance is highly dependent on the frequencies of importance in calibration. In **Case One**, the information is contained at high frequencies, while in **Case Two**, the information resides at low frequencies.

Table 4.1 *SEN* and *SEL* values for Case One and Case Two both before (d^2 - no), and after (d^2 - yes) treatment with a 13-point second-derivative filter. The results are shown for all three components. Case One shows an enhancement in the *SEL*'s upon differentiation, while Case Two shows a substantial decrease in *SEL*.

		Case One		Case Two	
	d^2	*SEN*	*SEL*	*SEN*	*SEL* (x 1E-3)
component 1	no	0.1316	0.1394	8.69E-04	0.9271
	yes	0.0019	0.8236	6.72E-07	0.2689
component 2	no	0.1315	0.1393	2.93E-04	0.3112
	yes	0.0023	0.9948	2.16E-07	0.0891
component 3	no	0.1317	0.1395	4.41E-04	0.4682
	yes	0.0019	0.8236	3.18E-07	0.1329

proves disastrous, since the low-frequency attenuation of the derivative filter eliminates the useful information in the pure-component spectra, and relatively enhances the uninformative higher frequencies. These effects are apparent in the actual values of the *SEN* and *SEL* for these two data sets, given in **Table 4.1**.

Although it is difficult to make definitive generalizations from these simulations alone, several points can be made. The effect of derivative filtering on the multivariate selectivity of an analyte is certainly difficult to predict even in the absence of drift noise. The change in *SEL* with filtering is related to the frequency composition of the pure-component spectra, and to the characteristics of the derivative filter used. Savitzky-Golay derivative filters will attenuate both low- and high-frequency components of the signals, and the change in *SEL* resulting from these signal modifications is entirely dependent on the location of the band-pass region of the filter with respect to the frequencies in the pure-component spectra that are important for successful calibration and prediction. When the additional complication of baseline drift is considered in the calculation of the multivariate *S/N*, the results are even more difficult to predict. The derivative filter band-pass, frequency composition of the signals, and now the relation of the error covariance matrix to the NAS vectors for the analytes are all

of fundamental importance, making it nearly impossible to state with any certainty *a priori* whether anything is to be gained by derivative filtering from a figures of merit standpoint.

Due to the uncertainty involved in the performance of derivative preprocessing from a figures of merit perspective, the analyst is resigned to using a trial-and-error approach to carefully match the filter band-pass to the situation at hand, a noted drawback of derivative preprocessing.

4.4.2 Derivative Filtering and Noise

It was proposed in the theoretical portion of this work that derivative filters can be thought of as attempting to diagonalize the error covariance matrix for the data and render the noise uncorrelated. **Figure 4.14a** shows a vector of noise which shows significant levels of drift. In **Figure 4.14b**, this noise vector has been differentiated using a 5-point quadratic second derivative filter, and in **Figure 4.14c**, a 13-point quadratic second-derivative filter function was used. As far as inspection can tell, the noise treated with the 5-point filter appears to be correlated to a much lesser degree than the original, and low-frequency character is certainly no longer blatantly obvious. Although the treated noise resulting from the 13-point filter looks to have a higher frequency composition than the raw data, some low-frequency drift is observed to persist. Although this example illustrates the visual reduction in correlated errors, a more rigorous evaluation must come through error covariance matrix comparisons.

In **Figure 4.14d**, the error covariance matrix for the drift-corrupted data (calculated from 50 repeat measurements of the noise sequence) is shown as a contour plot. The application of the 5-point quadratic second-derivative filter to the raw data results in the error covariance matrix shown in **Figure 4.14e**. From examination of the error covariance matrices, it is clear that error variation in this filtered data is almost exclusively characterized on the diagonal, implying that substantially less correlated error remains in the derivative filtered noise.

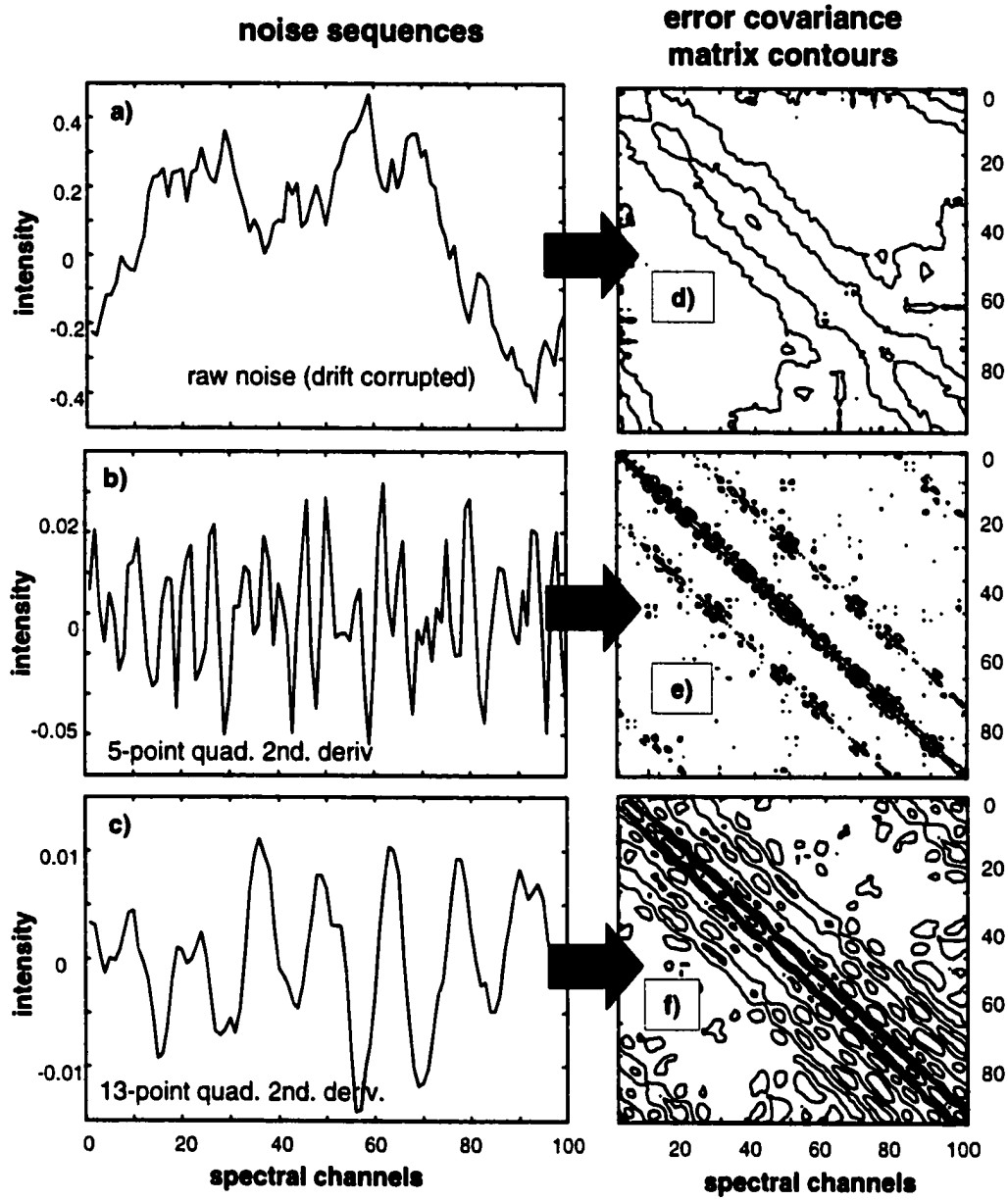


Figure 4.14 **a)** A noise sequence showing significant levels of drift noise. **b)** A derivative spectrum of **a** using a 5-point quadratic second-derivative filter. **c)** A derivative spectrum of **a** using a 13-point quadratic second-derivative filter. **d**, **e**, and **f)** Error covariance matrices corresponding to each noise sequence determined experimentally from 50 replicate measurements of the noise sequences.

The off-diagonal 'lines' that can be observed in this error covariance matrix result from the suboptimal treatment of the observed error covariance matrix by derivative filtering. **Figure 4.14f** shows the error covariance matrix resulting from treating the noise with the 13-point second derivative filter. Substantial error covariance remains after derivative treatment in this instance. It is apparent that the smaller sized filter does a better job of eliminating correlations among the errors and thus reducing the contribution of baseline drift.

These observations can be rationalized from a theoretical standpoint. **Figure 4.15** shows the noise power spectrum for the raw noise, and the NPS after filtering with a *difference filter* and a 13-point quadratic second-derivative filter. The NPS of the raw noise shows the low-frequency dominance characteristic of drift-noise. The 13-point quadratic second-derivative filter treatment results in a colored NPS, with frequencies in the 0.1-0.2 range

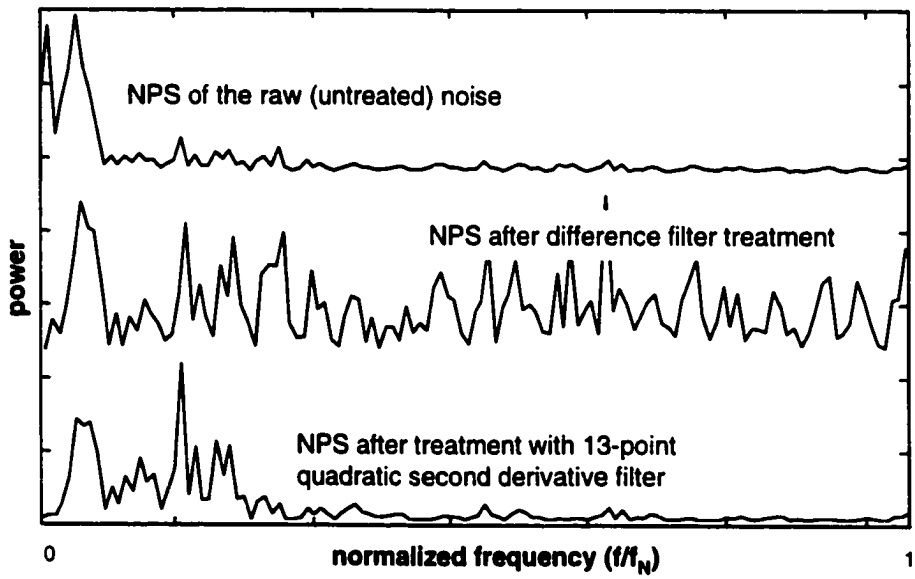


Figure 4.15 Noise power spectra (NPS) showing the frequency content of a raw noise sequence corrupted with drift, and the resulting NPS's after treatment of that noise with a difference filter and 13-point quadratic second-derivative filter.

dominating. The difference filter, however, leaves essentially white noise. From a drift noise perspective, then, the more narrow derivative filters tend to be more successful at reducing the low-frequency dominance in the noise power spectra without introducing other 'colors' in the noise.

4.4.3 Maximum Likelihood PCA and Drift Correction

The application of MLPCA and MLPCR to drift-noise corrupted calibration data was studied under two conditions: using the true error covariance matrix, and using estimates of the error covariance matrix.

