THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DECONVOLUTION AND QUANTITATION OF SEVERELY OVERLAPPED CHROMATOGRAPHIC PEAKS USING HIGH PERFORMANCE LIQUID CHROMATOGRAPHY WITH MULTI-CHANNEL ELECTROCHEMICAL DETECTION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

By

DIANA RUTH HAMILTON

Norman, Oklahoma

2005

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/876

UMI Number: 3164566

Copyright 2005 by
Hamilton, Diana Ruth

All rights reserved.



UMI Microform 3164566
Copyright 2005 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECONVOLUTION AND QUANTITATION OF SEVERELY OVERLAPPED CHROMATOGRAPHIC PEAKS USING HIGH PERFORMANCE LIQUID CHROMATOGRAPHY WITH MULTI-CHANNEL ELECTROCHEMICAL DETECTION

A DISSERTATION

APPROVED FOR THE DEPARTMENT OF CHEMISTRY

AND BIOCHEMISTRY

By

C. Leroy Blank

R. W. Taylor

R. A. White

D. K. Watson

G. Richter-Addo

Copyright 2005, Diana R. Hamilton

University of Oklahoma, Norman OK

Acknowledgements

There are many people to acknowledge and to be thankful for in my life; my parents, Augustus and Shirley, instilled a deep sense of pride in the accomplishment of a difficult goal. I wish to thank John Nail for the encouragement (and pushing) he gave me when I really needed it.

Dedication

To Mom and Dad: Thanks for everything.

Table of Contents

Abstract xxii

Chapter 1 — Introduction 1

The Problem 1

Introduction to Chromatography and Electrochemistry 1

Separation: Chromatography 3

Electrochemistry 5

Flow and Electrochemistry in Thin-Layer Flow-Through Cells 9

Electrode Material 10

Mobile Phase 11

The Detector: Thin-Layer Cells 12

The Recorded Data and Resultant Peak Shapes 14

Resolution 15

Quantitation Issues 19

Internal Standards 21

Deconvolution of Overlapping Peaks 21

Chromatographic Peak Shape Functions 24

Haarhoff – Van der Linde (HVL) 31

Exponentially Modified Gaussian (EMG) 32

Giddings Equation 33

Half-Gaussian Modified Gaussian (GMG) 34

4 Parameter EMG-GMG hybrid (GEMG4) 35
5 Parameter EMG-GMG hybrid (GEMG5) 35
EMG + GMG 36
Bi-Gaussian (BG) 36
Fraser-Suzuki (FS) 37
Log-Normal 38
Cauchy-Gaussian 38
Chesler-Cram 39
Principal Components - Background 40
Factor Analysis 48
Error 52
Residual Standard Deviation (RSD) 53
Root Mean Square (RMS) 53
Average error criterion 53
Chi-squared criterion 54
Distribution of misfit 54
Cumulative percent variance 54
Average eigenvalue 55
Scree test 55
Exner function 55
Imbedded error function 55
Factor indicator function 56
Malinowski F-test 56

Cross-validation 57
Principal Component Regression. 58
Partial Least Squares. 59
Development of the Cramer's Rule Approach 63
Chapter 2 — Experimental 72
NEUBA® Chromatographic System 72
Materials Used 76
Column 77
Mobile Phase 78
Standards 78
Epi 78
Dopa 79
Dopamine 79
Mixture Sets 79
Experimental Conditions 81
Experimental Data 81
Preprocessing of the data 82
Determining baseline correction 83
Time Corrections to the Data 84
Determination of time t_M 86
Adjustment of the data for the non-retained species 88
Curve Fitting (time-signal domain) 96
Determination of the chromatographic peak shape 96

Fitting the data to the model 96
Application of the model and obtaining the analyte amounts 96
Using Chemometric Methods 97
Cramer's Rule approach in the signal-potential domain 97
Fitting the Data to Electrochemical "Peak" Shape 97
Determining the amounts of the analytes 98
Chapter 3 — Curve Fitting Results and Discussion 100
Channel 1 101
Results 102
Mix 1 103
Mix 2 105
Mix 3 108
Mix 4 110
Mix 5 112
Discussion 113
Channel 2 115
Results 115
Mix 1 116
Mix 2 118
Mix 3 120
Mix 4 123
Mix 5 124
Discussion 127

Channel 3 127
Results 128
Mix 1 129
Mix 2 131
Mix 3 133
Mix 4 136
Mix 5 138
Discussion 140
Channel 4 141
Results 141
Mix 1 142
Mix 2 144
Mix 3 146
Mix 4 148
Mix 5 151
Discussion 153
Curve-Fitting Methods: Summary and Conclusions 153
Chapter 4 — 3-Dimensional Chemometric Methods: Results and Discussion 157
Principal Component Analysis and Associated Chemometric Analyses for the 481 Time Data Points 162
Results 173

Principal Component Analysis and Associated Chemometric Analyses for the Reduced, 121 Time Data Points 175
Results 184
Principal Component Analysis and Associated Chemometric Analyses for Very Reduced, 21 Time Data Points 189
Results 195
Cramer's Rule Approach 200
Results 202
Chapter 5 — Conclusions 205
References 219

Table of Equations

Figure 1-1: Schematic of a liquid chromatographic instrument 2

Figure 1-2. Representative hydrodynamic voltammograms for compounds A and B..... 8

Figure 1-3: A typical cross-flow thin-layer electrode cell..... 13

Figure 1-4. The detector cell used with NEUBA®..... 14

Figure 1-5: Peak types. 15

Figure 1-6: A typical chromatogram, and determination of the retention times of a nonretained species, t_M, and a retained species, t_R..... 17

Figure 1-7: Determination of the width at half height, and the width of the peak at the base. 18

Figure 1-8: Resolution of two peaks 19

Figure 2-1. Flow diagram for the NEUBA® system. 74

Figure 2-2: Representative Injection..... 85

Figure 2-3: Solvent Front Pattern..... 86

Figure 2-4: Retention time of Epi as a function of time corrections..... 88

Figure 2-5: Retention time of Dopa as a function of retention time corrections 89

Figure 2-6: Retention time of Dopamine as a function of time corrections 90

Figure 2-7: Overlap of Epi and Dopa..... 95

Figure 3-1: SSR results for Channel 1..... 102

Figure 3-2: Percent Difference for Channel 1, Mix 1 103

Figure 3-3: %RSD for Channel 1, Mix 1 105

Figure 3-4: Percent Difference for Channel 1, Mix 2 106
Figure 3-5: %RSD for Channel 1, Mix 2 107
Figure 3-6: Percent Difference for Channel 1, Mix 3 108
Figure 3-7: %RSD for Channel 1, Mix 3 109
Figure 3-8: Percent Difference for Channel 1, Mix 4 110
Figure 3-9: %RSD for Channel 1, Mix 4 111
Figure 3-10: Percent Difference for Channel 1, Mix 5 112
Figure 3-11: %RSD for Dopa, Channel 1, Mix 5 113
Figure 3-12: SSR results for Channel 2..... 116
Figure 3-13: Percent Difference for Channel 2, Mix 1 117
Figure 3-14: %RSD for Channel 2, Mix 1..... 118
Figure 3-15: Percent Difference Channel 2, Mix 2..... 119
Figure 3-16: %RSD for Channel 2, Mix 2..... 120
Figure 3-17: Percent Difference for Dopa, Channel 2, Mix 3 121
Figure 3-18: %RSD for Dopa, Channel 2, Mix 3 122
Figure 3-19: Percent Difference for Dopa, Channel 2, Mix 4 123
Figure 3-20: %RSD for Dopa, Channel 2, Mix 4 124
Figure 3-21: Percent Difference for Dopa, Channel 2, Mix 5 125
Figure 3-22: %RSD for Dopa, Channel 2, Mix 5 126
Figure 3-23: SSR results for Channel 3..... 128
Figure 3-24: Percent Difference for Dopa, Channel 3, Mix 1 129
Figure 3-25: %RSD for Dopa, Channel 3, Mix 1 130
Figure 3-26: Percent Difference for Dopa, Channel 3, Mix 2 132

Figure 3-27: %RSD for Dopa, Channel 3, Mix 2 133
Figure 3-28: Percent Difference for Channel 3, Mix 3 134
Figure 3-29: %RSD for Channel 3, Mix 3..... 135
Figure 3-30: Percent Difference for Channel 3, Mix 4 136
Figure 3-31: %RSD for Channel 3, Mix 4..... 137
Figure 3-32: Percent Difference for Channel 3, Mix 5 138
Figure 3-33: %RSD for Channel 3, Mix 5..... 139
Figure 3-34: SSR results for Channel 4..... 142
Figure 3-35: Percent Difference for Channel 4, Mix 1 143
Figure 3-36: %RSD for Channel 4, Mix 1..... 144
Figure 3-37: Percent Difference for Channel 4, Mix 2 145
Figure 3-38: %RSD for Channel 4, Mix 2..... 146
Figure 3-39: Percent Difference for Channel 4, Mix 3 147
Figure 3-40: %RSD for Channel 4, Mix 3..... 148
Figure 3-41: Percent Difference for Channel 4, Mix 4 149
Figure 3-42: %RSD for Channel 4, Mix 4..... 150
Figure 3-43: Percent Difference for Dopa, Channel 4, Mix 5 151
Figure 3-44: %RSD for Channel 4, Mix 5..... 152
Figure 4-1: A three-dimensional matrix 158
Figure 4-2: Scree plot for the original data set 163
Figure 4-3: Scores plot for original data set, PC1 166
Figure 4-4: Loadings plot for original data set, PC1 167
Figure 4-5: Scores plot for original data set, PC2 168

Figure 4-6: Loadings plot for original data set, PC2 169
Figure 4-7: Scores plot for original data set, PC3 170
Figure 4-8: Loading plot for original data set, PC3 171
Figure 4-9: Scores plot for original data set, PC4 172
Figure 4-10: Loadings plot for original data set, PC4 173
Figure 4-11: Scree plot for the reduced data set 176
Figure 4-12: Scores plot for PC1, reduced data set 178
Figure 4-13: Loadings plot for PC1, reduced data set 179
Figure 4-14: Scores plot for PC2, reduced data set 180
Figure 4-15: Loadings plot for PC2, reduced data set 180
Figure 4-16: Scores plot for PC3, reduced data set 181
Figure 4-17: Loadings plot for PC3, reduced data set 182
Figure 4-18: Scores plot for PC4, reduced data set 183
Figure 4-19: Loadings plot for PC4, reduced data set 183
Figure 4-20: Scree plot for the very reduced data set 189
Figure 4-21: Scores plot for PC1, reduced data set 191
Figure 4-22: Loadings plot for PC1, reduced data set 191
Figure 4-23: Scores plot for PC2, reduced data set 192
Figure 4-24: Loadings plot for PC2, reduced data set 193
Figure 4-25: Scores plot for PC3, reduced data set 194
Figure 4-26: Loadings plot for PC3, reduced data set 194
Figure 5-1: Hydrodynamic voltammograms of Epi and Dopa 211

Table of Figures

Figure 1-1: Schematic of a liquid chromatographic instrument 2

Figure 1-2. Representative hydrodynamic voltammograms for compounds A and B..... 8

Figure 1-3: A typical cross-flow thin-layer electrode cell..... 13

Figure 1-4. The detector cell used with NEUBA®..... 14

Figure 1-5: Peak types. 15

Figure 1-6: A typical chromatogram, and determination of the retention times of a nonretained species, t_M, and a retained species, t_R..... 17

Figure 1-7: Determination of the width at half height, and the width of the peak at the base. 18

Figure 1-8: Resolution of two peaks 19

Figure 2-1. Flow diagram for the NEUBA® system. 74

Figure 2-2: Representative Injection..... 85

Figure 2-3: Solvent Front Pattern..... 86

Figure 2-4: Retention time of Epi as a function of time corrections..... 88

Figure 2-5: Retention time of Dopa as a function of retention time corrections 89

Figure 2-6: Retention time of Dopamine as a function of time corrections 90

Figure 2-7: Overlap of Epi and Dopa..... 95

Figure 3-1: SSR results for Channel 1..... 102

Figure 3-2: Percent Difference for Channel 1, Mix 1 103

Figure 3-3: %RSD for Channel 1, Mix 1 105

Figure 3-4: Percent Difference for Channel 1, Mix 2 106
Figure 3-5: %RSD for Channel 1, Mix 2 107
Figure 3-6: Percent Difference for Channel 1, Mix 3 108
Figure 3-7: %RSD for Channel 1, Mix 3 109
Figure 3-8: Percent Difference for Channel 1, Mix 4 110
Figure 3-9: %RSD for Channel 1, Mix 4 111
Figure 3-10: Percent Difference for Channel 1, Mix 5 112
Figure 3-11: %RSD for Dopa, Channel 1, Mix 5 113
Figure 3-12: SSR results for Channel 2..... 116
Figure 3-13: Percent Difference for Channel 2, Mix 1 117
Figure 3-14: %RSD for Channel 2, Mix 1..... 118
Figure 3-15: Percent Difference Channel 2, Mix 2..... 119
Figure 3-16: %RSD for Channel 2, Mix 2..... 120
Figure 3-17: Percent Difference for Dopa, Channel 2, Mix 3 121
Figure 3-18: %RSD for Dopa, Channel 2, Mix 3 122
Figure 3-19: Percent Difference for Dopa, Channel 2, Mix 4 123
Figure 3-20: %RSD for Dopa, Channel 2, Mix 4 124
Figure 3-21: Percent Difference for Dopa, Channel 2, Mix 5 125
Figure 3-22: %RSD for Dopa, Channel 2, Mix 5 126
Figure 3-23: SSR results for Channel 3..... 128
Figure 3-24: Percent Difference for Dopa, Channel 3, Mix 1 129
Figure 3-25: %RSD for Dopa, Channel 3, Mix 1 130
Figure 3-26: Percent Difference for Dopa, Channel 3, Mix 2 132

Figure 3-27: %RSD for Dopa, Channel 3, Mix 2 133
Figure 3-28: Percent Difference for Channel 3, Mix 3 134
Figure 3-29: %RSD for Channel 3, Mix 3..... 135
Figure 3-30: Percent Difference for Channel 3, Mix 4 136
Figure 3-31: %RSD for Channel 3, Mix 4..... 137
Figure 3-32: Percent Difference for Channel 3, Mix 5 138
Figure 3-33: %RSD for Channel 3, Mix 5..... 139
Figure 3-34: SSR results for Channel 4..... 142
Figure 3-35: Percent Difference for Channel 4, Mix 1 143
Figure 3-36: %RSD for Channel 4, Mix 1..... 144
Figure 3-37: Percent Difference for Channel 4, Mix 2 145
Figure 3-38: %RSD for Channel 4, Mix 2..... 146
Figure 3-39: Percent Difference for Channel 4, Mix 3 147
Figure 3-40: %RSD for Channel 4, Mix 3..... 148
Figure 3-41: Percent Difference for Channel 4, Mix 4 149
Figure 3-42: %RSD for Channel 4, Mix 4..... 150
Figure 3-43: Percent Difference for Dopa, Channel 4, Mix 5 151
Figure 3-44: %RSD for Channel 4, Mix 5..... 152
Figure 4-1: A three-dimensional matrix 158
Figure 4-2: Scree plot for the original data set 163
Figure 4-3: Scores plot for original data set, PC1 166
Figure 4-4: Loadings plot for original data set, PC1 167
Figure 4-5: Scores plot for original data set, PC2 168

Figure 4-6: Loadings plot for original data set, PC2 169
Figure 4-7: Scores plot for original data set, PC3 170
Figure 4-8: Loading plot for original data set, PC3 171
Figure 4-9: Scores plot for original data set, PC4 172
Figure 4-10: Loadings plot for original data set, PC4 173
Figure 4-11: Scree plot for the reduced data set 176
Figure 4-12: Scores plot for PC1, reduced data set 178
Figure 4-13: Loadings plot for PC1, reduced data set 179
Figure 4-14: Scores plot for PC2, reduced data set 180
Figure 4-15: Loadings plot for PC2, reduced data set 180
Figure 4-16: Scores plot for PC3, reduced data set 181
Figure 4-17: Loadings plot for PC3, reduced data set 182
Figure 4-18: Scores plot for PC4, reduced data set 183
Figure 4-19: Loadings plot for PC4, reduced data set 183
Figure 4-20: Scree plot for the very reduced data set 189
Figure 4-21: Scores plot for PC1, reduced data set 191
Figure 4-22: Loadings plot for PC1, reduced data set 191
Figure 4-23: Scores plot for PC2, reduced data set 192
Figure 4-24: Loadings plot for PC2, reduced data set 193
Figure 4-25: Scores plot for PC3, reduced data set 194
Figure 4-26: Loadings plot for PC3, reduced data set 194
Figure 5-1: Hydrodynamic voltammograms of Epi and Dopa 211

Table of Tables

Table 2-1: Amounts of working standards used to make the standards and mixtures..... 80

Table 2-2: Resultant amounts regarded at true values for standards 80

Table 2-3: Electrode Potentials..... 81

Table 2-4: Statistics for Channel 4 data 93

Table 3-1: Summary of the data..... 154

Table 4-1: Percent variance captured by PCA model for the original data set..... 164

Table 4-2: Representative data for the entirely inclusive model, channel 2..... 174

Table 4-3: Percent variance captured by the PCA Model, reduced (121 points) data set 177

Table 4-4: Outcome for reduced data set, channel 1..... 185

Table 4-5: Outcome for reduced data set, channel 2..... 186

Table 4-6: Outcome for reduced data set, channel 3..... 187

Table 4-7: Outcome for reduced data set, channel 4..... 188

Table 4-8: Percent variance captured by the PCA Model, very reduced (21 points) data set..... 190

Table 4-9: Results for the PCA Model, very reduced (21 points) data set, channel 1 196

Table 4-10: Results for the PCA Model, very reduced (21 points) data set, channel 2 197

Table 4-11: Results for the PCA Model, very reduced (21 points) data set, channel 3 198

Table 4-12: Results for the PCA Model, very reduced (21 points) data set, channel 4 199

Table 4-13: Summary of the data for the Cramer's Rule Approach 204

Abstract

We wanted to create a simple technique that provided a high degree of accuracy in the deconvolution of two chromatographic peaks which were severely overlapped in the usual time-response domain and substantially overlapped in the potential-response domain. The two compounds we selected for investigation which fulfilled these initial criteria were epinephrine (Epi) and L-3,4-dihydroxyphenylalanine (Dopa). Five two-component mixture sets were prepared in the following molar response ratios, i.e., the moles of compound injected as a function of detector response: (1) 90% Epi to 10% Dopa; (2) 70% Epi to 30% Dopa; (3) 50% Epi to 50% Dopa, (4) 30% Epi to 70% Dopa and (5) 10% Epi to 90% Dopa. Six individual injections were performed for each mixture. These six injections were interspersed between coupled injections of Epi controls and Dopa controls. Thus, we had three injections of an Epi control, three injections of a Dopa control, and six injections of a mixture set; this procedure was followed for all five mixture sets, and ended with three injections of an Epi control and three injections of a Dopa control. The tyotal number of injections completed were, thus, six for each of mixtures 1 through 5, 18 Epi controls, and 18 Dopa controls. The current response was collected versus time using four electrochemical detectors in series having sequentially increasing potentials.

Our hypothesis was that a simple deconvolution in the potential-response domain at a carefully selected time following injection would yield sufficiently accurate determination of the two analytes to 95% confidence levels using the Student's t-test.

Our simple potential-response approach was thus compared to standard and commonly employed alternative approaches.

Twelve common straight forward time-response curve-fitting methods were initially examined. These included bi-Gaussian, exponentially modified Gaussian (EMG), Fraser-Suzuki, log-normal, Haarhoff-Van der Linde, Cauchy-Gaussian, Chesler-Cram, Giddings, (Gaussian Modified Gaussian (GMG), the 4-Parameter Gaussian-Exponentially Modified Gaussian (GEMG4), the 5-Parameter Gaussian-Exponentially Modified Gaussian (GEMG5), and the Exponentially Modified Gaussian + Gaussian Modified Gaussian (EMG+GMG).

Three common chemometric procedures using information from only two dimensions (time-response) or information from all three dimensions (time-potential-response) were next employed. These chemometric procedures included (1) principal component regression (PCR), (2) partial least squares (PLS) and (3) the Statistically Inspired Modification to the PLS algorithm (SIMPLS).

For each method investigated, we determined a mean and standard deviation for the calculated/predicted levels of both Epi and Dopa for each of the five separate mixtures. For the determinations involving only the two standard dimensions of information (i.e., time-response), this procedure was repeated for each of the four electrochemical channels separately. For the chemometric approaches using information from all the dimensions, only one such determination was available for each mixture. Likewise, the simple initially described procedure using potential-response time was only capable of providing a single outcome (mean ± s.d.) for each of Epi and Dopa for each of the five mixtures. Finally, the individual means were compared to the known standard

values in all cases using Student's t-test to determine if the means were the same as the standards at the 95% confidence level.

Among the twelve standard two-dimensional (time-response) approaches, analysis of the residual sum-of-squares for each clearly showed only five to be reasonably capable of approaching this problem. These five were the EMG, GEMG4, GEMG5, EMG+GMG and Cauchy-Gaussian. To provide these five with the best opportunity to yield the desirable deconvolution, we separately examined both the area and simple peak height data for each. Unfortunately, the most accurate of these approaches (EMG using peak heights) only predicted with 95% certainty the known content for both Epi and Dopa in only 5 out of 20 attempts.

The chemometric approaches employing two-dimensional time-response information were a bit more reliable in providing accurate outcomes. Each of the PCR, PLS, and SIMPLS approaches accurately predicted with 95% certainty the Epi and Dopa levels in 11 out of 20 cases.

The chemometric attempts employing information from all three dimensions simultaneously surprisingly were capable of accurately predicting the known standard values for both Epi and Dopa for only two of the five mixtures. Perhaps this is a reflection of the modest amount of instrumental noise contained in the data in combination with an overly inclusive approach.

Finally, the initially proposed single analysis of the data using the potential-response domain and Cramer's Rule for deconvolution proved to be much easier and more accurate than any of the alternatives listed above. In particular, this approach accurately predicted the content of both Epi and Dopa in all of the five mixtures

examined at the 95% confidence level. At the same time, the deconvolution using matrix algebra was and is substantially simpler than any of the other alternatives examined.

Thus, deconvolution of two severely overlapped chromatographic peaks in the time-response domain has been simply and accurately achieved using very limited amounts of potential-response information.

Chapter 1 — Introduction

The Problem

Of major concern for any chromatographic method is the existence of possible overlapping peaks, stemming from insufficient resolution of the analytes by the column, and the use of nonselective and nonspecific detectors. The method discriminates between dissimilar analytes, but similar analytes exhibit similar retention times and similar responses at the electrodes, resulting in overlapping peaks. For quantitative measurement, positive identification of the peaks and amounts represented by the peaks are required. Methods for this qualitative and quantitative determination have been investigated for other detection methods, but not for electrochemical detection. Therefore, a simple deconvolution method is needed.

Introduction to Chromatography and Electrochemistry

The basis of chromatography is separation of analytes. From its humble beginnings by Tswett[1] in 1906, the process has since evolved to include the separation of a mixture into its components with subsequent detection by a variety of detectors, including electrochemical detection, which produce an electronic signal proportional to the amount of analyte present. The current analytical separation methods use a very small amount of the mixture and allow for a definitive determination of the amount of substances present with the use of a detector specific for the compounds of interest.

In short, a chromatographic system consists of an injection system ("injector"), a column to effect the separation, a detector to detect the analytes as they elute from the column, and a recorder to record the signals as a plot of electronic signal vs. time (a chromatogram). Figure 1.1 shows a chromatographic system used in liquid chromatography with electrochemical detection (LCEC).

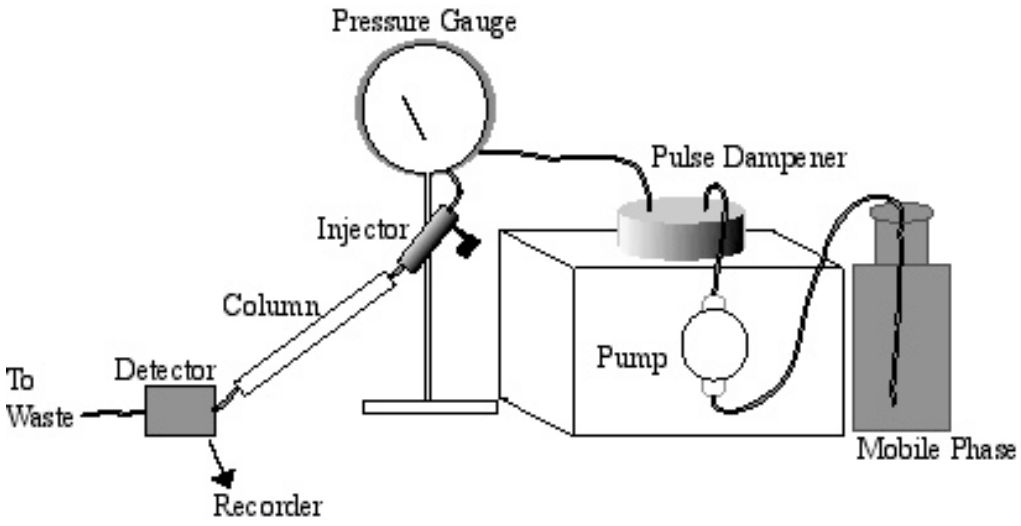


Figure 1-1: Schematic of a liquid chromatographic instrument

As can be seen in Figure 1.1 above, a liquid chromatographic system is fairly straightforward. The sample is introduced at the injector, and then carried through to the column, where the analytes are separated based on their relative affinities to the column substrate. They emerge at the end of the column, and are measured at the detector and recorded by the recorder. A liquid mobile phase, either a pure liquid or a solution, is pumped through the system and carries the analyte. Electrochemical detectors produce a signal from the pulses the pump produces. The pulse dampener reduces the pulses from the pump and thereby reduces the instrumental noise present at the detector. The

pressure gauge is used to monitor pressure in the system. It measures the backpressure generated by the column and is an indirect measure of how fast the mobile phase is being pumped through the system.

There are two aspects of the system to be discussed here, the separation of the analytes, and the detection of the analytes. Separation is the purview of chromatography, and the detection used in the system is electrochemical in nature. The mobile phase used in the system has characteristics of concern to both separation and detection, and a more full discussion of the mobile phase is held until the end of the electrochemistry section.

Separation: Chromatography

The stationary phase in the column is composed of packed silica particles coated with a compound designed to interact with the analytes. The amount of separation depends largely on the nature of the packing material itself—if the analytes have no affinity for the column, separation will not happen—but it also depends on the nature of the mobile phase. Analytes that strongly interact with the stationary phase over the mobile phase will be strongly retained and thus will slowly move through the column, resulting in long retention times. Analytes that weakly interact with the stationary phase are weakly retained and have short retention times. If the analyte does not interact at all with the column, it will move through the column at the same rate as the mobile phase and will have such a short retention time that it will be lost within the solvent front, which consists of all other non-retained substances giving a signal on the chromatogram. Ideally, the column will separate each analyte so that each analyte will exit the column at distinctly different times.

There are two components of the chromatographic system involved in separations one must consider, the mobile phase and the stationary phase as introduced earlier. The interactions between the stationary phase, the mobile phase, and the solute molecule are described by current chromatographic theory. In essence, this theory describes the process of separation as an equilibrium process with the solute molecule partitioning between the stationary phase and mobile phase. When an injection is made, the analyte mixture is introduced as a single plug which separates according to the affinities of the individual analytes, producing chromatographic zones or bands. Which phase the molecule is found in will depend on the concentration gradient established as the chromatographic zone passes through the column.

The speed at which a given chromatographic zone will pass through the column depends on several parameters. The first and foremost parameter is the retention characteristics of the analyte to the column, with structurally and chemically similar analytes having similar retention characteristics, and to a lesser degree the type and composition of the mobile phase.[2] Analytes similar in size, shape, and chemical properties tend to group together in time. As such, there are a limited number of possibly active sites available to the analyte molecule, and so there is some competition occurring for these sites among similar analytes.

The speed of the separation also depends on the flow rate of the mobile phase, [3, 4] with higher flow rates yielding shorter retention times. An unnecessarily low flow rate tends to be dominated by diffusional processes, and as such the chromatographic separation suffers with broader peaks than otherwise attainable. With a very high flow rate, however, the analytes may considerably overlap or even be lost in the solvent front

of the chromatogram. Another parameter is the temperature at which the separation is carried out[5], this can also have an effect on the retention characteristics of the analytes, since this process is governed kinetically. All these parameters must be considered to determine the optimal conditions for the chromatographic system. There is an optimal velocity for the mobile phase that will result in the greatest resolving power achievable for the column; this is usually determined experimentally and it is not necessary to locate it exactly in practice. In short, there is a trade-off between speed of the analysis, and the resolving power of the column.

Electrochemistry

Electrochemical detectors will respond to any compound that is electroactive at the potential of the electrode, and hence can be considered to be nonspecific and nonselective. These detectors are sensitive to changes in chromatographic conditions as well. A full description and explanation of the physics of the processes occurring at the electrode surface is, once again, beyond the scope of this work, and interested readers are referred to the abundant body of literature.[6-8]

The chemical reaction in question, of course, is the oxidation-reduction reaction:

$$Ox + ne^- \underset{k_a}{\overset{k_c}{\rightleftharpoons}} Red$$

Equation 1.1

where Ox and Red are the oxidized and reduced forms of the analyte, respectively, and k_c and k_a are the cathodic and anodic heterogeneous rate constants, respectively.

All faradaic electrochemistry depends, ultimately, on Faraday's Law:

$$Q = nFN$$

Equation 1.2

where Q is the number of coulombs passed, and N is the number of moles converted to product in the chromatographic zone. At any instant during the LCEC experiment, the current is reflected by the rate of conversion of the reactant to product.

Under equilibrium conditions, the net current is zero, and thus the electrode will adopt a potential based on the concentrations of the oxidized and reduced species, as dictated by the Nernst equation:

$$E = E^{o'} + \frac{RT}{nF} \ln\left(\frac{C_O^*}{C_R^*}\right)$$

Equation 1.3

The Nernst equation thus provides a link between electrode potential and the concentrations of participants in the electrode process. If an electrode system follows the Nernst equation or an equation derived from it, the electrode reaction is often said to be reversible, or Nernstian.

With further considerations and substitutions, the Nernst equation can be written in a form reflecting the limiting current for a given redox reaction:

$$E = E_{1/2} + \frac{RT}{nF}\ln\left(\frac{i}{i_l - i}\right)$$

Equation 1.4

where i_l is the limiting current, i is the response current, and $E_{1/2}$ is the half wave potential of the redox reaction involved. Rearrangement of this equation yields

$$i = \frac{i_l e^{\frac{nF}{RT}(E - E_{1/2})}}{1 + e^{\frac{nF}{RT}(E - E_{1/2})}}$$

Equation 1.5

A plot of the current vs. potential gives the normal S-shaped curve associated with the Nernst equation, seen in Figure 1.2 below showing the current-potential curves for two compounds. In this figure, the response current rises until it reaches a maximum, known as the limiting current. Problems with using potentials at half the limiting current include less current being generated at these potentials. At regions significantly below the limiting current, problems in reproducibility begin to appear.[9] The higher along the curve one measures, the less these reproducibility problems become but the greater the absolute contribution of noise to the signal. The optimum place along the curve is before the curve levels out, but rather high on the curve itself. The precise potentials used depend upon the compounds to be analyzed.