4.4.3.1 MLPCR with the True Error Covariance Matrix

Since the correlated error was introduced in the simulated data with known characteristics (see **Section 4.3.1.3**), it was possible to use this information directly in MLPCA, and MLPCR. **Figure 4.16** shows a set of simulated calibration spectra, heavily impaired by drift noise, both before, and after treatment with MLPCA. For comparison, the PCA of the data at the same rank is also shown. It is clear that the level of drift in the MLPCA estimated spectra has been significantly reduced relative to both the raw data, and the PCA estimated spectra. The wildly fluctuating drift noise has been corrected to a remarkable extent.

To properly compare the proposed method of optimal drift correction using MLPCA to derivative methods, large simulation studies were carried out in which the level of correlated noise was systematically varied while monitoring the calibration performance of PCR, derivative PCR (with a wide variety of derivative filters), and MLPCR. These calibration sets were generated from controlled spectral data with the standard deviation of the noise fixed at 0.005, and are consequently similar to the spectra shown in **Figure 4.9**. In **Figure 4.17**, the *RMSEP*'s for MLPCR and PCR, and a variety of derivative PCR methods are shown. **Figure 4.17a** displays sample results for derivative PCR with linear first-derivative preprocessing (varying filter widths), while **Figure 4.17b** shows the result of using quadratic second-derivative filters. The two figures are very

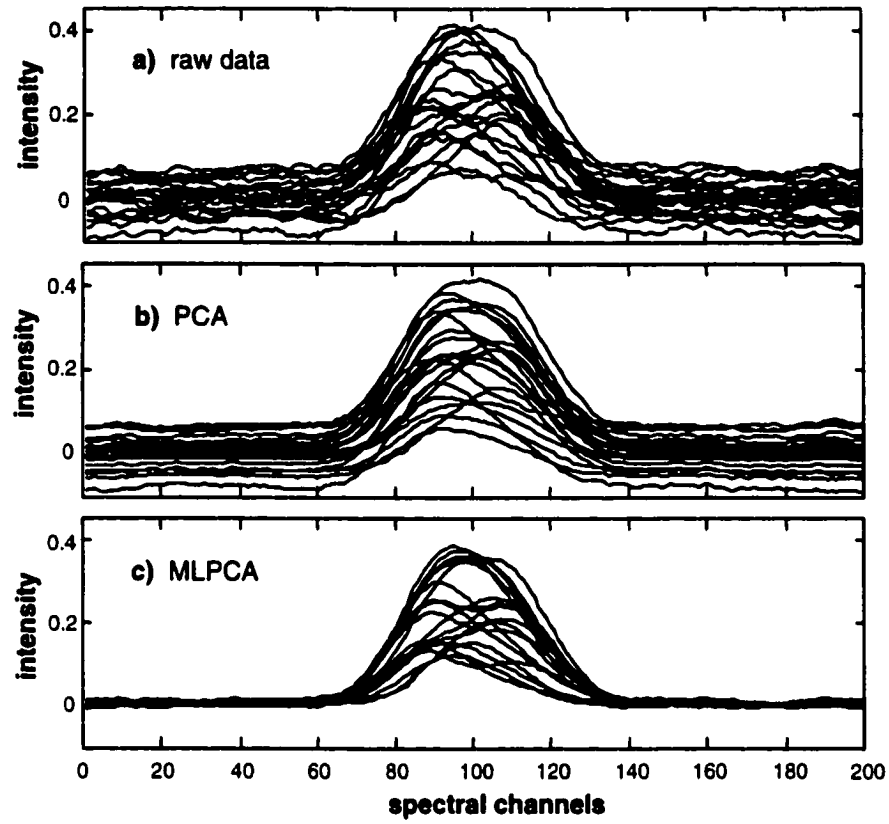


Figure 4.16 An illustration of the drift correction power of MLPCA compared to conventional PCA. **a)** 20 calibration spectra generated under controlled conditions and corrupted with substantial drift noise using a filter width of 95 channels (the noise has been scaled up to σ=0.05 to illustrate the point more clearly). **b)** The calibration data reconstructed from a rank 3 principal component subspace, and **c)** The calibration data reconstructed from a rank 3 space using MLPCA.

similar, except for the behavior of the very narrow filters. First-derivative filters achieve a greater degree of smoothing (although less drift reduction) than their second-derivative counterparts, and so the very narrow first-derivative filters still allow derivative PCR to perform reasonably well. Second-derivative filters, however, achieve a much lower degree of smoothing at narrow filter sizes than do first-derivative filters, and so derivative PCR calibration models built using narrow second-derivative filters have the potential to be heavily impaired by high-frequency noise, an effect often observed in these simulations, and in practice.

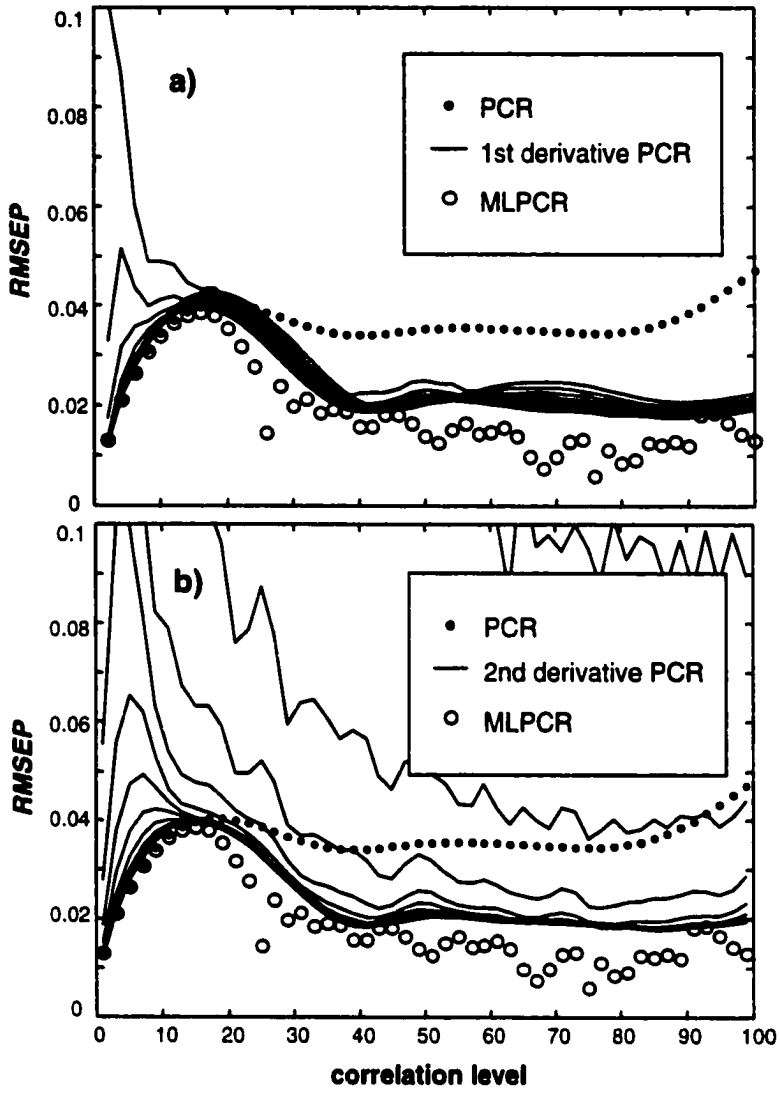


Figure 4.17 Simulation results comparing the performance of PCR to derivative PCR and MLPCR. **a)** PCR, MLPCR and linear first-derivative PCR, and **b)** PCR, MLPCR and quadratic second-derivative PCR.

When the level of correlated error is minimal, MLPCR can be seen to provide no enhancements over conventional PCR. This is, of course, expected since in the presence of uncorrelated measurement errors MLPCR reduces to simple PCR, and the two methods should perform equivalently. In contrast, derivative PCR often performs considerably worse than PCR when there is little or no error covariance. With no correlated error present, derivative filtering can

achieve no improvements in calibration performance by drift reduction. With the derivative filter matrix being applied to an effectively *iid* error covariance matrix, the derivative filter is, in these cases, simply *introducing* correlated measurement error into the system, and therefore creating noise conditions that are not well suited for PCR. In fact, with *iid* noise, the application of symmetric (even) derivative filters would be expected to adhere to the theoretical proof given in **Chapter 3**, suggesting that symmetric Savitzky-Golay filters of any sort cannot be expected to enhance calibration performance. These effects, coupled with the possible degradations in selectivity and *S/N* that can result from derivative filtering as previously noted, could lead to the observed poor performance of these derivative filters with low levels of drift noise.

As the level of correlation among the errors becomes more significant, both derivative PCR, and MLPCR are observed to surpass conventional PCR in predictive success. The enhancements observed for derivative PCR should result from drift reduction by **Equation 4.10** with possible added contributions in *SEL* and *S/N*; the improvements observed for MLPCR over PCR arise from MLPCR's use of the error structure in projecting the prediction spectra onto the calibration space, which conventional PCR ignores. In all simulations performed under these conditions, the performance of MLPCR was consistently found to be comparable to, or better than derivative PCR at its best (*i.e.*, when the best possible combination of filter characteristics were found).

4.4.3.2 MLPCR With an Estimated Error Covariance Matrix

Since the true error covariance matrix for the data is never known in practice, simulation studies were conducted to compare MLPCR using estimates of the error covariance to derivative PCR.

As discussed in **Chapter 1**, there are several simplifying assumptions which can be made regarding the error covariance structure of the spectral data. In the previously discussed simulation studies the known error covariance matrix was used (which is known also to be identical for all samples). In practice, replicate measurements may be acquired on each calibration sample, which

allows observation of the changes in the noise, while the chemical responses are (hopefully) held constant. This allows error covariance structure to be estimated for each sample available in replication. If one assumes the error covariance matrices for all of the samples are approximately the same (an assumption known to be valid in the simulation studies), a *pooled* estimated error covariance matrix can be obtained by averaging all of the individual sample-estimated error covariance matrices together. This pooled error covariance matrix estimate, in turn, can be used in MLPCR for drift-noise correction procedures and calibration. If the assumptions of equal row-covariance are valid, then the pooled error covariance estimate should be more accurate than simply using one set of replicates to ascertain error structure.

To test the utility of this approximation, replicate calibration spectra were generated for each of 20 calibration samples generated under controlled conditions. There are essentially two steps in this procedure: (1) calculating error covariance matrices for each sample from the available replicates, and (2) averaging all of these error covariance matrices together. As a controlling factor in the simulations, then, the number of repeat spectra involved in the estimation from each sample was used. In all cases, all 20 sample-specific error covariance estimates were averaged to yield a *pooled* error covariance matrix estimate. The replicate spectra were *only* used to estimate the error covariance structure, and were not used in the actual construction of the calibration model itself.

RMSEP's for MLPCR under these conditions are shown in **Figure 4.18.** The performances of some derivative PCR models are also included for direct comparison. It is apparent that, with this method of pooling the error covariance estimates, MLPCR still performs extremely well under these conditions. In this case, only a couple of repeats of each sample were necessary for MLPCR to surpass the predictive performance of PCR, and the derivative PCR methods shown. Although the performance of MLPCR appears to be optimal at a certain number of replicates (*ca.* 6 or 7), this minimum is merely a statistical aberration, and does not indicate a generalizable trend.

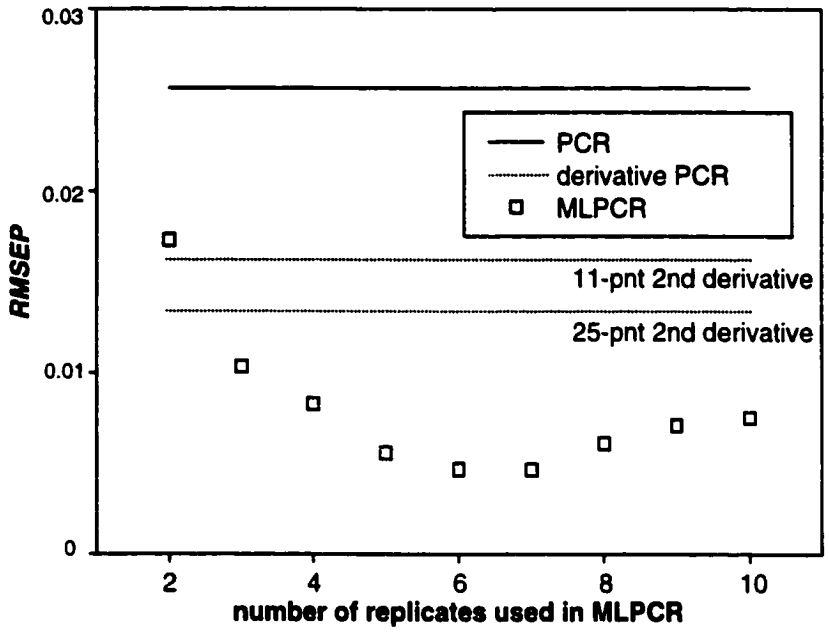


Figure 4.18 *RMSEP* for MLPCR as a function of the number of replicates of each calibration sample used to estimate the *pooled* error covariance structure. The performances of PCR and derivative PCR are shown for reference.

Pooling the error covariance estimates in this fashion still provides very reasonable estimates of the error covariance matrices provided there are a sufficient number of samples available over which to average the estimated error covariance matrices. In essence, using only 2 repeats per sample with pooling still yields an estimate which has been generated from 40 measurements. To truly challenge the performance of MLPCR with extremely poor estimates of error structure available, an experiment was simulated in which the 'analyst' only has replicates of a single sample available to estimate the structure of the drift – no pooling is possible. The number of replicate measurements used in the estimate was used as a control criterion. While the effect of the validity of the estimate on the performance of MLPCR is a very complex matter, some qualitative discussions can be made.

The results shown in **Figure 4.19** are for **a)** low, **b)** medium, and **c)** high levels of correlated errors. It is clear that with lower levels of correlated error, a

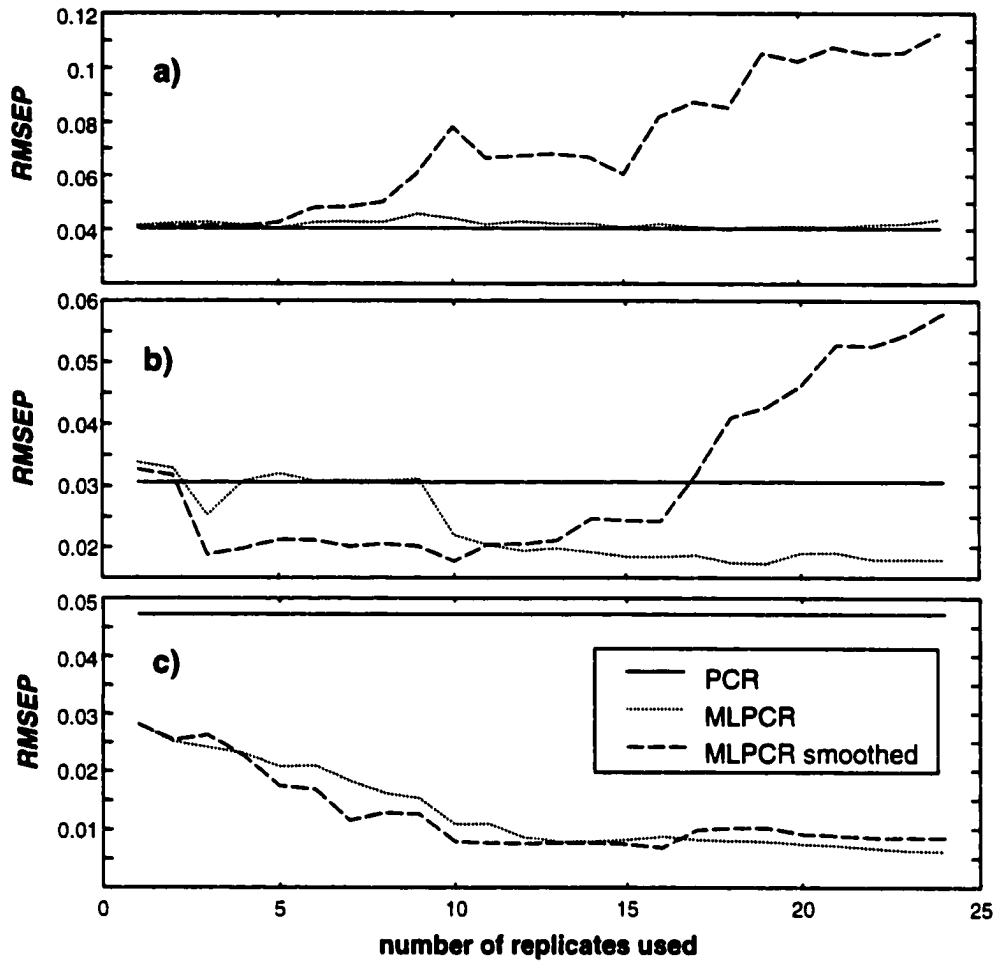


Figure 4.19 Performance of MLPCR as a function of the number of replicates used to estimate the error covariance structure of the data (*no pooling*). The performance of MLPCR on the same data when the error covariance estimate was smoothed (25-point block smooth) prior to use in MLPCR. Plots **a**, **b**, and **c** correspond to low, medium and high levels of drift (drift introduced with smoothing filter widths of 19, 59, and 99 channels).

large number of replicates is needed to achieve prediction errors significantly below those of PCR, while significant amounts of correlated error greatly reduced the number of replicates required for MLPCR to perform better than PCR. Presumably, subtle drift noise is of little detriment to PCR, and additionally, these subtleties are difficult to characterize from only a few replicates because of their relatively low 'signal-to-noise' ratio, whereas large correlations are significantly deleterious to PCR, and the error structure is prominent enough to be reasonably

estimated with only a few replicates. It will also be noted that with low and moderate error correlation, MLPCR performs comparably to PCR when only a few replicates are available. This is likely to be due to poor estimation of the covariance matrix which effectively is equivalent to the *iid* normal case.

The hazard with using poorly estimated error covariance matrices is that the estimated error covariance matrix has a very low signal-to-noise ratio. The performance of MLPCR with poorly estimated error covariance matrices can be improved, to some degree, by smoothing the error covariance estimate with a simple moving average filter. Given that the information in the error covariance matrices is 2-dimensional (2D) it is recommended that a 2D smoothing filter (a block smoother) be used. In 2D smoothers, the points in a block of size w x w (where w is preferably odd) are averaged to obtain a smoothed value at the middle of the block. This method can improve the performance of MLPCR if the available estimate of the error covariance matrix is of very poor quality (*e.g.*, few calibration samples and few replicates of each sample spectrum). In situations where the error covariance matrix is well-approximated, the smoothing procedure is unnecessary, and may in fact hinder the drift correction due to the distorting effects of the block smooth. This error covariance smoothing technique was applied to the same data used in the above simulations, with the results being included in **Figure 4.19** for comparison. The block filter size was chosen to be 5 x 5, and no attempt was made to find the best performing block filter under these circumstances. It is likely that better results *can* be achieved if some effort is made in selecting the block filter size, however for the purposes of this work, it was only deemed necessary to show that improvements can be achieved from the operation. In the low correlated error scenario (**Figure 4.19a**), it is apparent that distortion from the block smooth leads to artifacts in the estimated error covariance matrix which hamper calibration performance relative to MLPCR (without block smoothing), and PCR. With very high degrees of correlated error (**Fig. 4.19c**), the block smoothing procedure enhances the performance of MLPCR to some extent, but shows no great improvement in performance over

conventional MLPCR. It is likely that in these situations, error covariance structure is well estimated with a minimal number of samples, and thus the block smooth does little to improve the accuracy of the information contained in the estimate. With a mediocre amount of correlated error, however, the block smooth leads to a significant improvement in the performance of MLPCR with few replicates, and, in the simulations conducted in the course of this work, generally halves the number of replicates required to achieve a given *RMSEP*. These conditions are best suited for the application of the block smooth, since the error covariance structure is prominent enough to cause serious degradations in the *RMSEP* for PCR, but the signal-to-noise ratio of the error covariance matrices estimated from only a few replicates is still quite low. As expected, once a reasonable number of replicates are used to estimate the error covariance structure (*ca.* 12 in **Figure 4.19b**), the block smoothing procedure does little to enhance the quality of the estimated error covariance matrix, and thus, does little to improve the performance of MLPCR. Overall, the error covariance smoothing did enhance the performance of MLPCR when significant levels of correlated error were present, and tended to reduce the number of replicates required to achieve accuracy beyond that of derivative PCR methods.

4.4.3.3 Experimental Data

MLPCR and derivative PCR were compared to conventional PCR in handling baseline drift in the experimentally acquired diffuse reflectance spectra previously described. Examinations of the error covariance structure revealed very prominent drift effects, and that it was very similar between different samples of the calibration set, allowing for the pooling of the replicate estimates. Error covariance matrices were calculated for each set of sample replicates, excluding the set that were almost all removed in exploratory data analysis (spectra 51-55, see **Section 4.3.2**). Subsequently these 15 error covariance estimates were combined to yield a pooled estimate of the error covariance matrix (shown in **Figure 4.20**), which was used with MLPCR for multicomponent calibration.