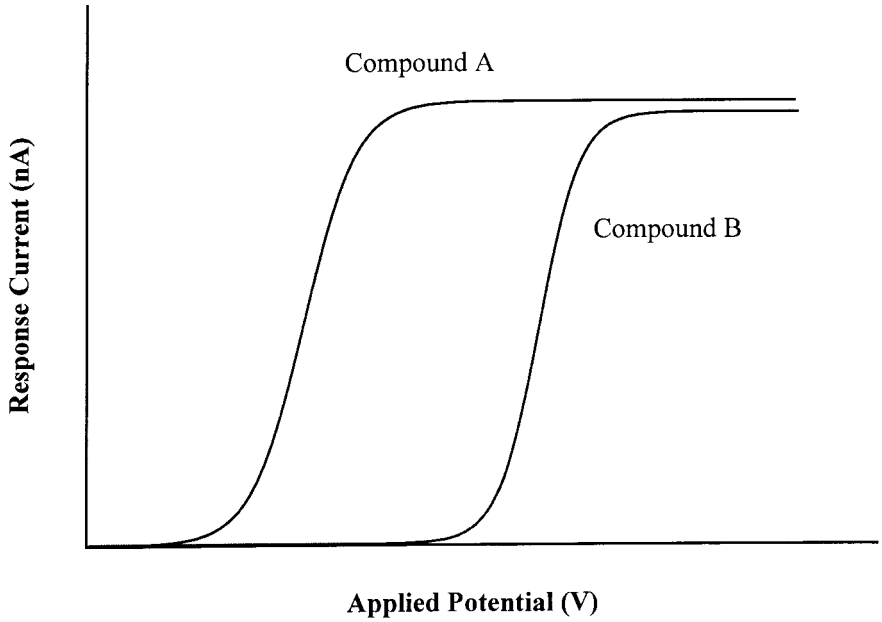


Figure 1-2. Representative hydrodynamic voltammograms for compounds A and B.

Figure 1.2 above shows two theoretical hydrodynamic voltammograms, one for compound A and one for compound B. Note that the current attributable to compound A reaches a maximum before current from compound B begins to manifest. In this case, compound A can be determined with no contribution from compound B; if one were looking only for compound A, an applied potential set lower than that required for compound B to become active will separate the two compounds.[10] In effect, only one compound will be seen. However, using only one electrode and setting the potential in the appropriate range where only A is active will not give a response for B unless the same sample is reanalyzed at a potential where both compounds are active. It is only at this point where the fact the sample contains two compounds becomes apparent. With increasing numbers of electrodes, increasing numbers of previously hidden analytes can be detected. This is a common use of multiple electrochemical detectors.

A second case can be seen if one imagines that the two curves are closer together, so that there is some overlap in currents developing from their respective compounds. In this case, a lower potential can still be applied, where one compound is absent or minimized, but the problem becomes larger still. In the worst case scenario, both curves overlap significantly enough that it is difficult or impossible to determine which compound, or what the percentage of the compounds, is responsible for the current seen using only one electrode.

Flow and Electrochemistry in Thin-Layer Flow-Through Cells

In a liquid chromatographic system using electrochemical detection (LCEC system), the chromatographic column provides a good amount of the selectivity due to the separating power of the column itself, as discussed above. As the eluents come from the end of the chromatographic column and to the electrochemical detector, the concentration gradient rises and falls as it passes through the electrochemical detector, yielding first a rising current and then a falling current.

Successful use of electrochemical detectors depend on well-defined hydrodynamic conditions, as well as a high rate of mass transport.[7, 11] The thickness of the diffusion layer at the surface of the electrode is related to the solution flow rate. The limiting steady-state response of flow-through electrodes has been determined, and solutions of the three-dimensional convective diffusion equation for the limiting currents under steady state conditions have been derived for various electrode geometries, using the appropriate boundary conditions. For an electrode array arranged in series, this solution holds for the first electrode, where reactants are being converted to products at

mass-transport controlled rates; in short, by diffusion. The diffusion layer does not have enough time to re-establish itself fully before it passes over the second electrode.[9, 12, 13] This results in less current generated and measured at the second and any subsequent electrodes in a series configuration, assuming that all the electrodes are exactly the same physically and are set for the same potential. Differences in the conditions on the individual electrode surfaces may make these differences in current more, or less, significant. A detailed examination of the case of dual series electrodes has been done, but this discussion is beyond the scope of this work.[12]

Electrode Material

The choice of electrode material is quite important, as the electrode must have the required mechanical ruggedness and long term stability required for the application to which it is being used. Glassy carbon, which has excellent mechanical and electrical properties, can be brought to a high surface polish with standard techniques, and is relatively free of impurities. In addition, the solvents used in the analyses normally performed have no deleterious effect on the electrode.[14] Glassy carbon can yield reproducible signal values over prolonged time periods;[15] virtually the same absolute signal values can be obtained after careful repolishing the electrode surface.[16] It is these characteristics that make glassy carbon electrodes a good choice for most applications.

Glassy carbon electrodes are not free from problems, however. The structure of glassy carbon is not homogeneous throughout the electrode, with ribbons of sp^2 hybridized carbons embedded in a sp^3 hybridized carbon matrix. Only the sp^2 hybridized carbons are electrochemically active. Thus, at the surface of the electrode there are be

regions with slightly different electrochemical properties. Electrode fouling is common, with Dopamine and other biogenic amines known to have a tendency to foul the electrode at higher potentials. Such compounds bond to the surface of the electrode and change the overall properties of the electrode itself. After repolishing, the electrode may not yield reproducible signals until the surface has been "conditioned," at which time the electrode begins to yield a reproducible signal. Thus, the use of glassy carbon may not be suitable for all analyses, and the diffficulties in its use must be weighed against the advantages.

Mobile Phase

The buffered aqueous mobile phase must dissolve the analytes, have sufficient ionic strength necessary for conductivity, and be chemically and electrochemically inert at the desired range of operating potentials. The concentration of electrolytes in the mobile phase is sufficient to minimize migration as a component in analyte movement. At the same time it should minimize the background current due to impurities found in the buffering agents[17] and dissolved oxygen[14] which contribute to baseline problems. Additionally, the mode of operation is a potential source of noise or problems; the oxidative mode of operation is much less susceptible to baseline problems due to reduction of the above mentioned trace impurities. The potential limits in an aqueous solution are from about -1.2V to about 1.2V.[15, 17] Any analyte that oxidizes or reduces outside of this range is not suitable for detection in an aqueous phase.

The mobile phase also should keep the compounds in the desired oxidative state. This requires some knowledge on the part of the experimenter. If the analytes are sensitive to oxygen, for example, then the mobile phase should be bubbled with an inert

but soluble gas to drive the oxygen from the mobile phase. Additionally, since the mobile phase is high in electrolytes, it sometimes can provide an excellent growth media for bacteria, and thus should be checked on a regular basis for bacterial growth. Some analysts will go to the extreme of preparing new mobile phase on a daily or weekly basis. Slight changes in the mobile phase, however, can cause significant problems in the elution of the compounds; additionally, the column must be conditioned to the new mobile phase prior to use.

The Detector: Thin-Layer Cells

The detectors used in many LCEC systems are of a cross-flow thin-layer design as shown in Figure 1.3. In this cell design, the electrochemically active analyte passes over the working electrode which is held at a potential sufficiently positive or negative enough to induce an electron transfer reaction (oxidation or reduction), producing an amperometric current proportional to the concentration of analyte that enters the thin-layer cell. The counter electrode supplies the electrons via the potentiostat to maintain a constant potential between the working electrode and the reference electrode. Amperometric detection can be used to react a negligible amount of the analyte, or nearly all of it, depending on the mass transport efficiency of the cell.

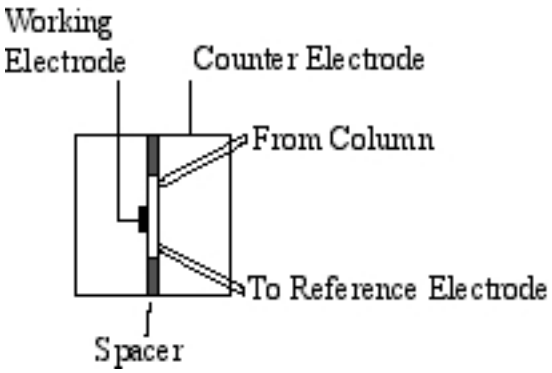


Figure 1-3: A typical cross-flow thin-layer electrode cell

The range of conversion efficiency comes from the area-to-volume ratio of the electrode surface to the volume of the cell. As the electrode surface area increases, the conversion efficiency improves. With the larger electrode areas and their increasing background current, detector noise also increases, and the signal-to-noise ratio decreases as well; this is problematic when it results in a lowering of the signal/noise ratio. With smaller electrodes, less material is converted, but lower detection limits can be attained with concomitant low noise levels. At extremely low conversion efficiencies, however, noise from the electronics becomes the limiting factor, as it becomes as great as, or greater than, the signal from the analyte.

The construction of the thin layer cell places the working electrode opposite the auxiliary or counter electrode, and is shown in Figure 1.3; this construction and the high ionic strength of the mobile phase minimizes the large iR drops commonly associated with amperometric thin-layer cells.[14] This configuration also ensures that the potential at all points on the working electrode surface will be uniform and that the uncompensated iR drop will be extremely small due to the thinness of the layer of solution between the

two plates.[17] Since no current passes along the thin layer, there is no resulting iR drop in the solution. The transit time for the molecules moving across the electrode surface is short when it is compared to the time it takes to diffuse across the thin layer of solution between one side of the electrode and the other, and thus no interference occurs between the reactions occurring at the working and auxiliary electrodes. The reference electrode for the cell used in NEUBA® is placed behind the auxiliary electrode, as shown in Figure 1.4.

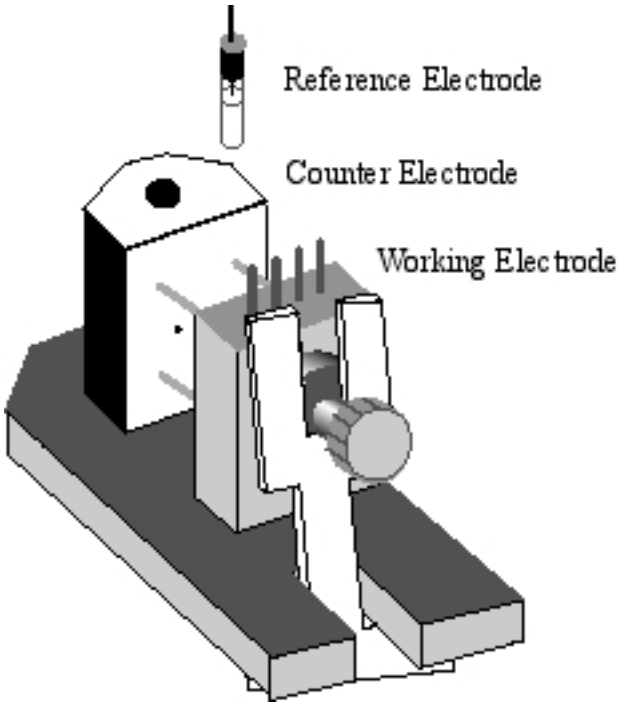


Figure 1-4. The detector cell used with NEUBA®

The Recorded Data and Resultant Peak Shapes

Initial chromatographic theory indicated that an ideal eluted peaks exhibited a Gaussian distribution. In truth many things affect the shape of the peak. A peak that is fronting, or showing an asymmetric first half of the peak, can be due to a dead volume

between the injector and the column, so that the injection is being made in one long plug instead of a short and narrow one. Tailing of a peak can be due to dead volumes in the connections from the column to the detector, or due to detector lag. Fronting and tailing are also functions of the flow rates in the system being used. A cold spot on a gas chromatographic column can have a similar effect. They can also be due to nonlinear distribution coefficients for the analytes. Distortions such as these lead to poor separations and less reproducible elution times, and thus should be avoided.

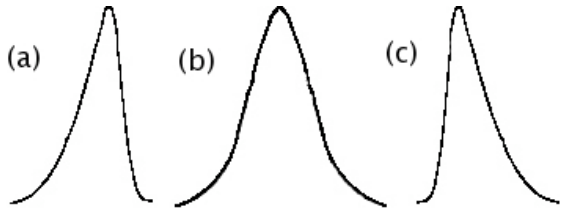


Figure 1-5: Peak types.

(a) Fronting peak; (b) symmetrical peak; (c) tailing peak

Resolution

Any column can be measured to determine the resolving power associated with that column. This quantity is known as the number of theoretical plates. The maximum number of separated peaks attainable for the column can also be determined. For a column of 10,000 theoretical plates yielding a resolution of 1, i.e., the peaks overlap by approximately 3%, the number of peaks the column can resolve is 25. In order to get 50 peaks to resolve on the column, the number of theoretical plates must be 20,000. Given

that biological samples can have hundreds of analytes present, a single column cannot resolve them all and thus overlapping peaks will be the rule rather than the exception.

Resolution (R_s) is the ability of the column to separate two analytes. It is a quantitative measurement of how well two peaks are separated in the analysis, and is determined by Equation 1.6:

$$R_s = \frac{2\left[\left(t_r\right)_B - \left(t_r\right)_A\right]}{W_A + W_B}$$

Equation 1.6

where A and B refer to the first and second peak, respectively, and the retention times of the two analytes are found as shown in Figure 1.6. The widths are measured at the base of the peak as shown in Figure 1.7.

The retention time of the analyte and as the width of the peak are values commonly used in peak shape equation determinations, as discussed later.

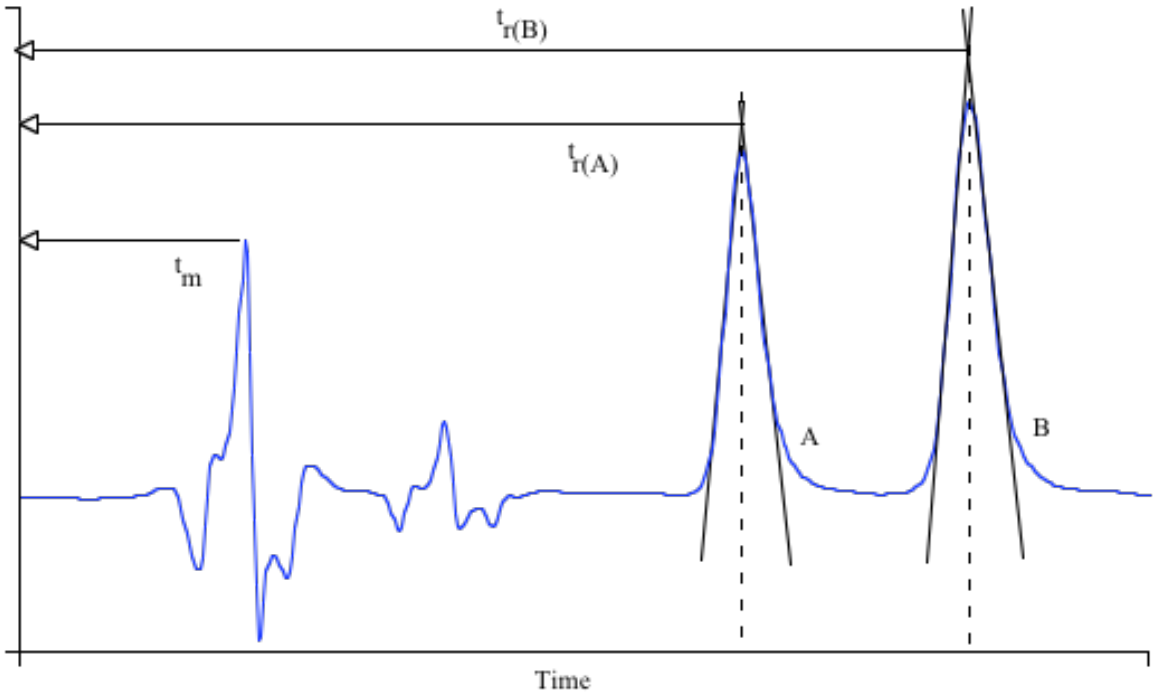


Figure 1-6: A typical chromatogram, and determination of the retention times of a nonretained species, t_M, and a retained species, t_R.

A resolution of 1.5 gives essentially a complete separation of the two components; at a resolution of 1.0, there is an overlap of the two analytes of about 4%. Increasing the number of plates of the column increases the resolution of the column; as the peak width narrows with retention times remaining constant, the resolution increases.