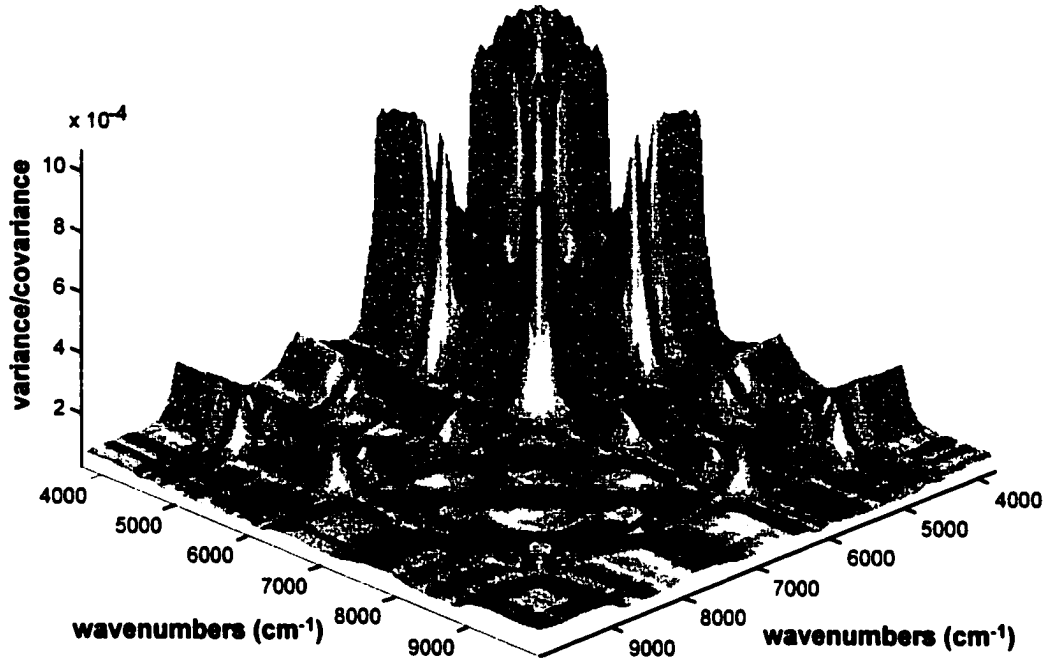


Figure 4.20 The estimated error covariance structure for the experimentally obtained NIR diffuse reflectance spectra.

Because a distinct prediction set was unavailable, cross-validation (CV) [49] was used to select model parameters and estimate the predictive performances of the competing methods. Cross-validation is a method for 'pretending' to have a set of prediction/validation samples without truly having them (for this reason, it is often referred to as internal validation). In the procedure, some of the samples are set aside as prediction samples while the calibration model is constructed from the remaining samples. The resulting model is then employed to predict the concentrations for the 'unknown' samples. A different lot of samples are removed, and the model is reconstructed, *etc.*, until all samples have been left out of the calibration process and predicted. With this method, a reasonable estimate of how well the calibration model is *predicting* can be garnered, without actually having a separate prediction set available. The number of samples set aside at each iteration is typically decided by the user, but in this case, one sample (and its replicates) was left out (leave-one-out CV)

during each iteration. The root mean-squared error of cross-validation (*RMSECV*) was used as an indicator function, which is defined as

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}{m}} \tag{4.30}$$

where $\hat{y}_i$ is the model's estimate of the concentration for the *i*th sample during the CV, y_i is the known concentration for that sample, and m is the total number of samples. The reader will note that **Equation 4.30** is very similar, although non-equivalent, to the formula for *RMSEP*. It is, however, often a reasonable estimator of the *RMSEP*.

Several model parameters had to be systematically altered in the cross-validation regime: number of latent variables (in all models), derivative filter width (derivative PCR), derivative filter order (derivative PCR), and first-, second-, *etc.*, derivative (derivative PCR). Lacking other information, the model parameters corresponding to the absolute best value of the *RMSECV* were taken to indicate the best performing model. For derivative PCR, the properties of the derivatives were set (*e.g.*, PCR with quadratic first-derivative preprocessing), and then the best filter width and number of latent variables were selected under these conditions by CV. As a caveat, it should be noted that the selection of optimal calibration characteristics (particularly the number of latent variables), and prediction error estimation is best done using a *true* (external) validation data set. That being said, however, it is the relative performances that are of primary interest in this work. Lacking a validation set, cross-validation error was the only reasonable measure available for this task.

Figure 4.21 depicts the result of a sample derivative preprocessing drift correction procedure (13-point quadratic second-derivative), and MLPCA drift correction on the 73 calibration spectra. The magnitudes of the derivative spectra have been dramatically reduced as can be seen by the scale of the absorbance axis, and some variability in the calibration spectra has been

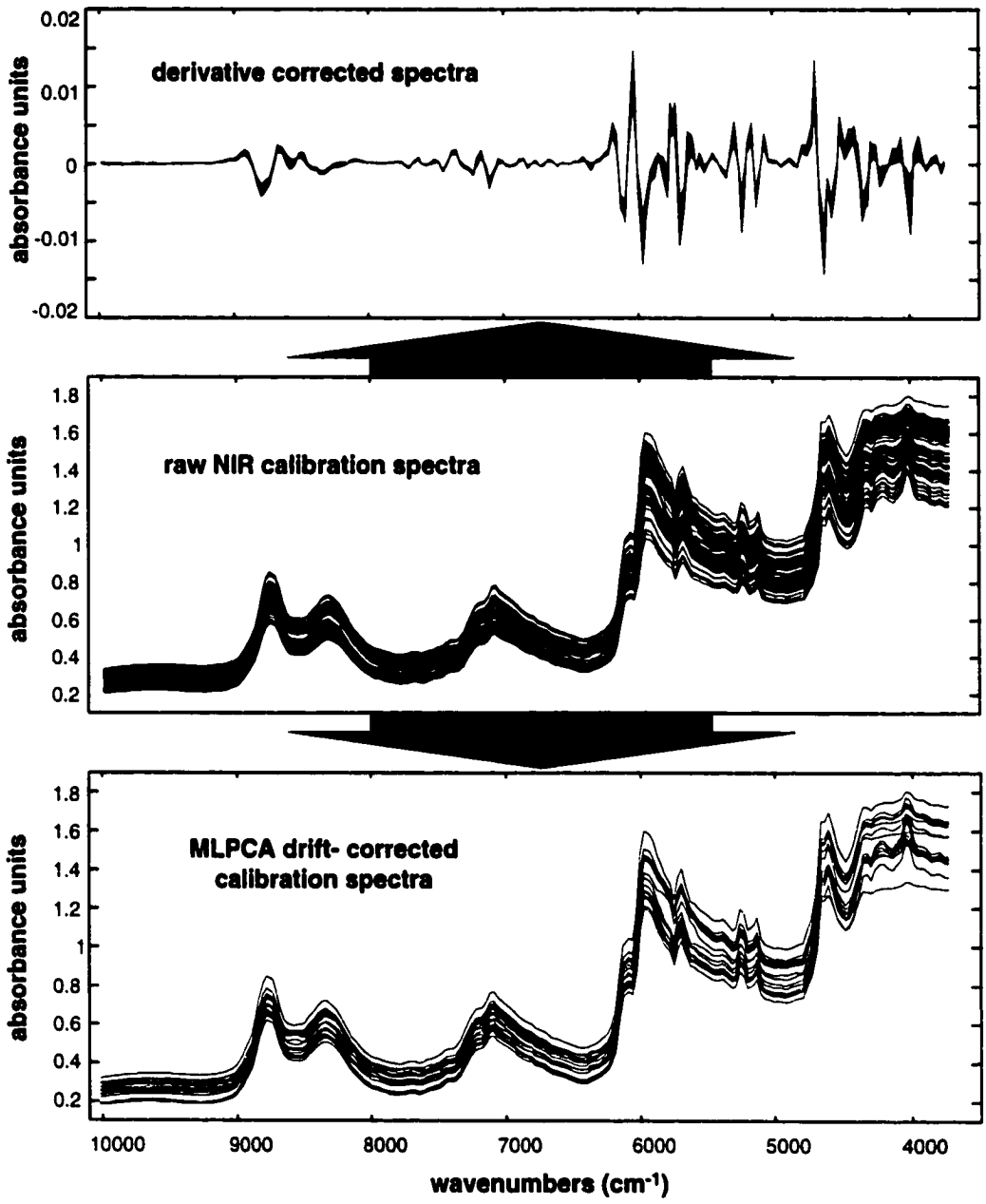


Figure 4.21 A visual comparison of the effects of drift correction on the 73 calibration spectra using a 13-point quadratic second-derivative filter and MLPCA.

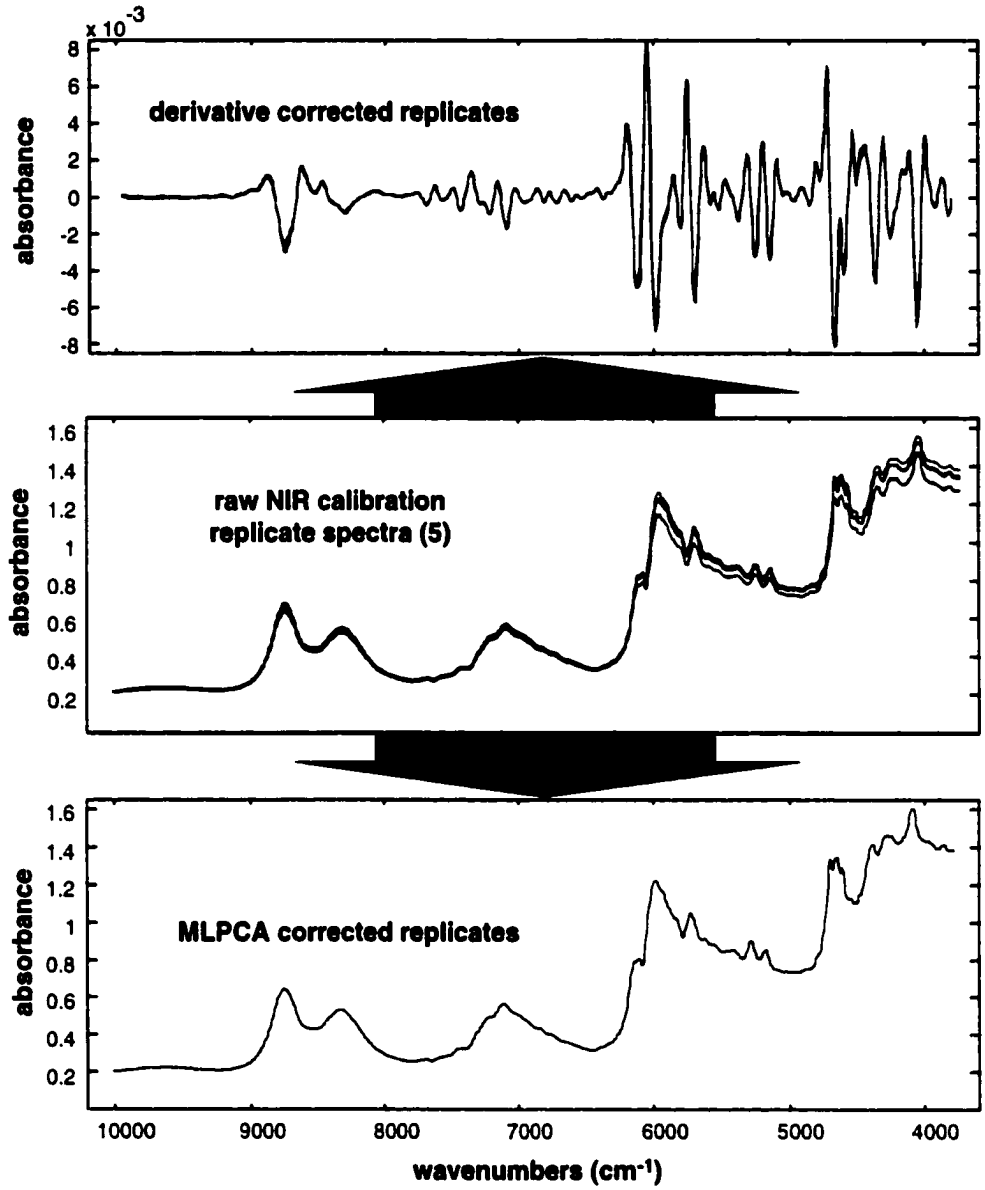


Figure 4.22 A close-up illustration of drift correction on 5 repeat spectra of the same sample using derivative preprocessing (13-point quadratic second-derivative) and MLPCA drift correction.

reduced. Although difficult to see, the greatest variation in the derivative spectra appears in the spikes that remain, while the variation in most of the flatter areas of the original spectra has been all but annihilated. Based on the prior discussion of derivative filter properties, this comes as little surprise, since the lower-frequency elements of the spectra are bound to be heavily attenuated. The MLPCA corrected spectra, however, exhibit rather different variations. Significant variance still exists, even in the very flat regions of the MLPCA corrected reflectance spectra. The variation, and information that has evidently been preserved in the MLPCA drift corrected spectra, was largely removed by derivative preprocessing as a result of the sharp low-frequency attenuation of these filters. **Figure 4.22** allows a closer inspection of a single set of replicate measurements corrected under these conditions. The MLPCA corrected replicates are virtually indistinguishable from one another and effectively overlap within resolution in the figure. The derivative preprocessed spectra, however, still exhibit some variance within the replicates.

The results of the cross-validation and calibration procedures are summarized in **Table 4.2** as *RMSECV*s and performance ratios, showing the performance of PCR, derivative PCR (under a variety of best-case filter conditions), and MLPCR. The performance ratio (*PR*) in these cases is defined as the relative performance of MLPCR to the other calibration methods, or

$$PR = \frac{RMSECV_{MLPCR}}{RMSECV_{other}} \tag{4.31}$$

The *PR* will exceed unity for calibration methods that demonstrate performance *superior* to that of MLPCR. For all three analytes of interest in the calibration set, MLPCR outperforms both the best PCR, and derivative PCR models in its predictions. In some cases, such as with component 2, some derivative PCR methods perform comparably to MLPCR. It must be kept in mind, however, that the expressed results are best-case scenarios for derivative filtering, and the result of extensive (and time-consuming) searches for optimal derivative PCR parameters.

Table 4.2 Summary of calibration performances for PCR, MLPCR and various forms of derivative preprocessing used in conjunction with PCR. 'Filter condition' (OD) refers to 'O' the polynomial order of the SG filter, and 'D' the first (1) or second (2) derivative. (LV: latent variables determined to be optimal, *PR*: performance ratio of *RMSECV* for MLPCR to the *RMSECV* of the method under inspection).

Calibration conditions		Results			
Component	**filter condition (OD)**	**filter width**	***RMSECV***	**LV**	***PR***
1	PCR	-	1.11	4	0.26
	deriv PCR - 1,1	3	0.82	4	0.35
	deriv PCR - 2,1	3	0.47	6	0.62
	deriv PCR - 2,2	7	0.44	6	0.66
	MLPCR	-	**0.29**	**7**	**1.00**
	MLPCR: ECV Block Average	9	0.33	9	0.88
2	PCR	-	0.95	6	0.29
	deriv PCR - 1,1	3	0.31	3	0.90
	deriv PCR - 2,1	5	0.32	5	0.88
	deriv PCR - 2,2	9	0.30	9	0.93
	MLPCR	-	**0.28**	**6**	**1.00**
	MLPCR: ECV Block Average	9	0.31	5	0.90
3	PCR	-	1.24	7	0.40
	deriv PCR - 1,1	7	0.66	7	0.74
	deriv PCR - 2,1	7	0.60	7	0.82
	deriv PCR - 2,2	9	0.55	9	0.89
	MLPCR	-	**0.49**	**7**	**1.00**
	MLPCR: ECV Block Average	9	0.45	9	1.09

Also shown in **Table 4.2** are the *RMSECV*'s for MLPCR using a smoothed error covariance matrix (denoted *MLPCR: ECV Block Average*). In this application, the smoothing operation resulted in little change in the MLPCR cross-validation error. This is unsurprising, however, since the error covariance structure for these calibration data is very prominent and appears to be well-estimated by pooling the estimated error covariance matrices. A visual inspection of the estimated error covariance matrix (**Figure 4.20**) confirms that it does appear to have a very high signal-to-noise ratio, implying that there is reasonable precision in the estimation.

4.5 Conclusions

The objective of the research outlined in this chapter was to investigate derivative preprocessing as a method of drift noise reduction in multivariate spectral data. This examination was carried out from the perspective that baseline drift can be characterized as correlated measurement errors, and that derivative filtering alleviates some drift noise by reducing the covariance terms in the error covariance matrices (*via* **Equation 4.10**). While this is often successful to some degree, derivative filters cannot be considered optimal, since the error covariance matrix can rarely be truly diagonalized by their application. In addition, the use of derivative filters substantially modifies the composition of the chemical signals in a fashion that is very difficult to predict *a priori*, making the effects on figures of merit in multivariate calibration largely unpredictable.

Derivative filters operate blindly in reducing drift noise and, therefore, must be chosen on a trial-and-error basis, but maximum likelihood PCA uses error covariance information to achieve simultaneous drift correction, and the maximum likelihood projection of the spectral data into a principal component space. It was shown that MLPCA is the 'optimal filter' from a drift reduction perspective, since MLPCA uses error covariance information to diagonalize the error covariance matrix, thereby eliminating drift-noise. The regression counterpart to MLPCA, MLPCR, is consequently an optimal calibration method to use when drift noise plagues the acquired data.

Baseline drift poses a significant threat to the precision and accuracy of many multivariate calibration methods. Derivative preprocessing has been widely employed to combat this problem in the past, but since it is suboptimal in terms of drift correction, its application requires time-consuming searches for the best filter characteristics for a given application. Unfortunately, the spectral interpretability also suffers upon differentiation. In this work, MLPCR was consistently found to perform as well as, or better than, derivative PCR when reasonable estimates of the error covariance structure were available. It is

therefore recommended that, provided error covariance information is obtainable, MLPCR be used as a calibration method for data corrupted by baseline drift.

5. Future Directions and Concluding Remarks

5.1 Future Avenues of Investigation

Drift noise, defined in **Chapter 4** as "any undesirable fluctuation in instrument response which results in correlated measurement errors", is becoming increasingly discussed in the literature regarding multivariate chemical analysis. This is due in part to the celebrity of instrumental techniques, such as NIR reflectance, which have features that are a tremendous asset in rapid sample analysis. Unfortunately these methods often have severe problems with drift noise arising principally from sample-specific effects, which are broadly referred to as scatter. A variety of methods have been proposed to combat these problematic fluctuations, some of which were listed in previous chapters [36-40]. Stark *et al.* provide a good review of the more common approaches [50]. Among the most popular of these correction methods is multiplicative scatter correction (MSC) [37], which is now often referred to as multiplicative signal correction due to its demonstrated utility in also correcting non-scatter related phenomena.

Multiplicative signal correction was first introduced in the chemistry literature in 1985 by Geladi *et al.* [37] as an extension of Norris's empirical ratioing method [36] with its purpose being to correct for the substantial variation due to light scattering that can often occur in reflectance measurements (sometimes as high as 99% [51]). In reflectance spectroscopy, the scatter noise is particularly difficult to deal with since it is typically considered a mixture of additive (*e.g.*, baseline offset) and multiplicative (*e.g.*, path length and light scattering level) effects. The most problematic of the two, multiplicative noise, arises from particle size variations within the sample, and the lack of pathlength control afforded in reflectance geometries. Each photon essentially experiences a different pathlength from source to detector due to the stochastic nature of the

internal reflections of sample particulates. As a result, a *distribution* of pathlengths contribute to the instrumental response, with the mean photonic pathlength often referred to as the *effective* pathlength. Since the pathlength is integral in the relation of concentration to absorbance, some correction must be applied to standardize these photon-level pathlength variations.

MSC was devised, in the author's own words [37], to preprocess the spectral data such that "all samples appear to have the same scatter level as [an] 'ideal' [sample]." Since 'ideal' samples tend to be lacking in practice, the mean spectrum from a collection of sample spectra is most often used. The correction method is meant to reduce the contributions of both baseline offsets and multiplicative noise by least-squares correction of the mixture spectra to the mean spectrum.

The term multiplicative noise as it is used here is *not* strictly what many have classically referred to as proportional/multiplicative noise. The term proportional noise is commonly taken to imply noise whose variance is proportional to signal, while multiplicative noise arising from the sample-specific effects discussed above is both proportional, *and* highly correlated since it is presumed to stem from the following model:

$$\mathbf{x}_i = \mathbf{x}_i^\circ + \mathbf{e}_i \tag{5.1}$$

where

$$\mathbf{e}_i = a_i\mathbf{1} + b_i\mathbf{x}_i^\circ + \boldsymbol{\varepsilon}_i \tag{5.2}$$

In this representation, a_i is an (additive) offset noise term (as in **Equation 4.2**), b_i is a (multiplicative) noise term which introduces a multiple of the true signal vector in the observed noise, and $\boldsymbol{\varepsilon}_i$ is a vector of noise terms arising from neither of the previous causes. If the scatter is allowed to be wavelength-dependent, then an additional term must be incorporated in **Equation 5.2**, but this wavelength-dependent scatter is often ignored. The geometric implications of such a noise structure are shown in **Figure 5.1**.