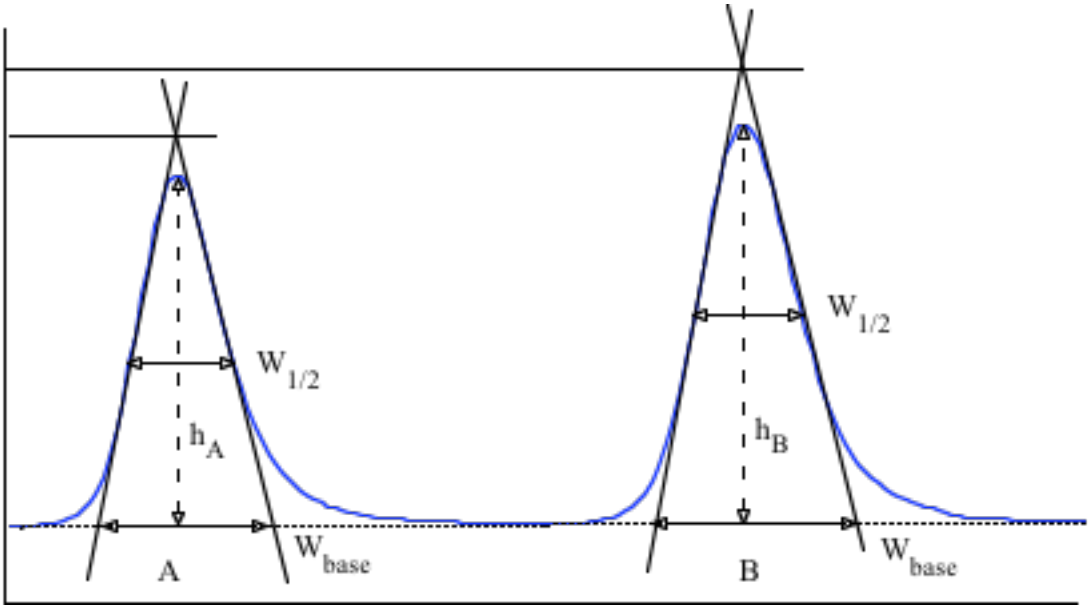


Figure 1-7: Determination of the width at half height, and the width of the peak at the base.

This is illustrated in Figure 1.8, where the resultant peak is the summation of the two peaks. In Figure 1.8(a), the peaks are well resolved and there is no difficulty in determining the two analytes. In 1.8(b), the peaks are moderately resolved, with a resolution of 0.50. The resultant peaks appear to be slightly wider, and the second peak begins to appear taller. In 1.8(c), the resolution is even poorer at 0.39. The second peak is now noticeably taller, and which parts of the doubled peak are attributable to which analyte becomes more difficult to determine. Part (d) shows the two peaks completely merged.

Resolution of Chromatographic Peaks

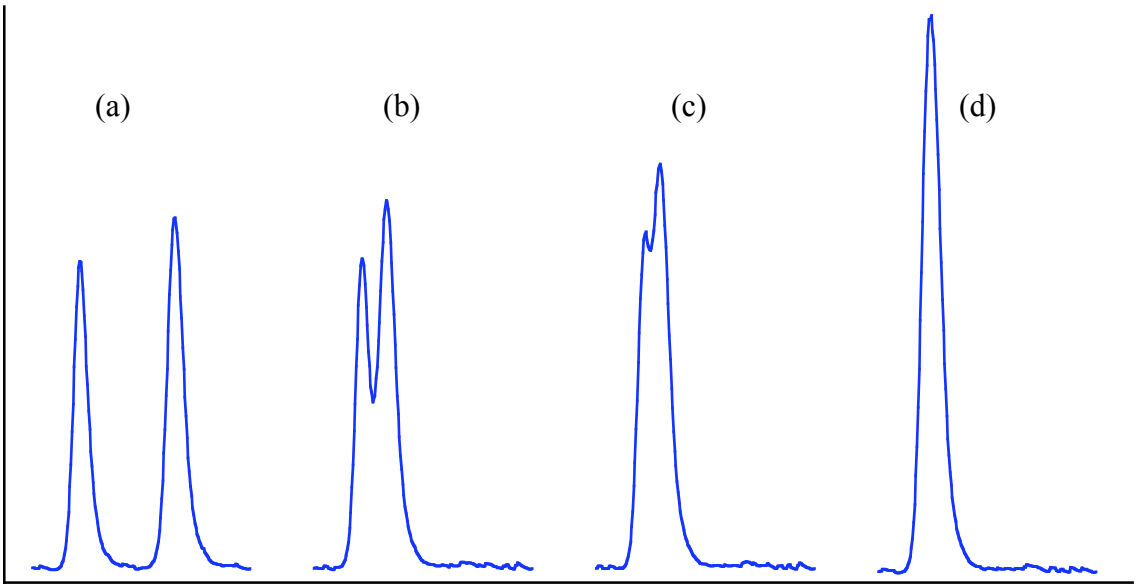


Figure 1-8: Resolution of two peaks

Key: (a) Well resolved, R_s = 2.78; (b) moderately resolved, R_s = 0.50; (c) poorly resolved, R_s = 0.39; (d) completely overlapping

All real chromatographic peaks exhibit skewed peak shapes, either as a fronting peak or a tailing peak. This skew in the peak shape complicates the quantification of the analyte. Figures 1.6, 1.7 and 1.8 were generated using real data. Each chromatographic peak in these figures exhibits tailing, which contributes to peak overlap.

Quantitation Issues

In the few cases where there is no overlap between the analytes, a direct measure of the quantitative amount for the analyte can easily be determined either as the peak height or as the integration of the area under the curve. If the recording device is a strip chart recorder, the easier method by far is to measure the peak height and from that value

determine the quantitative amount for the analyte. However, this method is not as precise or as reliable as is the integration of the area under the curve.

As peaks continue to overlap, the use of peak areas as a quantitative measurement becomes problematic. When the second peak overlaps the first, the region of overlap itself will be skewed since the signal will be a linear combination of the two analyte signals, thereby yielding a larger signal, as shown in Figure 1.8. For two peaks whose maxima are not involved in the overlap region, quantification based on peak heights rather than peak areas is quite straightforward. For moderately overlapped peaks, where the maximum of the second peak lies within the overlap region, there are two strategies commonly employed: the perpendicular drop method, and the tangent skimming method.

The perpendicular drop method draws a line between the two peaks based on where the minimum occurs in the valley between the peaks. This method is quick and easy, but tends to overestimate one peak while underestimating the other, depending on the relative amounts of the analytes and the degree of overlap. The tangent skimming method assumes that the leading or trailing edge of the peak is a continuation of the portion of the peak unaffected by the overlap itself. This method does not suffer as much from over- or underestimation of the relative areas of the two peaks. As the peaks become more and more overlapped, however, the valley disappears and any uncontaminated trailing edge of the preceding peak becomes less and less. Additionally, the adjacent peaks may not be similar in shape, and the points required to make these measurements may not be easily obtainable due to the overlap.

For severely overlapping peaks, the use of either perpendicular drop or tangent skimming is not appropriate. In these cases, the sample normally would be analyzed

again under different conditions, or additional steps must be taken to eliminate the undesired contaminant.

Internal Standards

The use of an internal standard is common practice for quantitative work. The purpose of an internal standard is two-fold: it serves as a peak against which one can quantitate, since it is a known amount, and it serves to indicate possible problems in the sample itself. An internal standard is a compound not normally found in the sample. It is added to the sample, either during extraction or just immediately prior to analysis, and is similar in structure and chromatographic behavior as the compounds of interest, but is chromatographically well separated from them.[18]

The internal standard is a known compound of known amount; hence, the response the detector yields for this compound is also known and should be consistent from injection to injection. Any deviations in the response of the internal standard is due to effects of the sample matrix. Since the sample matrix affected the internal standard, it is assumed to affect the analytes of interest in a similar manner. Thus, suppression of a signal can be accounted for and corrected.

Deconvolution of Overlapping Peaks

Two methods have been used in the past to resolve overlapping peak information: iterative curve fitting, and multivariate calibration methods; both are discussed below. These methods are not restricted just to chromatographic data.

Iterative curve fitting[19] assumes that a continuous function can be found that is the sum of one or more independent parts, each one of which can be itself described by a known function. The number parameters required to fit the chromatographic peak, however, has a tendency to rise as the peak shape becomes more and more complex. Additionally, a large number of equations can been used to fit a single peak to a high degree of precision. For accuracy, the number of analytes contributing to the peak must be known, else an inordinate number of "analytes" can be found leading to false positives.

Multivariate calibration methods have been used to resolve overlapping information in various analytical methods since before the birth of the field known as chemometrics. Mass spectrometers have enjoyed an explosion in use due to pattern recognition methods, enabling the identification of overlapping analytes.[20-22] NMR,[23] UV-Vis,[24] guided microwave spectroscopy,[25] and IR spectrometry[26] likewise have enjoyed improvements due to these methods. Recently, resolution of strongly overlapping peaks in capillary electrophoresis was reported.[27] The analysis of drinking water[28] and the characterization of supercritical adsorption processes between salicylic acid and various adsorbent materials using principal component analysis and partial least squares regression has also been reported. [29] Quantitative structure-activity relationships among carbonic anhydrase inhibitors were determined using multiple linear regression and computational neural network analyses.[30] These methods can also be used to determine the origin of various foodstuffs such as potatoes,[31, 32] honey,[33] the nitrogen content in wheat leaves,[34] and similarities between natural and accelerated browning of sherry wine,[35] as well as the source of illegal drugs[36] and pharmaceutical fingerprinting.[37]

It has also been used to determine the levels of veterinary drugs in edible tissues of animals.[38, 39]

Electrochemical determinations of analytes are not immune to overlapping peaks, irrespective of the presence of prior chromatographic separations. Various strategies and configurations have been investigated to overcome this problem. Voltammetric-amperometric dual electrode detection,[13] difference mode detection,[40] redox mode detection,[41] and the use of dual electrode response methods[42, 43] as strategies for LCEC detectors have been tried. Recently, chemometric methods have been applied to the problem. Artificial neural networks, along with partial least squares analysis, was applied to dropping mercury electrode polarography to determine the pesticides atrazine-simazine and terbutryn-prometryn in binary mixtures.[44] Coulometric and volumetric experiments involving slow titration reactions were also investigated.[45] Metal complexation determinations using differential-pulse polarography were performed and compared to traditional curve fitting least squares approaches.[46]

This work involves the investigation into iterative curve fitting as a method of deconvoluting two overlapping chromatographic peaks, as well as principal component analysis with linear regression to determine the actual components in the peaks. A simpler method of deconvolution may involve using a single well-chosen time point to investigate the data along the signal-potential domain. Therefore, a discussion of these methods is in order. A full discussion of these topics is well beyond the scope of this work.

Chromatographic Peak Shape Functions

It has been widely believed that if a function could be determined to exactly fit the data, then numerical methods could be used to determine the parameters attainable from the data, including the retention time of the species. This will be discussed in more detail below. Work has been done concerning thermodynamic data that can be determined from such functions after they have been fitted to the data. For such a scheme to work, however, a well defined peak shape function is required.[47, 48] Additional problems include: (a) poor convergence or ambiguous solutions are not uncommon, since the exact number of peaks is generally not known; (b) the position of the baseline itself is not exactly known, and in most cases is subjective in nature; and finally, (c) the initial estimates of the parameters are not accurate, leading to a solution at a local minimum as a result of the optimization method.[49] Fitting procedures that assume all the components have the same peak shape at different concentrations are subject to error.[50] It must be assumed that the detector response varies linearly with concentration at any time during the elution; this has been shown to not be the case except near the peak maximum.[50] For badly overlapped peaks where the equations being fitted to the data are not perfect representations of the data, the minimum found can be a very shallow function of the parameters,[51, 52] again leading to errors.

The method used to determine the curve fitting is usually a variant of the Marquardt method for least squares estimation for nonlinear parameters.[53] A very popular variant is the Levenberg-Marquardt method.[54, 55] Prior to Marquardt's method, most least-squares algorithms to determine the parameters used one of two methods; (a) either expanded the model as a Taylor series and made corrections to the several

parameters determined at each iteration on the assumption of local linearity, or (b) the method of steepest-descent was used. Steepest-descent methods,[49] also known as gradient methods, are still being used. Both methods often have difficulties, as the steepest-descent method suffers from slow convergence after the first few iterations, and the Taylor series suffers from divergence of successive iterations.[53] The sum of squares of the residuals is routinely used as a criterion for evaluating the peak shapes used in the peak deconvolution.[56]

There is no one single peak shape that can be ascribed as a "one-size-fits-all" equation, and the exact peak shape must be experimentally determined.[50, 57, 58] It has been suggested that different functions could fit the band envelope with equivalent statistical validity, and equal fits may be obtained with more than one set of parameters depending on the starting values and the methods used to compute the parameters.[59, 60] Because of the extra- and intra-column effects, it has been stated that the experimental retention time as well as the variance do not properly describe the shape of the peak under experimental conditions; it was further stated by Papoff et. al that these parameters do not allow one to predict how the shape will change when the composition of the eluent is varied[2] in gradient elution chromatography. Instrumental effects affect the chromatography; forms of the exponentially modified Gaussian equation have been used to model the instrumental contributions and subsequently subtracting them out of the data, allowing for sharpening of the peaks and thus attaining better separation of the peaks.[61]

The first and most simple theory developed for chromatography was the plate theory. This theory still accounts for much of what is generally observed in practice. It

describes the chromatographic process as a number of individual steps, and divides the chromatographic column into an infinitesimal number of discrete compartments which the chromatographic band would travel down in a stepwise fashion. The resultant band shape described by the model is Gaussian. It was quickly determined that Gaussian band shapes did not sufficiently describe the band shape[48, 62] for many kinds of chromatography and spurred much investigation into the exact nature of chromatographic processes. A stochastic model of chromatography has since evolved.

Several groups were working on a "universal" band shape equation. These groups attacked the problem on two competing fronts: a theoretical approach, and a practical approach. The theoretical approach utilized those equations with parameters describing real processes to the parameters in the equations but perhaps were not easy to implement. The practical approach uses those equations that while perhaps do not have parameters that are easily identifiable with any particular real phenomena, nevertheless fit the data very well and tend to be easier to implement.

The "theoretical" equations allow one to calculate thermodynamic data concerning the system under study; these equations include the Gaussian and those derived from it. Working with the existing theory led to using either a skewed Gaussian,[63, 64] a modified Gaussian function with a polynomial standard deviation,[2] or the Gaussian equation convolved with a variety of other peak shapes,[65] from an exponential decay[47, 66] to a Lorentzian (Cauchy)[67, 68] band shape. These include the exponentially modified Gaussian (EMG),[2, 3, 47, 48, 54, 55, 57, 58, 60, 62, 66, 68-83] a simplified exponentially modified Gaussian,[84] the Gaussian-exponentially modified Gaussian (GEMG),[67, 83] and the derivatives of these two or even the use of two Gaussian functions.[2] It also includes

the Cauchy-Gaussian[58, 59, 82, 83, 85, 86] (CG) equation which fits strongly tailed peaks. Of these equations, the EMG has been used rather extensively to describe chromatographic figures of merit[87-89] as well as thermodynamic information[88] concerning the system at hand, since the statistical moments for the equation are well known and this equation has sound footing in theory,[57, 61, 66] and has been used to study extra-columnar effects in chromatographic systems.[2] It tends to fit more peaks, including those that front as well as those that tail.[76, 77] The EMG function also has been examined as a function of more than one time constant.[83, 90] However, it does contain an integral form (the error function) that can be difficult to estimate numerically.[91] A hybrid of the exponential and Gaussian functions, named the exponential-Gaussian hybrid function (EGH), is:

> "mathematically simple, numerically stable, and its parameters are readily determined by making graphical measurements and applying simple equations … [and it] serves as a useful alternative to the exponentially modified Gaussian (EMG) for modeling slightly asymmetrical peaks since the two models produce nearly the same profile at relatively low asymmetries. The EGH also serves as an addition to the extensive list of alternative models that are sometimes better than the EMG at describing highly asymmetrical peaks."[92]

The peaks at low asymmetries are those that are best described by the EMG. For peaks at higher asymmetries, the EMG often does not describe the shape of the experimental peak very well and the function becomes numerically unstable.

Haarhoff and van der Linde[54, 55, 58, 81, 83] derived an equation based on the Gaussian equation that takes into account nonideality, nonlinearity,[93-95] and sorption effects[94] that will fit both tailing and fronting peaks. However, this equation shows changes in the elution curve with increasing amounts of injected solute; the peak maximum shifts to earlier times, and the asymmetry ratio decreases. Giddings derived an equation[54, 55, 83] firmly based in theory that can be used for peaks that have moderate tailing. There is even an equation describing nonlinear chromatography,[83] which has been considered for modeling ideal delta function analyte loading.[54]

The "practical" equations are those that fit the data well, even though there may not be any theoretical basis for them and the parameters may not describe real phenomena. These are normally used to determine the identity of the peak based on the retention time. Given that the elution profile is a statistical distribution of the peak, statistical functions such as the Poisson distribution,[55, 68, 73, 74, 83, 96] a modified Poisson distribution,[97] the Weibull function,[67, 81, 83, 98, 99] the Gamma function,[67, 81, 83, 91, 98, 99] the log-normal function,[67, 81, 83, 87, 98-101] and the Gram-Charlier[81, 87, 88, 102] and the Edgeworth-Cramer series[81, 83, 87, 102, 103] have been used. The latter two approaches are not very useful for peaks that are skewed and require series expansions which represent the determination of between seven and ten parameters per peak.

Additional functions that have been proposed and used in the past include linear combinations of two or more skewed Gaussian functions with differing parameters.[85] The bi-Gaussian function, where two half-Gaussian peaks with different widths are added, has been used[2, 58, 68, 82, 83, 96, 104] as well. This function will model both fronting and tailing peaks. An interesting parameter is the ratio of the width of the back peak to

the width of the front peak, σ_1/σ_0. When two overlapping bi-Gaussian peaks with differing ratios have a resolution of zero (complete overlap), the double bi-Gaussian peaks can be differentiated from a true single bi-Gaussian peak. This may show that the peak is not pure, but does nothing to determine the concentrations of the components. The bi-Gaussian equation does not yield correct first-order moment values, and since it contains two width values, the second order moments are not directly related to the peak shape.[2] A model similar to the bi-Gaussian equation was proposed by Li[105] wherein he transformed the leading and trailing edges of a Gaussian function into step functions to fit chromatographic peaks.

Chesler and Cram proposed convoluting a Gaussian, an exponential, and a hyperbolic tangent joining function, where the front of the peak can be fit with the Gaussian, the back side of the peak with the exponential function, and the hyperbolic tangent functions as a broadening function to enable the fit.[19, 58, 80, 83, 88] This function has been used to examine chromatographic processes.[106] It contains some eight parameters, and even with this large number of parameters used to fit the data, it is still not a very useful equation.[105] Additionally, derivative functions have been proposed,[85] as well as using the leading and trailing edges of sigmoid functions.[107, 108]

A recent review found about ninety empirical functions that have been used for the representation of chromatographic and spectroscopic peaks.[83] A count of those equations listed in this review that have been used for chromatography numbered some seventy-one. As noted by the authors, the literature contains instances of two apparently different equations representing the same function being presented with different names, as well as instances of printing errors, some serious, that have been propagated. Several

equations have more than one name, and some names are currently in use for more than one equation. Different forms of the same equation have been reported.

The parameters in the equations themselves are not uniform throughout the equations found in the literature. For instance, the width of a peak depends on where it is measured; different equations measure width at different heights. Thus, uniformity in the equations themselves is necessary in order to allow a more direct comparison between the different equations. The parameters are listed with each equation. The following parameters are of note. The height of the peak can be either exactly the height, in which a change in the magnitude of the parameter will change the height, or the height can depend on other parameters in the equation. The stated retention time is either the exact retention time, or it can depend on other parameters. A change in the magnitude of the retention time for an equation expressing the exact retention time, of course, will change the retention time with no changes to other parameters. If the retention time depends on other parameters, a change in the retention time will change the related parameters as well. Different functions measure the peak width at different heights. In some cases changes in the parameter itself changes the width; or in most cases, the width depends on other parameters. Ideally the symmetry parameter should be positive for fronted peaks, zero for symmetric peaks, and negative for tailed peaks, but this is not uniform for all the functions. The physical meaning of this parameter depends on the function itself. For some equations, the symmetry can be changed by changing the parameter itself, and for other equations it depends on other parameters.

Twelve equations are used in this work. Since there are some ninety possible functions that could be used, only the most common ones used in the literature and those

that appeared to be good candidates were evaluated in this work. These equations are discussed below. Depending on the author, several parameters in the equations may be reported in various ways. An attempt has been made to make the variables uniform to allow a more direct comparison between the different functions. These twelve equations are:

1. Haarhoff – Van der Linde (HVL)

$$y = \frac{\frac{a_0 a_2}{a_1 a_3 \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - a_1}{a_2} \right)^2 \right]}{\frac{1}{\exp\left(\frac{a_1 a_3}{a_2} - 1 \right)} + \frac{1}{2}\left[1 + erf\left(\frac{x - a_1}{\sqrt{2} a_2} \right) \right]}$$

Equation 1.7

where a_0 is the area, a_1 is the peak center, a_2 is the width, and a_3 is the distortion factor. Another form of this equation, in more familiar chromatographic terms, is:

$$c = \frac{\frac{A\sigma^2}{t_R \alpha_1} \exp\left[\left(\frac{t - t_R}{2\sigma^2} \right)^2 \right]}{\left[\exp\left(\frac{t_R \alpha_1}{\alpha^2} \right) - 1 \right] + \frac{1}{2}\left[1 + erf\left(\frac{t - t_R}{\sqrt{2}\sigma} \right) \right]}$$

Equation 1.8

where c is the concentration, A is area, t_R is the retention time, σ is the standard deviation of the Gaussian function, and α values are fitting parameters. This equation will yield a single maximum, and can be used on fronted, symmetrical or nearly symmetrical peaks, and also on tailed peaks. The zeroth statistical moment (area) can be determined exactly from this equation.

2. Exponentially Modified Gaussian (EMG)

In its bi-directional form:

$$y = \frac{a_0}{2\alpha_1}\exp\left(\frac{a_2^2}{2\alpha_1^2} + \frac{a_1 - t}{\alpha_1}\right)\left[erf\left(\frac{t - a_1}{\sqrt{2}a_2} - \frac{a_2}{\sqrt{2}\alpha_1}\right) + \frac{\alpha_1}{|\alpha_1|}\right]$$

Equation 1.9

where a_0 is the area, a_1 is the center, a_2 is the width, and a_3 is the distortion factor. This equation has many forms, among which is

$$c = \frac{\sqrt{2\pi}A\sigma}{2\alpha_1}\exp\left(\frac{\sigma^2}{2\alpha_1^2} + \frac{t_R - t}{\alpha_1}\right)\left[1 + erf\left(\frac{t - t_R}{\sqrt{2}\sigma} - \frac{\alpha}{\sqrt{2}\alpha_1}\right)\right]$$

Equation 1.10

where c is concentration of the analyte as a function of time, A is area, t_R is the retention time, σ is the standard deviation of the Gaussian function, and α values are fitting

parameters. According to a recent review,[83] almost 20 formulae for the EMG function have been reported in the literature, with the first seven listed are stated as independently derived and therefore equivalent to one another; the next four are approximate expressions, equivalent to the "true" EMG only when the retention time is much larger than the symmetry parameter, which is generally the case. The remainder in this review are all different from one another, and also different from the first twelve; printing errors have also crept into the literature of EMG equations. It has been suggested that the peak shape most commonly encountered with chromatographic data is the EMG.[75, 83] Much work has been done using this equation, and hence it is the best understood of the equations commonly in use. The calculation of the figures of merit for this equation were worked out by Foley and Dorsey[89] in 1983, and has since been used to determine thermodynamic properties of analytes.

3. The Giddings Equation. This equation does not fit highly skewed peaks. Since the definition of "highly skewed peaks" is not commonly encountered in the literature, the Giddings equation was included in this study for completeness. This equation takes the form:

$$y = \frac{a_0}{a_2}\sqrt{\frac{a_1}{t}}I_1\left(\frac{2\sqrt{a_1 t}}{a_2}\right)\exp\left(\frac{-t - a_1}{a_2}\right)$$

Equation 1.11

where a_0 is the area, a_1 is the center, a_2 is the width, and I_1 is the Bessel function for n=1, where t<<n:

$$I_n(t) = \frac{1}{n!}\left(\frac{t}{2}\right)^n, t \geq 0$$

Equation 1.12

4. Half-Gaussian Modified Gaussian (GMG)

$$y = \frac{a_0 \exp\left(-\frac{1}{2}\frac{(t-a_1)^2}{\alpha_1^2 + a_2^2}\right)\left[1 + erf\left(\frac{\alpha_1(t-a_1)}{\sqrt{2}\alpha_1\sqrt{\alpha_1^2 + a_2^2}}\right)\right]}{\sqrt{2\pi}\sqrt{\alpha_1^2 + a_2^2}}$$

Equation 1.13

where a_0 is the area, a_1 is the center, a_2 is the width, and a_3 is the distortion factor. As can be surmised, this is a convolution of a half-Gaussian equation with an exponentially modified Gaussian equation. This allows fitting to tailing peaks with more Gaussian characteristics in the main part of the peak itself.

5. 4 Parameter EMG-GMG hybrid (GEMG4)

$$y = \frac{a_0 \exp\left(-\frac{1}{2}\frac{\left(t - a_1 + \alpha_1\right)^2}{\alpha_1^2 + a_2^2}\right)\left[-1 + erf\left(\frac{a_2^2 - \alpha_1\left(t - a_1\right)}{\sqrt{2}a_2\sqrt{\alpha_1^2 + a_2^2}}\right)\right]}{\sqrt{2\pi}\sqrt{\alpha_1^2 + a_2^2}erf\left(\frac{\sqrt{2}}{2} - 1\right)}$$

Equation 1.14

where a_0 is the area, a_1 is the center, a_2 is the width, and a_3 is the distortion factor. This is another hybrid EMG peak with a Gaussian-Modified Gaussian peak using four parameters.

6. 5 Parameter EMG-GMG hybrid (GEMG5)

$$y = \frac{a_0 \exp\left(-\frac{1}{2}\frac{\left(\alpha_2 t - a_1\alpha_2 + \alpha_1^2\right)^2}{\alpha_2^2\left(\alpha_1^2 + a_2^2\right)}\right)\left[1 + erf\left(\frac{\alpha_1\left(-a_2^2 + \alpha_2 t - a_1\alpha_2\right)}{\sqrt{2}a_2\alpha_2\sqrt{\alpha_1^2 + a_2^2}}\right)\right]}{\sqrt{2\pi}\sqrt{\alpha_1^2 + a_2^2}erf\left(\frac{\sqrt{2}\alpha_1}{2\alpha_2} - 1\right)}$$

Equation 1.15

where a_0 is the area, a_1 is the center, a_2 is the width, and finally a_3 and a_4 are distortion factors. This is another hybrid EMG peak with a Gaussian-Modified Gaussian peak using five parameters; the use of additional fitting parameters should allow for a better fit

for the equation. A large number of parameters can only be justified when an even larger number of experimental data points are available.

7. EMG + GMG

$$y = \frac{a_0}{4\alpha_1}\exp\left(\frac{2\alpha_1 - 2\alpha_1 t + a_2^2}{\alpha_1}\right) erfc\left(\frac{\alpha_1 - \alpha_1 t + a_2^2}{\sqrt{2}a_2\alpha_1}\right) + \frac{a_0}{2\sqrt{2\pi}\sqrt{a_2^2 + \alpha_2^2}}\exp\left(-\frac{1}{2}\frac{\left(a_1 - t\right)^2}{a_2^2 + \alpha_2^2}\right) erfc\left(\frac{\alpha_2\left(a_1 - t\right)}{\sqrt{2}a_2\sqrt{a_2^2 + \alpha_2^2}}\right)$$

Equation 1.16

where a_0 is the area, a_1 is the center, a_2 is the width, and a_3 is the distortion1 factor, and a_4 is the second distortion factor. This is a summation between the EMG and the Half Gaussian-Modified Gaussian equation. As stated before, better fits to the experimental data can be found with increasing numbers of functions being used to fit a peak. However, this equation may be problematic for use when using it to deconvolute two overlapping peaks, since the two halves of the equation are not coupled. In this case, a total of four equations are actually being used.

8. Bi-Gaussian (BG), in its chromatographic form:

$$y = A\exp\left(-\frac{\left(t - t_R\right)^2}{2\sigma_L^2}\right), t < t_R$$

$$y = A\exp\left(-\frac{\left(t - t_R\right)}{2\sigma_R^2}\right), t > t_R$$

Equation 1.17

where A = area, t is time, t_R is the retention time, and σ_L and σ_R are the standard deviations for the totally symmetric function. This is a discontinuous and thus undifferentiable function; determination of the peak maximum via the first derivative of the equation is not available. However, since it is discontinuous, with the first half of the equation having a different standard deviation than the last half, the retention time reported by the equation will be the peak maximum.

9. Fraser-Suzuki (FS)

$$c = A\exp\left(-\ln 2\left\{\frac{\ln\left[1 + \alpha_1\frac{\left(t - t_R\right)}{\sigma}\right]}{\alpha_1}\right\}^2\right)$$

Equation 1.18

where A = area, t is time, t_R is the retention time, and σ_L and σ_R are the standard deviations for the totally symmetric function. This equation has been stated as being the same as the log-normal equation.

10. Log-Normal

$$c = A\exp\left[\frac{-\ln 2}{\ln^2\left(\frac{\alpha_1}{\sigma}\right)}\ln^2\left(\frac{t-t_R}{\sigma+\alpha_1}\frac{\frac{\alpha_1}{\sigma^2}-1}{\frac{\alpha_1}{\sigma}}\right)+1\right]$$

Equation 1.19

where A is the area, t is time, t_R is the retention time, and σ is the standard deviation, supposedly for the totally symmetric function. This equation has been shown in the literature to fit slightly skewed peaks, and has been promoted as a "universal" peak fit equation.[83]

11. Cauchy-Gaussian

This function is the result of a convolution between a Gaussian function and a Lorentzian, or Cauchy, function.

$$c = A\left\{f_L\exp\left[\frac{-\left(t-t_R\right)^2}{2\sigma_L^2}\right]+\frac{\left(1-f_L\right)\sigma_L^2}{\sigma_L^2+\left(t-t_R\right)^2}\right\}, t < t_R; 0 < f_L < 1$$

$$c = A\left\{f_R\exp\left[\frac{-\left(t-t_R\right)^2}{2\sigma_R^2}\right]+\frac{\left(1-f_R\right)\sigma_L^2}{\sigma_L^2+\left(t-t_R\right)^2}\right\}, t > t_R; 0 < f_R < 1$$

Equation 1.20

where A = area, t is time, t_R is the retention time, and σ_L and σ_R are the standard deviations for the totally symmetric function. This is another discontinuous function, and as such the retention time is expected to be exact.

12. Chesler-Cram, in its chromatographic form:

$$c = A\left\{\exp\left[\frac{-\left(t - t_R\right)}{2\sigma^2}\right] + \left[1 - \tfrac{1}{2}\left(1 - \tanh\left\{\alpha_1\left(t - \alpha_2\right)\right\}\right)\right]a_3 \exp\left[-\tfrac{1}{2}\alpha_4\left\{\left|t - \alpha_5\right| + \left(t - \alpha_5\right)\right\}\right]\right\}$$

Equation 1.21

where A = area, t is time, t_R is the retention time, and σ is the standard deviation, supposedly for the totally symmetric function.

Several authors reported the band broadening found in severely tailing peaks to be more an instrumental band-broadening effect rather than specific to the physical processes of adsorption and desorption of an analyte to the column. If the peaks are very closely eluting, as are the peaks in the present study, this asymmetry can be attributed to the chromatographic system, and thus the parameters describing the asymmetry will be the same for both peaks. Initially, both peaks were allowed to find independent asymmetry parameters, and these results were compared to the results obtained when the asymmetry parameters were the same for both peaks. The results of this work are presented later.

Principal Components - Background

Principal component analysis is an approach based on linear algebra which allows for the reduction in the data but still keep much of the variation in the data set. The data is analyzed in reference to various axes, with the axis of the first principal component along the axis describing the most variance in the data; the second principal component is then placed orthogonal to the first and is thus uncorrelated. The result can be considered in two ways: first, the old variables can be considered in terms of linear combinations of the new variables, and as such, the old variables can be explained by a smaller set of new variables, called factors, and thus reduces the data set. Alternatively, the new variables can be considered as linear combinations of the old ones.