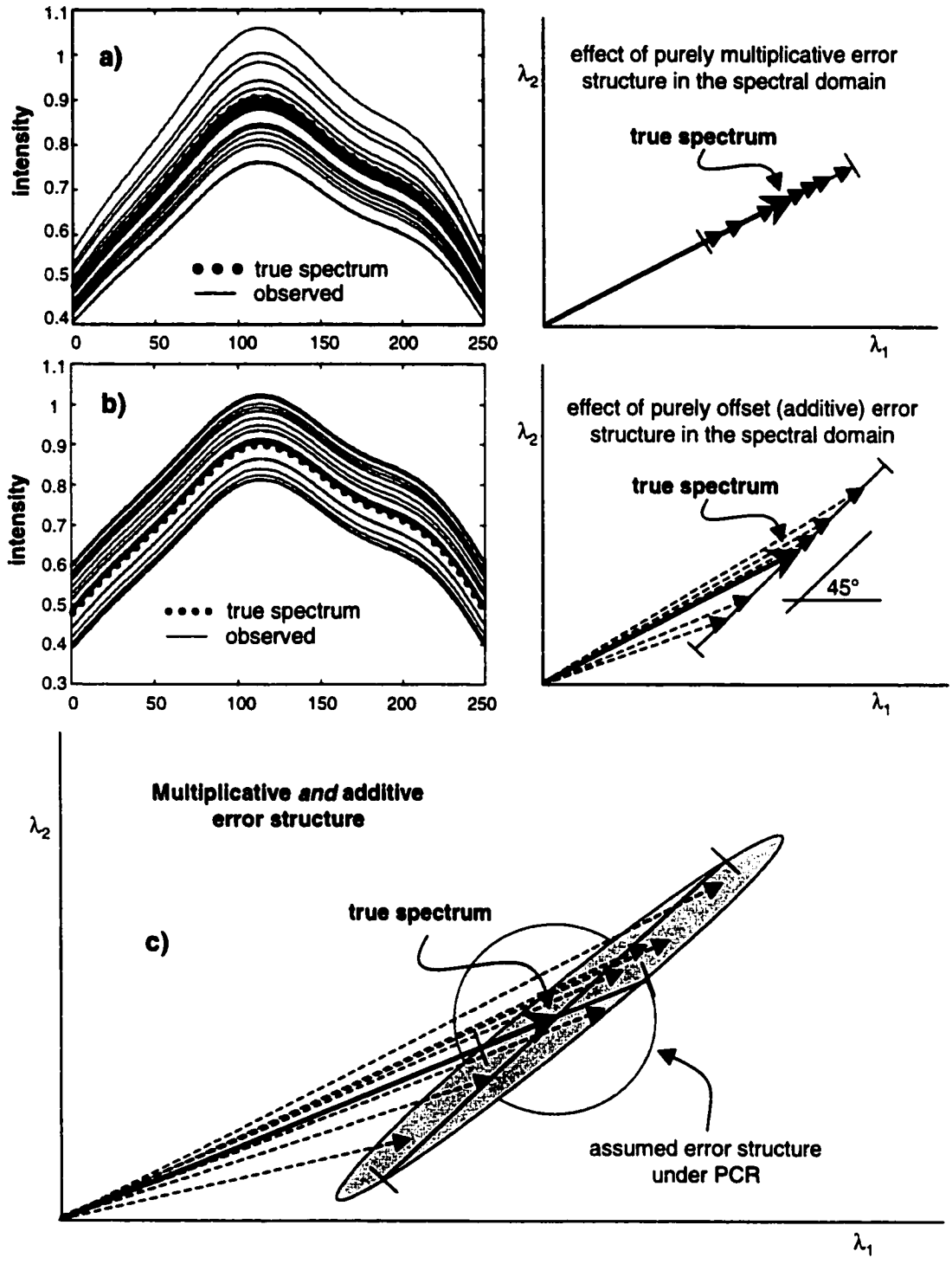


Figure 5.1 **a)** The effect of multiplicative error in the spectral domain, **b)** the effect of offset errors shown in a complimentary fashion, and **c)** the combined effects illustrated geometrically.

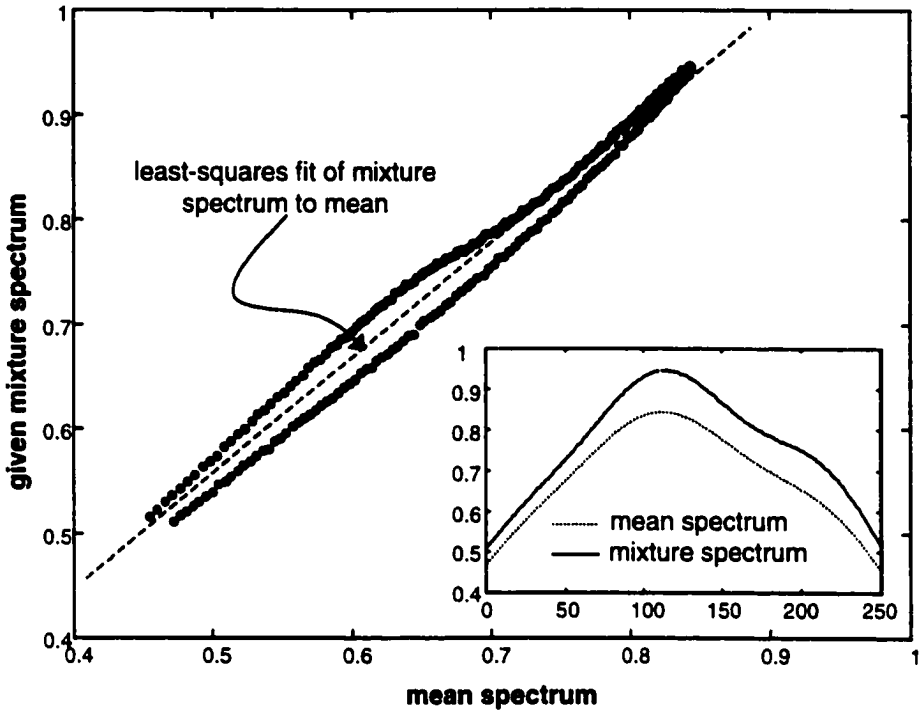


Figure 5.2 An illustration of the least-squares fitting procedure used to correct each of the mixture spectra in a calibration or prediction set to a mean scattering level. The actual mixture spectrum, and the mean spectrum are shown in the inset.

MSC assumes that the following model accurately describes the *i*th spectrum in a set of calibration data.

$$\mathbf{x}_i = a_i \mathbf{1} + b_i \bar{\mathbf{x}} + \mathbf{i}_i \tag{5.3}$$

In **Equation 5.3**, a_i and b_i are the additive and multiplicative factors from **Equation 5.2**, $\bar{\mathbf{x}}$ is the mean spectrum from the set samples measured, and $\mathbf{i}_i$ is an information vector, which represents the chemical variation not accounted for by offset variation, or multiples of the mean spectrum. The correction is achieved by regressing each $\mathbf{x}_i$ onto the mean vector to estimate the additive and multiplicative factors $\hat{a}_i$ and $\hat{b}_i$, as illustrated in **Figure 5.2**. To remove these undesirable effects from the observed spectral vector, $\mathbf{x}_i$ is adjusted to the scatter corrected $\hat{\mathbf{x}}_i$ in the following fashion.

$$\hat{\mathbf{x}}_i = \frac{(\mathbf{x}_i - \hat{a}_i)}{\hat{b}_i} \tag{5.4}$$

which is equivalent to

$$\hat{\mathbf{x}}_i = \bar{\mathbf{x}} + \frac{\hat{\mathbf{i}}_i}{\hat{b}_i} \tag{5.5}$$

An illustration of the effect of this correction on spectral data is given in **Figure 5.3** for simulated spectral data with additive, multiplicative and white noise, and a geometric interpretation is also offered in **Figure 5.4**.

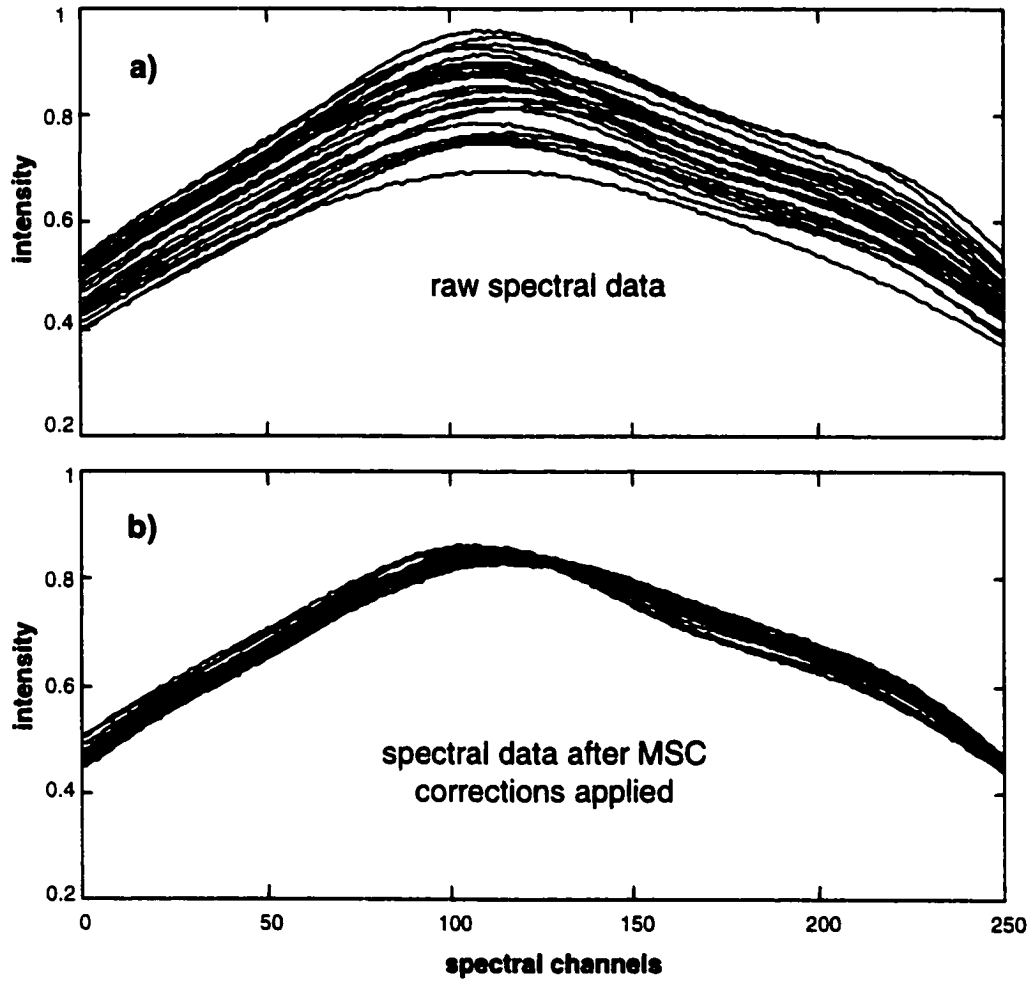


Figure 5.3 **a)** Simulated raw spectral data (25 spectra) corrupted with additive, multiplicative, and white noise. **b)** Those same spectra after MSC application.

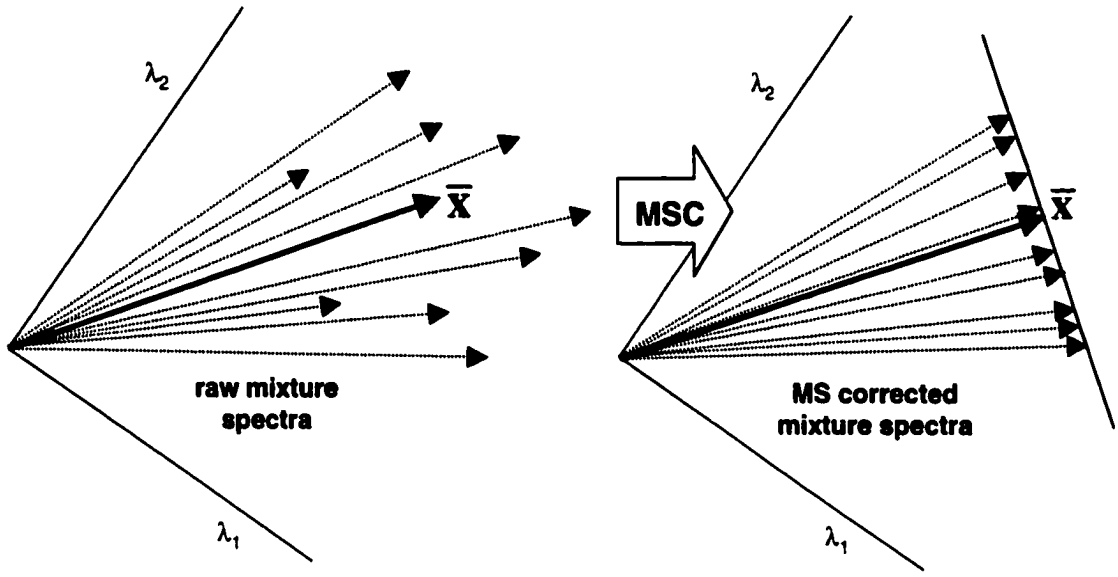


Figure 5.4 Geometric interpretation of MSC in two dimensions.