To explain this method, consider the problem in two-dimensional space. The linear transformations are:

$$u_1 = a\left(y_1 - \overline{y_1}\right) + b\left(y_2 - \overline{y_2}\right) = ax_1 + bx_2$$

Equation 1.22

$$u_2 = c\left(y_1 - \overline{y_1}\right) + d\left(y_2 - \overline{y_2}\right) = cx_1 + dx_2$$

Equation 1.23

The variance in u can be made arbitrarily large by multiplication of the coefficients a and c, as well as b and d, by some constant value. Thus, a restriction on these coefficients must be made, the so-called normalizing constraint:

$$a^2 + b^2 = 1$$

Equation 1.24

$$c^2 + d^2 = 1$$

Equation 1.25

Equations 1.22 and 1.23 make the new coordinate system (u_1, u_2) orthogonal; Equations 1.24 and 1.25 make the transformation vectors of unit length. Furthermore, these conditions make the angle θ between u_1 and x_1 the same as that between u_2 and x_2 so that

$$u_1 = \cos\theta x_1 + \sin\theta x_2$$

Equation 1.26

$$u_2 = -\sin\theta x_1 + \cos\theta x_2$$

Equation 1.27

and thus it can be seen that

$$a = d = \cos\theta$$

Equation 1.28

$$b = -c = \sin\theta$$

Equation 1.29

In matrix form this can be expressed as:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \bullet \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Equation 1.30

Equation 1.30 represents the rotation in the coordinate system from the old variables to the new. We wish to maximize the variation encompassed for each of the components, and so we look to the general variance-covariance matrix for the *u* variables

$$\begin{bmatrix} \sigma^2_{u_1} & \sigma_{u_1 u_2} \\ \sigma_{u_2 u_1} & \sigma^2_{u_2} \end{bmatrix}$$

Equation 1.31

which can be written as

$$\begin{aligned} \sigma^2_{u_1} &= \cos^2\theta\sigma^2_{x_1} + \sin^2\theta\sigma^2_{x_2} + 2\sin\theta\cos\theta\sigma_{x_1 x_2} \\ &= \text{variance along } u_1 \end{aligned}$$

Equation 1.32

Since the new variables are restricted to be uncorrelated, $\sigma_{u_1u_2}$ and $\sigma_{u_2u_1}$ are both zero. The angle where $\sigma_{u_1}{}^2$ is maximized can now be found by setting the derivative of $\sigma_{u_1}{}^2$ to zero, and this angle is given by:

$$\theta = \frac{1}{2} Arc\tan \frac{2\sigma_{x_1x_2}}{\sigma_{x_1}^2 - \sigma_{x_2}^2} \text{ if } \sigma_{x_1}^2 > \sigma_{x_2}^2$$

Equation 1.33

and

$$\theta = 90 + \frac{1}{2} Arc\tan \frac{2\sigma_{x_1x_2}}{\sigma_{x_1}^2 - \sigma_{x_2}^2} \text{ if } \sigma_{x_1}^2 < \sigma_{x_2}^2$$

Equation 1.34

If the variances are equal, the angle becomes 45^o: positive for positively correlated variables and negative for negatively correlated variables. At this point, several terms should be mentioned. Each vector that defines a principal component is called an eigenvector of the variance-covariance matrix, and the variance along this vector is called its eigenvalue. These terms will be used again later in this discussion.

Now consider the problem in multiple dimensions, *m*. The general transformations are:

$$
\begin{aligned}
u_1 &= v_{11}x_1 + v_{12}x_x + \cdots + v_{1m}x_m \\
u_2 &= v_{21}x_1 + v_{22}x_x + \cdots + v_{2m}x_m \\
&\vdots \\
u_m &= v_{m1}x_1 + v_{m2}x_x + \cdots + v_{mm}x_m
\end{aligned}
$$

Equation 1.35

The coefficients v can be written into a matrix of coefficients, V:

$$
V = \begin{bmatrix}
v_{11} & v_{12} & \cdots & v_{1m} \\
v_{21} & v_{22} & \cdots & v_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
v_{m1} & v_{m2} & \cdots & v_{mm}
\end{bmatrix}
$$

Equation 1.36

and the u and x vectors are column vectors of the new and old variables:

$$
u = \begin{bmatrix}
u_1 \\
u_2 \\
\vdots \\
u_m
\end{bmatrix}
$$

Equation 1.37

$$
x = \begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_m
\end{bmatrix}
$$

Equation 1.38

The transformation can be shown in matrix form as:

$$u = V \cdot x$$

Equation 1.39

Once again, there are two conditions that need to be satisfied: orthogonality and the vector having unit length. Thus, for each new pair of coordinates u_k and u_r:

$$v_{k1}v_{r1} + v_{k2}v_{r2} + \cdots + v_{km}v_{rm} = 0$$

Equation 1.40

and for each component u_r:

$$v_{r1}^2 + v_{r2}^2 + \cdots + v_{rm}^2 = 1$$

Equation 1.41

These two conditions, taken together, are equivalent to:

$$V \cdot V^T = I$$

Equation 1.42

where I is the identity matrix. For the two-dimensional case, V was satisfied by the variance-covariance matrix as seen in Equation 1.31 above. For the multidimensional case, however, the use of this approach yields very complex expressions. However, this problem can be approached using matrix mathematics. The variance-covariance matrix of the new variables u_r can be stated as a function of the variance-covariance of the x variables:

$$C_u = V \cdot C_x \cdot V^T$$

Equation 1.43

where C_u and C_x are the variance-covariance matrices of the u and x variables, respectively. The expression to maximize then becomes

$$Var(u_1) = v_1^T \cdot C_x \cdot v_1$$

Equation 1.44

where $v^T \bullet v = 1$ and $v_1{}^T$ is a row vector representing the transformation by which we get u_1; u_1 is a linear transformation of the first row of matrix V (the x variables)

$$v_1^T = \begin{pmatrix} v_{11} & v_{12} & v_{13} & \cdots & v_{1m} \end{pmatrix}$$

Equation 1.45

This equation is equivalent to

$$C_x \cdot v = \lambda v$$

Equation 1.46

The vector v is the eigenvector of the variance-covariance matrix C_x and λ is the eigenvalue. The eigenvalue λ can be determined by

$$v_1^T \cdot C_x \cdot v_1 = \lambda_1 \left(v_1^T \cdot v_1 \right) = \lambda_1$$

Equation 1.47

Maximizing Equation 1.47 means maximizing λ, thereby taking the largest eigenvalue and its corresponding eigenvector.

As stated before, the new variables u_r are the principal components, and as such are uncorrelated linear functions of the original variables. The coordinates of the corresponding eigenvectors are the coefficients of the original variables for the principal component. The loading for the variable is this coordinate multiplied by the square root of the eigenvalue of this principal component. The term loading is also often applied to the coefficients themselves. The value an object takes for a principal component is the score of the object for that principal component.

The next step is to decide how many principal components to retain in the model. One would usually retain enough of the first principal components to account for 80-90% of the total variation. The total variation of all the variables can be determined by:

$$\sum_{r=1}^{m} \lambda_r$$

Equation 1.48

and so the required amount of variance can be determined by

$$\frac{\sum_{r=1}^{p} \lambda_r}{\sum_{r=1}^{m} \lambda_r} \geq 0.8 \text{ or whatever criterion selected}$$

Equation 1.49

Usually only the first two or three principal components are enough to explain an important part of the variance. The variance and hence the error associated with the data will be further discussed below.

Factor Analysis

Taking this discussion further, a determination of how many factors, or principal components, are necessary to explain the variance needs be done. This determination is called factor analysis. Recall that for the two dimensional space, the transformations are (Equations 1.26 and 1.27 respectively):

$$u_1 = \cos\theta x_1 + \sin\theta x_2$$

$$u_2 = -\sin\theta x_1 + \cos\theta x_2$$

Generalizing for *m*-dimensions, the transformation can be written as

$$u = V \cdot x$$

Equation 1.50

and the conditions as

$$V \cdot V^T = I$$

Equation 1.51

Any matrix that satisfies Equation 1.51 above also has the property that its transpose is equal to its inverse:

$$V^T = V^{-1}$$

Equation 1.52

and thus Equation 1.50 can be rewritten as

$$u = V^{-1} \cdot x = V^T \cdot u$$

Equation 1.53

Equation 1.53 can be written out as

$$\begin{aligned} x_1 &= v_{11}u_1 + v_{12}u_2 + \cdots + v_{1m}u_m \\ x_2 &= v_{21}u_1 + v_{22}u_2 + \cdots + v_{2m}u_m \\ &\vdots \\ x_m &= v_{m1}u_1 + v_{m2}u_2 + \cdots + v_{mm}u_m \end{aligned}$$

Equation 1.54

and so, in vector notation,

$$x_i^T = u_i^T \cdot V$$

Equation 1.55

Keep in mind that since

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{i2} & x_{i2} & \cdots & x_{im} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

Equation 1.56

Equation 1.55 therefore can be decomposed into two matrices, multiplied together:

$$X = U \cdot V$$

Equation 1.57

The matrix U are the scores for the data. The first row contains the scores for the first object, the second row for the second, and so on. The entire matrix can be found using multiple linear regression (MLR):

$$U^T = \left(V^T \cdot V\right)^{-1} \cdot V \cdot X^T$$

Equation 1.58

Since the eigenvectors are orthogonal, $(V^T \cdot V) = I$, the identity matrix, and thus

$$U^T = V \cdot X^T \text{ or } U = X \cdot V^T$$

Equation 1.59

Since not all of the eigenvectors need be included in V in order to account for the unextractable noise inherent in the data, and hence in X, only the first several eigenvectors, p, may be required:

$$X = U_p \cdot V_p + E$$

Equation 1.60

where V_p is the matrix of eigenvectors for the first p rows of V, U_p are the corresponding scores, and E is the matrix of errors or residuals, which are the difference between the values the data set gives and what the model gives for the same value of x.

Error

All real data sets contain some error, either random or experimental. The sources of the error are generally ascribed to noise, including both experimental and instrumental. The amount of error associated with the data set can actually mask the true dimensionality of the data set. The total error associated with a given data set can be divided into two sources: imbedded error and extracted error. Extracted error is error that is found within the minor principal components; it can be removed by discarding the minor principal component. Imbedded error, however, is contained within the first several major principal components, and thus cannot be completely removed from the data set. The amount of imbedded error will therefore affect the reproduction of the data set.

While the imbedded error cannot be removed and is therefore included in the reduced data set, the extracted error can and is usually removed. Thus, the correct number of principal components to be retained becomes important; if it is underestimated, then important information concerning the structure of the data has been removed as extracted error, but if it is overestimated, experimental error is included in the principal component model and could cloud or mask the underlying structure contained in the data set. The question to be answered is: How many principal components does one retain?

If the objective is just a quick analysis to get an understanding of the data, then the additional dimensions and their associated error may not be important. However, if one has mixture data and wants to identify the constants and estimate their proportions in

the mixtures, an accurate estimate of the number of dimensions to include becomes important. There are two categories of techniques that have been developed to determine the true dimensionality of the data: those based on knowledge of the experimental error associated with the data and techniques that do not require such knowledge and are thus approximate. The various techniques are summarized below. Among those techniques requiring prior knowledge are:

1. Residual Standard Deviation (RSD). The RSD of a matrix is a measure of the lack of fit between the principal component model and the data set. When the number of dimensions that are required to reduce the RSD to approximately equal to the estimated experimental error of the data, the true dimensionality of the data set is revealed.

2. Root Mean Square (RMS). Like the RSD above, the true dimensionality of the data set is found to be the number of dimensions that are required to reduce the RMS to approximately that of the experimental error. The RMS and RSD are related to one another; however, the RSD measures the difference between the raw and the pure data that contains no experimental error, and the RMS measures the differences between the raw and the reproduced data using a given number of principal component dimensions. The method of RSD is to be preferred, since it gives a measure of the actual error.

3. Average error criterion. The average error is the average of the absolute values of the differences between the raw and reproduced data, with the true dimensionality being the number of dimensions that are required to reduce the average error to be approximately the same as the estimated average error in the data.

4. Chi-squared criterion. For data sets where the standard deviation changes from data point to data point and is not constant, a chi-squared technique has been developed. It takes into account the variability in error from one data point to the next; however, one must be reasonably accurate in the error estimate for each data point.

5. Distribution of misfit. In this method, the number of misfits between the observed and reproduced data sets are studied as a function of the number of principal components employed. Misfits are defined as data points whose deviation from the observed value is more than 3σ, estimated from experimental information. The true dimensionality of the data is the number of principal components required so that either none of the data points, or some user specified proportion of data points, are classified as misfits.

There are methods that do not require prior knowledge of the experimental error in the data set and thus can be used to estimate the amount of error in the data set. These methods include:

1. Cumulative percent variance. This method is a measure of the total variance accounted for by using p principal component dimensions. The sum of the square of the reproduced data is divided by the sum of the square of the raw experimental data, times 100%. It can also be expressed in terms of the eigenvalues of the data matrix, and thus is also known as the percent variance in the eigenvalue. The criterion is to accept the set of largest eigenvalues that account for a specified proportion of the data. However, this method can only be used if an accurate estimate is made of the true variance in the data, and this cannot be done without knowledge of the errors in the data.

2. Average eigenvalue. The principal components whose eigenvalues are above the average eigenvalue is retained and the others are discarded. If the analysis is done using its correlation matrix, the average eigenvalue will be unity, since the variance of each variable is unity. Thus, only those components whose eigenvalues are more than one should be retained, and so this method is also known as the eigenvalue-one criterion.

3. Scree test. This test assumes that the residual variance will level off before those dimensions that contain random error will be included in the data. When the residual percent variance is plotted against the number of principal component dimensions, the curve will drop rapidly and level off at some point. The dimensionality of the data set is shown to be where the curve levels off or where a discontinuity in the curve occurs; successive eigenvalues explain less and less variance in the data and result in a drop in the residual percent variance. In situations where the errors are not random, discontinuities in the plot will be seen, since principal component analysis may exaggerate any non-uniformity in the data in trying to explain the variation.

4. Exner function. The Exner psi (ψ) function can be used as a method to identify the dimensionality of the data set. This function can vary from zero, which gives the best fit, to infinity. A value of 1.0 gives the upper limit of significance; this means that reproducing the data using *p* principal components is the same as saying each point is equal to the mean of the data. The largest acceptable ψ value has been proposed at 0.5. A ψ value of 0.3 is considered a fair correlation, 0.2 is a good correlation, and 0.1 an excellent correlation.

5. Imbedded error function. This function was developed to identify the principal component dimensions that contain error, but without relying on an estimate of

the error associated with the data. The value of the function should decrease as the true dimensions are being used in reproducing the data; when error dimensions are included in the reproduction, the function should increase, because the error dimensions are the sum of squares of the projections of the error points on the error axis. The imbedded error function method, however, has difficulties as a steady increase is rarely seen, as principal component analysis exaggerates any non-uniformity that exists in the error distribution, and hence will result in non-equal error eigenvalues. The minimization in the function may not be clearly seen if systematic errors exist, uniformity in the error distribution in the data is not seen, or the errors are not truly random. In any of these cases, local minima may be seen.

6. Factor indicator function. The factor indicator function appears to be more sensitive than the imbedded error function. It also reaches a minimum when the correct number of principal components have been reached which is more pronounced and can often be seen in cases where the imbedded error function does not show a minimum.

7. Malinowski F-test. This test is based on the Fisher variance ratio test (F-test); the F-test is the ratio of two variances, obtained from two normally distributed data samples. If the errors in the data are not uniform, or if systematic errors exist, the variance expressed by the error eigenvalues will not follow a normal distribution. Since the true eigenvectors for the data contain both structural information and imbedded experimental error, they have eigenvalues that are statistically greater than the pooled variance of the error eigenvalues. The error of the first eigenvector is set equal to the principal component with the smallest eigenvalue. The next smallest eigenvalue is tested

for significance by comparing its variance to the variance of the error set; if this quantity exceeds the tabulated value of F for the chosen level of significance, it marks the division between the true and error eigenvectors. If it does not exceed the tabulated value of F, the eigenvalue being tested is added to the error eigenvector set and the process is repeated with the next smallest eigenvalue until the division line is found. The advantage of the F-test is that it takes into account a model for the noise or errors in the data, even if the model is a normal distribution.

8. Cross-validation. The cross-validation method tries to determine which principal components have the best predictive ability, and those that contain experimental and random errors as do the above methods. It does not identify the breaking point for the dimensions that contain sample structure and information. Cross validation is used in the formulation of the principal component model for samples from a known class, and then the principal component model is used to classify unknown samples. The data matrix X is divided into a number of small groups; each group is deleted in turn from the data and a principal component analysis is performed on the remaining data set. The deleted group is then predicted using the model just found. The process is then repeated until all the groups have been subsequently deleted and predicted. A suitable criterion of a goodness of fit is selected, relating the actual and predicted values, and summed over all the groups; the optimal principal component model is then identified using the goodness of fit criterion. This method is used when a large number of possible dimensions are inherent in the data matrix. In the current investigations, there are only four possible dimensions, and thus cross-validation as normally used is not necessary, as examination of the number of principal component analysis will reveal.

In the present work, three different algorithms were used in conjunction with principal component analysis: partial least squares (PLS), and principal component regression (PCR) using the traditional NIPALS (*N*on-*I*terative *Pa*rtial *L*east *S*quares) and the SIMPLS (*S*tatistically *Im*proved *P*artial *L*east *S*quares) algorithms. Each of these methods are a two-step procedure, with principal component analysis being performed first to obtain the reduced number of variables and the factor scores, and the subsequent algorithm is used with these factor scores to determine the analyte concentration or amount. Each of these methods and their respective algorithms will be discussed in turn.

Principal Component Regression.

Principal Component Regression (PCR) methods do not require that details of all the components in the system be known, but does require a sensible estimate of how many significant components are in the mixture.[109] A normal regression technique relates the original variables to the concentrations; PCR relates a principal component to that concentration. This method is a solution to the factor analysis problem discussed above, giving:

$$t = Xw + e$$

Equation 1.61

where t is a column vector containing n_1 rows of t_1 followed by n_2 rows of t_2, where t is the target vectors for the first and second components, respectively, and X is a vertical concantenation of data matrices X_1 and X_2, where each column is the reading from an

individual detector in the detector array. The component *w* is the weight vector. The quantity *e* is the error vector, which is minimized to solve the regression problem. The solution is

$$w = \left(X^T X\right)^{-1} X^T t$$

Equation 1.62

If the measurements are highly correlated or are noisy, finding a good value of the weight vector can be difficult because of the inverse function in Equation 1.51. However if principal component analysis is first performed on the data, a reduced data set can be determined, and directions that correspond to noise in the system can be discarded. The predicted values can now be found using the reduced data set; the projected data are commonly called the "score matrix." This approach to the problem is called principal component regression, and as an alternate solution to Equation 1.51 may be better behaved than the standard MLR approach.[110]

Partial Least Squares.

Partial Least Squares (PLS) is related to PCR. While PCR finds factors that capture the greatest amount of variance in the predictor variables, and MLR (discussed previously) tries to find a single factor that best correlates predictor variables with predicted variables, PLS tries to do both: it tries to capture the greatest amount of variance and it tries to achieve correlation.[111] It requires no prior knowledge of the system under investigation, and is applicable to non-linear systems.[112] It is often

regarded as the major regression technique for multivariate data.[109] When applied to chromatographic data, this method takes into account errors in both the concentration estimates and the chromatograms, instead of only in the dependent data (the chromatograms).[113] This method, however, must be used with caution, as the predictions are essentially statistical in nature. The underlying factors in the method have little or no physical meaning, and so a linearly additive model where each underlying factor can be physically interpreted is not expected. Even with these problems, if future samples contain similar features as the original data, the model can be very robust.

PLS uses both the target vector and the measurements to determine the lower dimensional subspace, using an iterative procedure.[110] This subspace is then used to make the predictions. Two commonly used algorithms are the traditional NIPALS, for *N*on-*Iterative Par*tial *L*east *S*quares, and its relative, SIMPLS (*S*tatistically *Im*proved *P*artial *L*east *S*quares). NIPALS carries the x variance information in the loads, not the scores; the X-block (predictor variable) scores are orthogonal to one another and the X-block loadings are normalized. SIMPLS has the X-block scores being orthonormal, i.e., both orthogonal and normalized, and the X-block loadings are not normalized. In fact, the explicit objective of the SIMPLS algorithm is to maximize covariance, while NIPALS does not. SIMPLS relies on orthogonalization of a Krylof sequence to calculate the PLS weights, which is a very fast procedure.[111]

Each of the methods under consideration rely on principal components as the first step in the process. Thus, one of the parameters to be determined was the number of principal components required to adequately describe the data. Two commercial packages have been used to determine this parameter, Matlab (The MathWorks Inc.,

Natick NY) and a module to Matlab, PLS_Toolbox 2.1 (Eigenvector Research, Inc., Manson, WA). These packages have been used in the investigation of antibiotics in animal tissues,[38] the determination of acetonitrile and ethanol in water,[25] the characterization of impregnated materials,[29] the determination of aluminum and iron in plant extracts,[114] the amount of dermatan sulfate contamination in heparin samples,[23] the determination of the strongly overlapping peaks of ebrotidine and its metabolites in capillary electrophoresis,[27] and in standardization techniques for ion-selective sensor arrays.[115] Other investigations include the determination of tryptophan in feed samples,[116] and the characterization of dental composites.[117] Matlab, with its associated functions and toolboxes, has also been used in the analysis of voltammetric data,[46] and the determination of the hydrocarbon content in naphtha by MIR spectroscopy.[118]

There are two ultimate criteria to consider in determining the applicability of any method, including chemometric methods, for determination of quantities of compounds: the accuracy and the precision. Accuracy is defined simply as how well a result compares to the known, real value; the precision of a method is a determination of the spread in the data. The determined result may be very accurate, but the precision may be so low that the result itself may be called into question. When there are replicates in the theoretical data set for each determination, and the values for each determination is measured, accuracy and precision are easily determined. The issue of accuracy is addressed by two methods. One is the use of a commonly employed diagnostic RMSEC, the root mean square error of calibration:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{n}}$$

Equation 1.63

where n is the number of samples, $\hat{y}_i$ are the predicted y-values when all samples are included in the model formation, and y_i are the actual values. The percent RMSEC is then determined by

$$\%RMSEC = RMSEC\frac{100}{\bar{y}_i}$$

Equation 1.64

where $\bar{y}_i$ is the mean of the y-values. A similar criterion is used to determine the goodness of fit of the predicted values, which uses an equation of the same form, but uses the predicted values from a new data set, not the calibration data set, and is termed the root mean square error of prediction, RMSEP.

Another simpler criterion used was the residual percentage, which is simply the difference between the known value and the determined value, expressed as a percent of the known value. This value can be considered also as a difference percentage. A positive residual percentage means that the determination is higher than the known value; a negative residual percentage means that the determination is lower than the known value. Additionally, the difference percentage will be examined using clustering analysis

to determine which data set is closest to the model set. The issue of precision is expressed as a percent relative standard deviation (percent RSD).

Development of the Cramer's Rule Approach

The Cramer's Rule Approach depends on the responses of the individual electrodes to a compound or compounds at a given time. It is only used on the potential-signal domain. Thus, it requires a description of the individual currents of the analytes. We start with the Nernst equation restated here in terms of the concentrations of the reduced and oxidized species in a redox reaction:

$$E = E^{o'} - \frac{0.05916}{n} \log \frac{C_{R(x=0)}}{C_{O(x=0)}}$$

Equation 1.65

where "x = 0" denotes distance from the electrode surface. Rearrangement of Equation 1.65 yields:

$$C_{R(x=0)} = C_{O(x=0)} 10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}$$

Equation 1.66

for the reduced species and

$$C_{O(x=0)} = \frac{C_{R(x=0)}}{10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}$$

Equation 1.67

for the oxidized species. The model also requires an expression for the concentration in terms of the reduced species in the bulk solution. With only the reduced species in the bulk of the solution, at the electrode surface we have

$$C_{O(x=0)} + C_{R(x=0)} = C_R^*$$

Equation 1.68

Prior to any potential being applied, at time zero (t=0), i.e., no reaction has taken place yet, $C_{O(x=0)} = 0$ and so $C_{R(x=0)} = C_R^*$. Substituting Equation 1.67 into Equation 1.68 yields

$$\frac{C_{R(x=0)}}{10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}} + C_{R(x=0)} = C_R^*$$

Equation 1.69

which reduces to

$$C_{R(x=0)}\left[\frac{1}{10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}} + 1\right] = C_R^*$$

Equation 1.70

The concentration of the reduced species at the surface of the electrode thus can be written as a function of the concentration of the bulk reduced solution, the applied potential, the standard potential, and the number of electrons in the balanced chemical equation:

$$C_{R(x=0)} = \frac{C_R^*}{\left[\frac{1}{10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}+1\right]}$$

Equation 1.71

which is further simplified to

$$C_{R(x=0)} = \frac{C_R^*}{\left[\frac{1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}{10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}\right]} = \frac{C_R^* 10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}{1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}$$

Equation 1.72

The model also requires an expression for the current in terms of the reduced species in the bulk solution. This is expressed as a function of the moles of electrons involved in the reaction, the Faraday constant, the area of the electrode, and the mass constant:

$$\frac{i}{nFA} = m_R\left[C_{R\,(x=0)} - C_R^*\right]$$

Equation 1.73

Equation 1.72 is then substituted into Equation 1.73, yielding:

$$\frac{i}{nFA} = m_R\left[\frac{C_R^* 10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}{1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}} - C_R^*\right]$$

Equation 1.74

or

$$i = nFAm_R\left[\frac{C_R^* 10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}{1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}} - C_R^*\right]$$

Equation 1.75

which simplifies to:

$$i = nFAm_R C_R^*\left[\frac{10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}{1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}} - 1\right]$$

Equation 1.76

This terms in the bracket simplify and the equation becomes

$$i = nFAm_R C_R^* \left[\frac{10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]} - \left(1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}\right)}{1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}} \right] = nFAm_R C_R^* \left[\frac{-1}{1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}} \right]$$

$$= \frac{-nFAm_R C_R^*}{1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}$$

Equation 1.77

Equation 1.77 above is in terms of the concentration of the reduced species in the bulk solution. Since the purpose of this study is to deconvolute the overlapping peaks and not to determine constants related to the electrode, Equation 1.77 can be simplified by collecting the constants together:

$$K_R = nFAm_R$$

Equation 1.78

and so we use the simplified equation

$$i = \frac{-K_R C_R^*}{1+10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}}$$

Equation 1.79

to model the current at a given electrode for a single species. Since the potential of the electrode is set is also a constant, and the number of electrons involved in the reaction as well as the standard potential are also constants, we can further combine these into the overall constant yielding

$$i = -K_R^{'} C_R^*$$

Equation 1.80

Due to the conventional nomenclature used, Equation 1.80 above appears to yield a negative current; in practice, however, the current is positive, and hence further development of the model using Equation 1.79 will show the results as positive with the K'_R now being $K^{\ddagger}_R$.

For the oxidized species, we use Equation 1.66, repeated here:

$$C_{R(x=0)} = C_{O(x=0)} 10^{-\left[\frac{\left(E-E^{o'}\right)n}{0.05916}\right]}$$

Equation 1.66

and the bulk equation for the oxidized species:

$$C_{O(x=0)} + C_{R(x=0)} = C_O^*$$

Equation 1.81

along with the current equation for the oxidized species

$$\frac{i}{nFA} = m_O\left[C_O^* - C_{O(x=0)}\right]$$

Equation 1.82

A similar treatment as above using the appropriate equations yields

$$i = nFAm_O C_o^*\left[\frac{10^{-\left[\frac{n(E-E^{0'}}{0.05916}\right]}}{1+10^{-\left[\frac{n(E-E^{0'}}{0.05916}\right]}}\right]$$

Equation 1.83

for the current equation for the oxidized species. Using the corresponding equation to collect the electrode parameters into a single constant, and, since the potential is fixed during the run for the electrode, and the number of electrons involved is constant as well as the standard potential, the constant for the oxidized species takes on the form of:

$$K_O = nFAm_O\left[\frac{10^{-\left[\frac{n(E-E^{0'}}{0.05916}\right]}}{1+10^{-\left[\frac{n(E-E^{0'}}{0.05916}\right]}}\right]$$

Equation 1.84

gives the final equation as

$$i = K_o^{'} C_o^*$$

Equation 1.85

for the oxidized species. Equations 1.80 and 1.85 above imply concentrations; however, the amount injected could just as easily be used since the respective constant collects up all the respective terms.

To finish the model, the total signal is taken as a sum of the individual currents for each potential in the time domain:

$$i_{tot,Ch1} = \left(K_R^{\neq} C_R^*\right)_{Ch1,Cmpd1} + \left(K_R^{\neq} C_R^*\right)_{Ch1,Cmpd2} + \cdots + \left(K_R^{\neq} C_R^*\right)_{Ch1,Cmpd(n)}$$

Equation 1.86

shown here for the reduced species in bulk. For each electrode, there is a separate equation. This set of equations form the simultaneous linear equation set that is to be solved to determine the concentration of the respective species. Thus, four electrodes would involve four equations:

$$\begin{aligned}
i_{tot,Ch1} &= \left(K_R^{\neq} C_R^*\right)_{Ch1,Cmpd1} + \left(K_R^{\neq} C_R^*\right)_{Ch1,Cmpd2} + \cdots + \left(K_R^{\neq} C_R^*\right)_{Ch1,Cmpd(n)} \\
i_{tot,Ch2} &= \left(K_R^{\neq} C_R^*\right)_{Ch2,Cmpd1} + \left(K_R^{\neq} C_R^*\right)_{Ch2,Cmpd2} + \cdots + \left(K_R^{\neq} C_R^*\right)_{Ch2,Cmpd(n)} \\
i_{tot,Ch3} &= \left(K_R^{\neq} C_R^*\right)_{Ch3,Cmpd1} + \left(K_R^{\neq} C_R^*\right)_{Ch3,Cmpd2} + \cdots + \left(K_R^{\neq} C_R^*\right)_{Ch3,Cmpd(n)} \\
i_{tot,Ch4} &= \left(K_R^{\neq} C_R^*\right)_{Ch4,Cmpd1} + \left(K_R^{\neq} C_R^*\right)_{Ch4,Cmpd2} + \cdots + \left(K_R^{\neq} C_R^*\right)_{Ch4,Cmpd(n)}
\end{aligned}$$

Equation 1.87

where the amounts or concentrations of the respective species are expected to remain the same. This method was previously outlined but not described fully by Cao and Hoshino[119] in 1996; they showed that the method worked, and used it to detect 3,4-dihydroxyphenylglycol (DHPG).[120] The chromatographic traces for the analytes shown in both papers indicated a discernable peak, with a minimum in the chromatographic trace both before and after it. This is a case that has been addressed many times by fitting the data using the time/signal domain. They calculated the amounts for the analyte on a point-by-point basis over all the elution time points using a least-squares method, but did not indicate any problems with possible column overloading. This method should work as long as the potentials used to differentiate between the analytes are different, since the separation power of the method depends on the different electrochemical behavior of the analytes.

Chapter 2 — Experimental

NEUBA® Chromatographic System

NEUBA®, a "neurobiological analyzer",[121-126] was developed in the late 1980's for the analysis of complex mixtures such as neurochemicals in brain tissue and metabolites in urine. Multiple electrode detectors such as those in NEUBA® have recently found use in various applications, from the analysis of beer and wine[127] to mapping regions of the brain,[128] and the use of multi-channel and multi-electrode systems are becoming more popular.[129-136]

NEUBA® is a liquid chromatograph system having multiple columns and multiple electrochemical detectors. It has four individual chromatographic columns, with thirteen working electrodes arranged in four thin layer cells, all controlled by a single potentiostat and using one autosampler to make injections. Of the four columns, three are used to separate most of the desired analytes by dividing them from a simply chromatographic standpoint into weakly, moderately, and strongly retained components. The fourth column is devoted to the separation and detection of acetylcholine and choline by use of an immobilized enzyme post column reactor.

The three primary systems use an electrochemical array detector consisting a series of four 1.0 mm glassy carbon electrodes in one thin layer flow through cell. System 4, however, has a single 2.0 mm platinum working electrode. The three-dimensional nature of the multiple detector LCEC separation thus allows for identification and quantification of potentially hundreds of compounds.