With MSC becoming increasingly popular as a method for drift and scatter reduction, it would be useful to examine MSC in the same theoretical light as previous investigations of digital filtering methods. Some preliminary studies have been conducted, and the results will be briefly summarized here.

The experimental NIR reflectance data discussed in **Chapter 4** are heavily corrupted by drift noise, as can be observed from the error covariance matrix exhibited in **Figure 4.20**. Provided **Equation 5.2** is valid and a and b are uncorrelated, it can be shown that the error covariance structure for a given spectral vector, $\mathbf{x}^{o}$, corrupted by additive and multiplicative scatter effects, as well as other sources of white noise is

$$\boldsymbol{\Sigma}_{tot} = \boldsymbol{\Sigma}_{a} + \boldsymbol{\Sigma}_{b} + \sigma^{2}\mathbf{I}_{n} \tag{5.6}$$

where $\boldsymbol{\Sigma}_{a}$ is the offset contribution, which, from **Equation 4.4**, is

$$\boldsymbol{\Sigma}_{a} = \sigma_{a}^{2}\mathbf{1}\mathbf{1}^{\mathrm{T}} \tag{5.7}$$

and the multiplicative contribution, Σ_b, is the outer product of the true spectrum with itself, or

$$\Sigma_b = \sigma_b^2\left(\mathbf{x}^o\mathbf{x}^{oT}\right) \tag{5.8}$$

An inspection of the NIR reflectance error covariance structure reveals that multiplicative scattering effects are extremely dominant, as the structure of the error covariance matrix largely mirrors that of the spectral shapes, and offset noise, if present, contributes little to variation between sample replicates.

The performance of MSC in correcting for multiplicative scattering effects in these NIR reflectance data was compared to the performance of MLPCR. The MSC preprocessed NIR reflectance data are shown in **Figure 5.5**. The variation between samples is significantly reduced relative to the untreated spectral data, although, as with derivative filters, this isn't a sure sign of performance enhancement. The cross-validation results are summarized in **Table 5.1**.

Table 5.1 Summary of results for cross-validation studies of PCR and MLPCR models generated from unprocessed ABS polymer data, and from MSC corrected data. (LV = number of latent variables, *RMSECV* = root mean-squared error of cross-validation)

		Raw data		MSC data	
Component	**Model Type**	**LV**	***RMSECV***	**LV**	***RMSECV***
1	**PCR**	4	1.11	3	0.92
	MLPCR	7	0.29	3	0.43
2	**PCR**	6	0.95	8	0.55
	MLPCR	6	0.28	4	0.26
3	**PCR**	7	1.24	4	0.88
	MLPCR	7	0.49	4	0.41

Table 5.1 lists the *RMSECV*'s for the raw spectral data using both PCR and MLPCR (identical to the results given in **Chapter 4**). The spectral data were also treated using MSC and submitted for calibration and cross-validation using

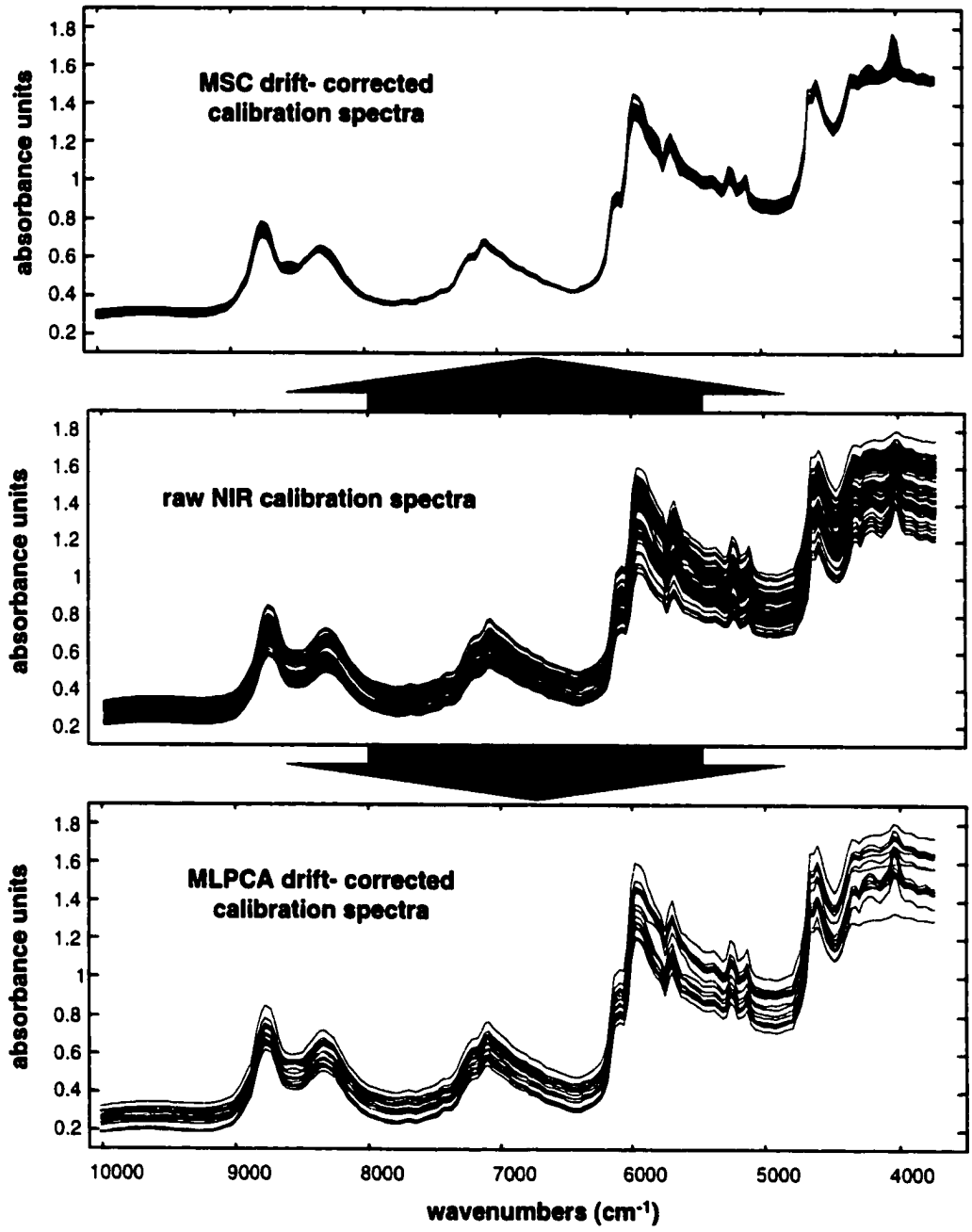


Figure 5.5 A visual comparison of the MSC treated NIR data, to the raw spectral data, and the previously discussed MLPCA correction.

PCR. The MSC treatment improves the performance of PCR with all three components; however, the improvement does not reach the level of performance observed for MLPCR with the raw spectral data. For all three components, MLPCR (with the raw spectra) performed at least twice as well as MSC-PCR. MSC was also used as a preprocessing method with MLPCR to see if MLPCR could benefit from such standardizations. Although marginal improvements can be seen for components 2 and 3, no significant changes in MLPCR performance can be observed in this case. The reader may also notice that the number of latent variables required for MLPCR appears greater than the MSC treated methods in **Table 5.1**. While this may be the case here, it is not generally extensible, and it is necessary to reiterate that the number of latent variables was determined in these studies from the *absolute* minimum of the *RMSECV*, which can potentially give spurious results for the 'optimum' number of latent variables.

Although MSC-PCR does not appear to perform as well as MLPCR in scatter correction and handling, it does raise an interesting possibility for future avenues of investigation. If the error model in **Equation 5.6** is valid as written, then it may be feasible to use hard error models with NIR data in MLPCA and MLPCR, eliminating the need for replicate measurements. If multiplicative effects dominate the error covariance structure, the mixture spectra as obtained from experiment could be used as surrogates for **Equation 5.8**. While the absolute values of σ_a^2 and σ_b^2 are unimportant (MLPCA and MLPCR are invariant to the scale of the error covariance matrix), their relative magnitudes *are* liable to have an influence on the validity of the model. An anticipated complication would therefore involve establishing the proper ratio of offset variance-covariance to multiplicative variance-covariance, although it is likely that these parameters could be roughly estimated using MSC itself. This possibility was investigated using several data sets available on the web for download [52]. Offset contributions were typically negligible compared to the multiplicative scattering effects, and so it was assumed that the error covariance structure is entirely multiplicative in nature. Preliminary investigations involving these online data

sets have been conducted, although MLPCR (using this hard error model) has not been found to perform significantly better than PCR in most cases. The ABS polymer data was also reexamined using this MLPCR approach. In this instance, the offset noise was *not* assumed negligible; the offset and multiplicative variance terms where estimated (via MSC) to be $\sigma_a^2 = 0.0008$, and $\sigma_b^2 = 0.0060$. Using these parameters, **Equation 5.6** was employed to estimate the error covariance structure of the data. The mean sample spectrum was used in **Equation 5.8** in the absence of knowledge of the true spectra, and equal row covariance structure was assumed for MLPCR. Results for MLPCR were found to be, at best, only marginally better than PCR. While the reasons for this require further investigation, it is likely that the MSC-estimated parameters σ_a^2 and σ_b^2 are not accurate enough for MLPCR calibration performance improvement over standard PCR. Further investigations may yield insights into the hard error models which enhance MLPCR performance, and allow precise calibration and prediction without the requirement of replicate measurements.

It is likely that MLPCR using NIR data would be amenable to the simplification used throughout this work—the assumption of equal row error covariance structure. Provided the mixture spectra are reasonably similar in shape, **Equation 5.8** suggests that the error covariance structure for the samples would be highly similar as well. In situations in which this was deemed not to be the case, individual projections of the mixture spectra onto the MLPCA estimated space could still be achieved with relative ease using different error covariance matrices for each sample.

While MSC is quite a commonly used scatter correction method, several other recently introduced techniques are gaining in popularity. Methods such as the standard normal variate (SNV) transformation [40], and orthogonal signal correction [53] are preprocessing methods which are touted to substantially reduce the undesirable influence of drift and scatter effects, and consequently improve calibration performance. The field would most certainly benefit if theoretical studies, similar to those outlined in this work, were initiated on these

sorts of methods. Other more complex preprocessing methods such as Fourier filtering and Wavelet denoising and smoothing could also be examined from the calibration perspective.

5.2 Conclusions

The ideal analytical instrument would be infinitely precise, accurate, free, and furnish results instantaneously—a whimsical notion not likely to materialize in the near future; however, the field of analytical chemistry is evolving at a tremendous rate and producing simpler, faster, and cheaper analytical techniques on a daily basis. Chemometric methods have played an invaluable role in realizing these advances, since the analytical cost associated with rapid and inexpensive methods—*precision*—can often be recovered using mathematical methods which make full use of multichannel data. This recovery is somewhat hampered, however, by the increasing complexity of both the chemical and physical properties of analytical samples which can often introduce deleterious artifacts in the measurement data which are particularly troublesome for conventional chemometric methods. In an attempt to alleviate these undesirable effects, it is common practice to employ preprocessing methods which, optimistically, condition the measurement data to a form which is better suited for use in classical chemometric methods. Surprisingly, the theoretical and practical consequences of these pretreatment methods in multivariate analyses are not well understood, or even worse, misunderstood. A sound theoretical understanding of these preprocessing methods will certainly engender more rational approaches to preprocessing, and invariably lead to improved methods for handling injurious artifacts in multivariate analysis. It was with these points in mind that this research was conducted.

In **Chapter 3** the use of digital smoothing filters in multivariate calibration was examined. These filters are typically applied with aspirations of reducing the noise level of the data, and thereby reducing the prediction error of calibration models using these data. A theoretical examination of symmetric digital

smoothing filters was undertaken from the perspective of the net analyte signal. This divorced the investigation from the calibration method employed and allowed the theoretical derivation of the multivariate signal-to-noise ratio for the filtered and unfiltered data. Under the assumptions used (*iid* errors in the original data and negligible error in the calibration step), it was found that no enhancement in the multivariate *S/N* ratio can be expected by digital smoothing with a symmetric filter and, therefore, no enhancement in *RMSEP* could be anticipated. Gains in performance were sometimes observed in the practical experimental evidence presented, a result that was shown to arise from improvements in the estimation of the calibration model by smoothing. These benefits were consistently marginal.

In **Chapter 4** the problem of drift correction was explored, again from a theoretical perspective. Derivative filters, one of the more popular drift correction methods, were discussed and shown only to approach ideal drift correction. Optimal drift correction (complete elimination of drift in the limit) could only be achieved if the derivative filter matrix satisfied **Equation 4.11**, a condition that is highly unlikely in practice. An optimal filter was derived which provides optimal drift correction provided knowledge of the error covariance structure is available. This optimal filtering method was shown to be a special case of maximum likelihood PCA, and its use in calibration procedures constitutes a special case of maximum likelihood PCR. The premier benefit of this approach, arguably, is that it is a *direct* and *pointed* mode of drift correction, which consistently performs as well as, or better than the *indirect* derivative methods which often require time-consuming parameter optimizations. From a figures-of-merit perspective it was shown that the effect of differentiation was unpredictable, and highly dependent on the characteristics of both the measurement errors and the pure-component spectra. Studies with experimental data (NIR reflectance) confirmed the results of the simulations.

On the whole, this body of research has contributed substantially to the theoretical knowledge of two of the more widely applied advanced preprocessing

methods for multivariate calibration, and demonstrated that the theoretical comprehension gained by such studies leads to more cogent strategies for preprocessing. These insightful investigations have hopefully filled-in a substantial portion of the knowledge gap that currently exists regarding the theoretical implications of such preprocessing techniques in multivariate calibration. As a result, more direct and rational approaches to preprocessing can be adopted, avoiding the costly trial and error approaches of days past. Similar advances are likely to be made if other preprocessing methods are examined in the same light; this research will hopefully act as a springboard for future investigations.

References

1 S. Wold, *Kem. Tidskr.*, **3**, 34 (1972)

2 P. Sheperd, *J. Chemometrics.*, **1**, 3 (1987)

3 K. Esbensen and P. Geladi, *J. Chemometrics*, **4**, 389 (1990)

4 G. M. Hieftje, *Anal. Chem.*, 72, 309A (200)

5 P. Geladi, D. MacDougall and H. Martens, *Appl. Spectrosc.*, **39**, 491 (1985)

6 A.L.Cauchy, *OEvres*, **IX**, 172 (1829)

7 K. Pearson, *Phil. Mag.*, **6**, 559 (1901)

8 R.J. Adcock, *The Analyst*, **5**, 53 (1878)

9 P. Paatero and U. Tapper, *Chemom. Intell. Lab. Syst.*, **18**, 183 (1993)

10 P.D. Wentzell, D.T. Andrews, D.C. Hamilton, K. Faber, and B.R. Kowalski, *J. Chemometrics*, **11**, 339 (1997)

11 P.D. Wentzell, D.T. Andrews and B.R. Kowalski, *Anal. Chem.*, **69**, 2299 (1997)

12 P.D. Wentzell, and M.T. Lohnes, *Chemom. Intell. Lab. Syst.*, **45**, 65 (1999)

13 E. Sanchez, and B.R. Kowalski, *J. Chemometrics*, **2**, 247 (1988)

14 A. Lorber, *Anal. Chem.*, **58**, 1167-1172 (1986)

15 K.S. Booksh and B.R. Kowalski, *Anal. Chem.*, **66**, 782A (1994)

16 C.D. Brown and P.D. Wentzell, *J. Chemometrics*, **13**, 133 (1999)

17 A. Lorber, K. Faber, and B.R. Kowalski, *Anal. Chem.*, **69**, 1620 (1997)

18 K. Faber, A. Lorber, and B.R. Kowalski, *J. Chemometrics*, **11**, 419 (1997)

19 P.D. Wentzell and C.D. Brown, *Encyclopedia of Analytical Chemistry*, R. A. Meyers, Ed., John Wiley & Sons - *in press*

20 S. N. Deming, J.A. Palasota and J.M. Nocerino, *J. Chemometrics*, **7**, 393 (1993)

21 R.W. Hamming, *Digital Filters*, 2nd Edition, Prentice-Hall: Englewood Cliffs, NJ (1983)

22 A. Savitzky and M.J.E. Golay, *Anal. Chem.*, **36**, 1627 (1964)

23 A. Proctor, P.M.A. Sherwood, *Anal. Chem.*, **52**, 2315 (1980)

24 R.A. Leach, C.A. Carter and J.M. Harris, *Anal. Chem.*, **56**, 2304 (1984)

25 P.D. Wentzell, T. P. Doherty and S.R. Crouch, *Anal. Chem.*, **59**, 367 (1987)

26 C.G. Enke and T. A. Nieman, *Anal. Chem.*, **48**, 705A (1976)

27 a. G.M. Hieftje, *Anal. Chem.*, **44**, 69A (1972)
b. D. Binkley and R. Dessy, *J. Chem. Educ.*, **56**, 148 (1979)
c. M.U.A. Bromba and H. Ziegler, *Anal. Chem.*, **53**, 1583 (1981)

28 a. T.H. Edwards and P.D. Wilson, *Appl. Spectrosc.*, **28**, 541 (1974)
b. R. J. Larivee and S.D. Brown, *Anal. Chem.*, **64**, 2057 (1992)

29 K. Faber and B.R. Kowalski, *J. Chemometrics*, **11**, 181 (1997)

30 K.R. Betty and G. Horlick, *Anal. Chem.*, **49**, 351 (1977)

31 O. E. de Noord, *Chemom. Intell. Lab. Syst.*, **23**, 65 (1994)

32 N. M. Faber, *Anal. Chem.*, **71**, 557 (1999)

33 M. B. Seasholtz and B. R. Kowalski, *Anal. Chim. Acta*, **277**, 165 (1993)

34 C. D. Brown and P. D. Wentzell, *in preparation*

35 T. Hirschfeld, D. Honigs, and G. Hieftje, *Appl. Spectrosc.*, **39**, 430 (1985)

36 a. K. H. Norris and R. F. Barnes, in *Proceedings, 1st International Symposium on Feed Composition, Animal Nutrient Requirements and Computerization of Diets*, International Feedstuffs Institute: Utah State University, Logan, Utah (1977)
b. K. H. Norris, in *Food Research and Data Analysis*, H. Martens and H. Russwurm, Jr., Eds., Applied Science: London (1983)

37 P. Geladi, D. MacDougall, and H. Martens, *Appl. Spectrosc.*, **39**, 491 (1985)

38 a. I. S. Helland, T. Næs, T. Isaksson, *Chemom. Intell. Lab. Syst.*, 29, 233 (1995)
b. H. Martens and E. Stark, *J. Pharm. Biomed. Anal.*, **9**, 625 (1991)
c. T. Isaksson and B. R. Kowalski, *Appl. Spectrosc.*, **47**, 702 (1993)
d. J. L. Ilari, H. Martens, and T. Isaksson, *Appl. Spectrosc.*, **42**, 722 (1988)
e. S. Schönkopf, H. Martens, and B. Alsberg, in *Making Light Work: advances in NIR spectroscopy*, I. A. Cowe, and I. Murray, Eds., VCH: New York, NY (1992)

39 I. Murray and P. A. Hall, *Anal. Proc.*, **20**, 75 (1983)

40 R. J. Barnes, M. S. Dhanoa and S. J. Lister, *Appl. Spectrosc.*, **43**, 772 (1989)

41 a. V. J. Hammond and W. C. Price, *J. Opt. Soc. Am.*, **43**, 924 (1953)
b. E. Tannenbauer, P. B. Merkel and W. H. Hammill, *J. Phys. Chem.*, **21**, 311 (1953)
c. J. D. Morrison, *J. Chem. Phys.*, **21**, 1767 (1953)

42 a. T. C. O'Haver and G. L. Green, *Anal. Chem.*, **48**, 312 (1976)
b. J. E. Cahill, *Am. Lab.*, **11**, 79 (1979)
c. T. C. O'Haver and T. Begley, *Anal. Chem.*, **53**, 1876 (1981)

43 a. T. R. Griffiths, K. King, H. V. St. A. Hubbard, M. -J. Shwing-Weill, and J. Meullemeestre, *Anal. Chim. Acta*, **143**, 163 (1982)
b. D. G. Cameron and D. J. Moffatt, *Anal. Chem.*, **41**, 539 (1987)
c. W. F. Maddams and W. L. Mead, *Spectrochim. Acta Part A*, **38**, 437 (1982)

44 L. L. Juhl and J. H. Kalivas, *Anal. Chim. Acta*, **207**, 125 (1988)

45 C. D. Brown, L. Vega-Montoto and P. D. Wentzell, *Appl. Spectrosc.*, **57**, *in press*

46 J. Steiner, Y. Termonia and J. Deltour, *Anal. Chem.*, **44**, 1906 (1972)

47 R. J. Pell and B. R. Kowalski, J. Chemometrics, 5, 375 (1991)

48 Industrial co-investigators on the project: Drs. R. Pell, and M. B. Seasholtz, Analytical Division, Dow Chemical Company, Midland, MI. Instrumental analysis was carried out on the resin samples by Dave Albers, also of Dow Chemical Company.

49 H. Martens and T. Næs, *Multivariate Calibration*, John Wiley & Sons: New York, NY (1989)

50 E. Stark, K. Luchter, and M. Margoshes, *Appl. Spec. Rev.*, **22**, 335 (1986)

51 E. Stark, in *Analytical Applications of Spectroscopy*, C. S. Creaser and A. M. C. Davies, Eds., Royal Society of Chemistry: London (1988), p. 28

52 Corn data set available from Eigenvector Research, Inc. (www.eigenvector.com); grass data set available from W. F. McClure (mcclure@eos.ncsu.edu), temperature fluctuation data available from A. Smilde (asmilde@its.chem.uva.nl)

53 S. Wold, H. Antii, F. Lindgren, J. Ohman, *Chemom. Intell. Lab. Syst.*, **44**, 175 (1998)