NEUBA® consists of a computer interfaced to both an autoinjector and the multi-electrode potentiostat/controller. Each sample is injected and a portion directed through a six port column selection valve, which then sequentially injects a set amount to each of the four analytical columns. A detector is located immediately at the output of each column; to help minimize electronic noise, the working electrodes of each detector are contained in a Faradaic cage. The potentiostat/controller controls the applied potential and monitors the current at each of the working electrodes. Control of the six-port selection valve and the autoinjector is routed through the potentiostat/controller from the computer. The solvent delivery system has a number of conventional LC connections to the six port valve. There are also a series of flow-through by-pass lines which regulate the solvent flow through the column selection valve. When the valve is not turned so as to allow flow through it for a particular system, the by-pass lines allow solvent to continue to flow through the affected system.

Figure 2-1 below shows the flow diagram for the NEUBA® system. The heart of the system is the six-port tandem injection valve, marked 7066A and 7066B. This tandem rotary valve is used to direct the required mobile phase through the autoinjector and into the appropriate column. Flow continues in the other systems through the by-pass lines. The valve then switches to position 2 for the prescribed length of time as determined by the autosampler program and flow continues through the other systems, again, through their by-pass lines. The sequence continues all the way around the valve. Narrow bore tubing (0.005") begins from the point of the autosampler through to the exit from the electrochemical cell. Stainless steel tubing is used from the pumps throughout

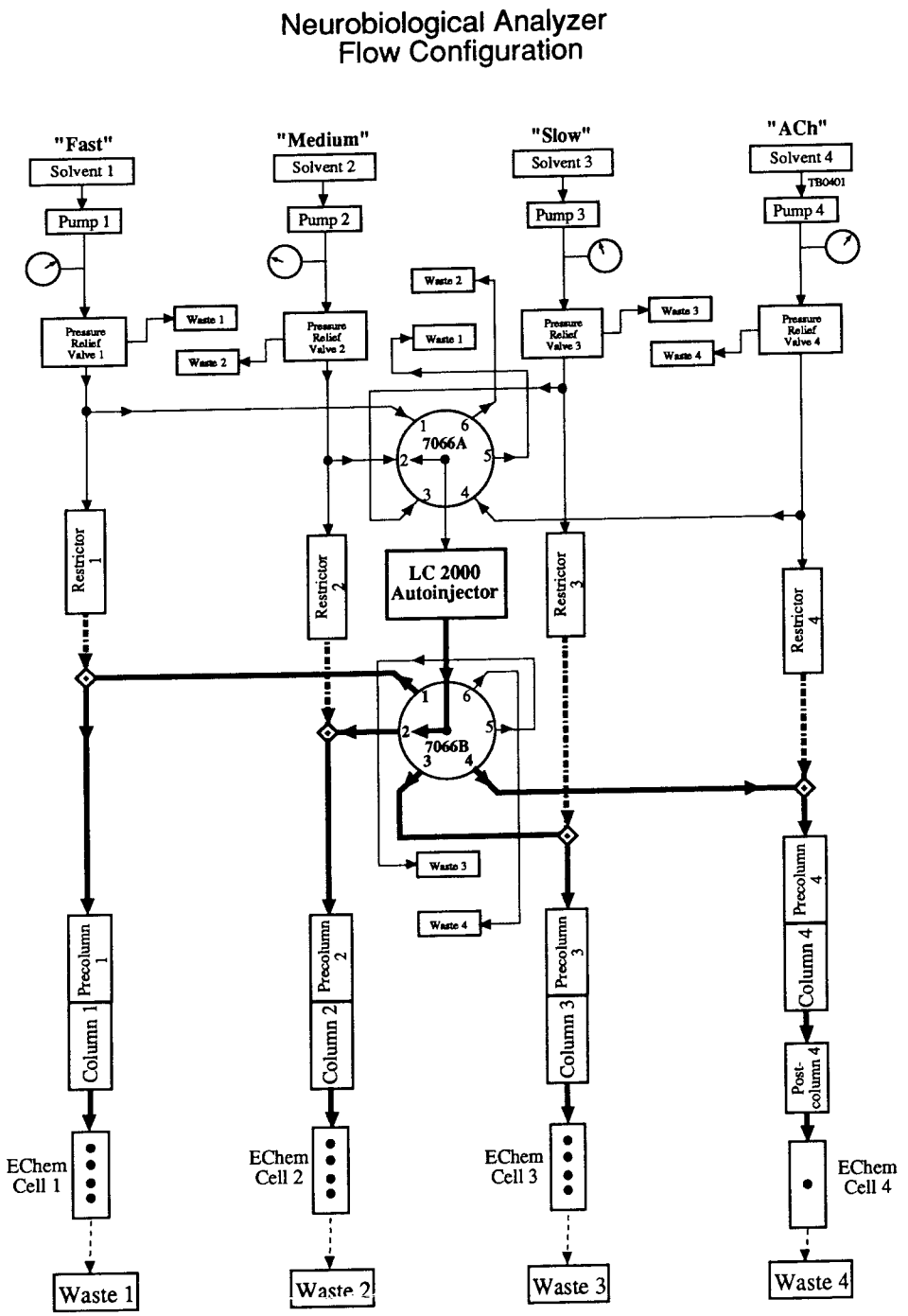


Figure 2-1. Flow diagram for the NEUBA® system.

From: D.J. Turk, "Development and Utilization of a Multi-Column, Multi-Electrode Detector for the Simultaneous Determination of Multiple Neurochemical Species," Ph.D. Dissertation, University of Oklahoma, Norman, OK 1991.

the rest of the system, with the exception of the tubing from the analytical columns to the electrochemical cells. This tubing, although narrow-bore, is teflon tubing to ensure electrical isolation of the cell block assembly and still minimize dead volume in the system.

The components of NEUBA® are, for the most part, commercially available. The PM-48 solvent delivery systems are from Bioanalytical Systems; the QUAD electrochemical detectors were purchased commercially (special order), and the autoinjector unit is a Dynatech LC2000. Pressure gauges for monitoring system back pressures are also installed. The potentiostat/controller, however, was manufactured in-house. Components were placed to minimize dead volume associated with the LC tubing connections that would interfere with the analysis.

Multiple electrode cells potentially can use the selectivity of the electrodes to resolve overlapping peaks into different chromatogram. NEUBA® uses a series configuration for all four working electrodes on each main electrochemical cell.

Any cell that uses more than one electrode set close together, whether the cell is configured as a wall-jet, parallel, or serial, is subject to crosstalk between the electrodes. Furthermore, if the downstream electrode is set to a lower potential than the upstream one, it is subject to interference from the upstream electrode. The downstream electrode has been shown to provide complex biphasic peak patterns, with a negative peak corresponding to the positive peak of the upstream electrode, followed by a positive peak.[137] The electrodes used in this application of NEUBA® were set from a low potential successively to higher ones, and thereby eliminated this possible problem.

NEUBA® is interfaced to a 32 bit 16 MHz Macintosh IIx® computer system, with a 68020 processor with 5 megabytes of RAM. The software used in the acquisition and initial analysis was specially written for the system. The electronics for NEUBA® were either assembled in the laboratory using standard procedures, with readily available chips and electronic parts, or were commercially purchased, depending on the requirements.[121]

For this project, only one chromatographic system was used, and a 1.0-mm four-electrode glassy carbon cell used in the detector assembly, yielding a simplified instrument. The work-up of the data, however, would still be applicable to using additional systems. Also, during the course of the investigations, the autosampler had to be discontinued, and all subsequent injections made by hand using a manual injector (Rheodyne Model 7010, Rheodyne LLC, Rohnert Park CA).

Materials Used

All aqueous solutions used 18 MΩ distilled/deionized water, prepared by passing distilled water through a Milli-Q Reagent Water System (Millipore, Continental Water Systems, El Paso TX). The solutions made were stored in a Revco Model ULT1985-B-J-K ultra-low freezer (Rheem Refrigeration Products Division, West Columbia SC), and was maintained at –80^{o}C. Mobile phase solutions were filtered through a Millipore membrane filter (Millipore, Bedford MA). All reagents used in this work were obtained from commercial sources in the highest available purity (≥98%) and used without further purification.

The following reagents were used:

3-hydroxytyramine, hydrochloride salt (3,4-dihydroxyphenehylamine, dopamine); Sigma Chemical Company, St. Louis MO

ethylaminediaminetetraacetic acid, disodium salt; Sigma Chemical Company, St. Louis MO

L-3,4-dihydroxyphenylalanine (L-dopa, levodopa, DOPA); Aldrich Chemical Company, Milwaukee WI

sodium hydroxide, Mallinckrodt Analytical Reagent, Mallinrockrodt Baker Inc., Phillipsburg NJ

epinephrine hydrochloride (adrenaline, Epi); Aldrich Chemical Company, Milwaukee WI

citric acid, monohydrate, ACS Reagent Grade; Sigma Chemical Company, St. Louis MO

diethylamine; Sigma Chemical Company, St. Louis MO

octyl sulfate, sodium salt; Sigma Chemical Company, St. Louis MO

acetonitrile, HPLC grade; Fisher Scientific Company, Fair Lawn NJ

Column. The column used in this study was packed in house using a downward feed slurry packer. The 10 cm x 4.6 mm i.d. column was packed with C18 3μm Adsorbosphere (Alltech Associates Inc., Deerfield IL) at 6710 psi for 25 minutes. After packing, the column was rinsed while still on the column packer with about 50 mLs of distilled water followed by about 200 mLs of mobile phase prior to connecting the column to the chromatographic system. The column was further allowed to equilibrate for at leaset 24 hours prior to use in an analysis.

Mobile Phase. The mobile phase used in this study was made according to the following procedure. The mobile phase was de-bebubbled by stirring for one hour prior to use. When it was placed on the system, the first 500 mLs of the mobile phase pumped through the column was discarded, thereby allowing for equilibration of the column and the components of the mobile phase.

To 1850 mL of distilled water add with stirring:

Amount	Final Concentration
0.25 – 0.50 g NaOH	
37.22 mg $Na_2H_2EDTA•2H_2O$	0.05 mM

After dissolving, add:

42.02 g citric acid monohydrate	0.10 M
1.7 mL diethylamine	0.06% w/v
118.45 mg sodium octyl sulfate	0.255 mM

Adjust to pH 2.45 using concentrated NaOH solution

Filter through 0.20 μm type FA Millipore filter

Add this solution to 150 mL acetonitrile, prefiltered	7.5% v/v
using a 0.20μm type FH Millipore filter	

Standards

Epi. 3.34 mg of Epinephrine were dissolved in distilled water in a 10.00-mL volumetric flask to make a 1.002 mM working solution. This solution was then divided

into approximately 1-mL portions and stored in 1-mL polyethylene vials at –80°C until just prior to use. When needed, vials were thawed, and the standard solutions were made according to Table 2-1 below using distilled water in 100.00-mL volumetric flasks. The standards were then divided as before and again stored at –80°C. The appropriate number of samples and mixes were thawed just prior to running on NEUBA®.

Dopa. 1.97 mg of 4-hydroxylphenylalanine (L-Dopa) were dissolved in distilled water in a 50.00-mL volumetric flask to make a 1.00 mM working solution. This solution was then divided, stored, and later working standards were made as shown in Table 2-1 in the same manner as for Epi.

Dopamine. Dopamine was used as an internal standard in the study. 1.92 mg of Dopamine•HCl were dissolved in distilled water in a 10.00-mL volumetric flask to make a 1.01 mM working solution. This solution was then divided and stored as described previously, and the working standards were made as shown in Table 2-1.

Mixture Sets. Standard solutions were made as shown below in Table 2-1. Two standards, one containing Epi and Dopamine, and the other containing Dopa and Dopamine, were made as check standards for the duration of the experiment. The mixtures contained all three compounds in the amounts shown below.

The resultant solutions were injected using a 20.00-μL sampling loop on the manual injector, resulting in the amounts injected on column as shown below in Table 2-2; these are the amounts recognized as the true values for the deconvolution procedure. To enable comparison to data and papers previously mentioned, the same mixtures are presented in pg amounts.

Table 2-1: Amounts of working standards used to make the standards and mixtures

	Amt Epi (μLs)	Amt Dopa (μLs)	Amt Dopamine (μLs)
Epi Std	171	0	231
Dopa Std	0	333	231
Mix 1	154	33.3	231
Mix 2	119.8	99.8	231
Mix 3	85.5	166.2	231
Mix 4	51.3	233	231
Mix 5	17.1	300	231

Table 2-2: Resultant amounts regarded at true values for standards

	Amt Epi		Amt Dopa		Amt Dopamine	
	pmol	pg	pmol	pg	pmol	pg
Epi Std	68.5	0.206	0	0	92.6	0.488
Dopa Std	0	0	133.	0.677	92.6	0.488
Mix 1	61.7	0.185	13.3	0.067	92.6	0.488
Mix 2	48.0	0.144	40.0	0.203	92.6	0.488
Mix 3	34.3	0.103	66.6	0.338	92.6	0.488
Mix 4	20.5	0.062	93.4	0.474	92.6	0.488
Mix 5	6.85	0.021	120.	0.610	92.6	0.488

Experimental Conditions.

Two pure standard solutions of the neurotransmitters Epi and Dopa were made to known concentrations, and these two standards were subsequently analyzed on NEUBA®. The experimental conditions used for the electrodes are shown in Table 2-3 below.

The mobile phase composition was sufficient to allow for the elution of the two compounds within one minute from the solvent front. The amounts of the compounds were as shown in Figure 2-2: 68.5 pmol for Epi, 133. pmol for Dopa, and 93.6 pmol for Dopamine. The standard injections were all made from the same working solutions, and were not independent samples. Thus, there is no information concerning variance in concentrations. The mobile phase was run at a rate of 1.2 mLs min^{-1}.

Table 2-3: Electrode Potentials

Channel	Potential (mV)
1	600
2	700
3	800
4	900

Experimental Data

The five mixtures as described were comprised of the three compounds Epi, Dopa, and Dopamine. The ratios used were based on the amounts previously used for the pure standards for Epi and Dopa. The experiment was designed such that up to twice the actual amounts of the standards used would not cause the system to switch to a higher

attenuation level and thereby cause a signal spike on the system which results in difficulties in measuring the signal.

Six replicates for each mixture set were injected into NEUBA® in the following manner: three injections of Epi were run followed by three injections of Dopa, then six injections of the mix were run. Each mix set was bracketed by the three injections of Epi and three injections of Dopa, and the entire run was ended with three injections of Epi and three of Dopa. The calculated amounts injected on column are shown above in Table 2-2. These are the values used in the determinations of the accuracy of the deconvolutions.

The experimental data consisted of standard mixtures of Epi and Dopa, with Dopamine as the internal standard. As the chromatographic peaks did not contain valleys or other visual indications of more than one analyte present, the resolution between the two analytes was calculated from their respective standards. The resolution between the two peaks was measured at 0.13, indicating severe overlap between these peaks.

Preprocessing of the data.

In many cases, the chromatographic traces do not have the same baselines, and thus baseline corrections are commonly performed. Also, it is common that the injection starting and ending times are not exactly the same, and thus, for any method that relies on exact retention times this problem must be addressed. An initial examination of the data showed that preprocessing of the data was indeed necessary before any deconvolution could be attempted. First, the baselines were not consistent between the injections or even between the responses from the electrodes within a single injection. Secondly, the

samples were injected manually on the system. Thus, each injection started at a slightly different time. Both of these problems would have to be addressed before any additional deconvolution could be attempted.

Determining baseline correction. All the data from NEUBA® had offset baselines which allow visualization of the signals from each of the four electrodes independently from one another. The actual offsets, or baselines, are variable in nature. Additionally, baseline drift during the course of the injection was present. Therefore, each injection required a baseline correction. In many cases, the baseline correction was determined to be not a single offset number, but rather a linear function with time. Three areas in the chromatographic run were selected for pivot points in the data: at the beginning of the chromatographic run, after the solvent front, and at the end of the run. These areas are short regions where the data has a slope of approximately zero, and are located just before the solvent front or the analyte peak. The number of points in time that were selected were sometimes variable but never less than 20 points, corresponding to a two-second time window.

A baseline was determined in the form of $y=mx+b$, where y is the signal and x is the time, underneath the part of the chromatogram under consideration, as discussed above. The signal associated with the baseline was subtracted from the total signal, thereby yielding a baseline corrected signal. A plot of the baseline corrected signal and the time therefore yields a baseline corrected peak that should begin and end at zero. However, this baseline is a relative baseline, and the valley between the analyte peak and the internal standard peak in many cases did not reach this relative baseline. All further data manipulations were done on the baseline corrected data.

Time Corrections to the Data. There is known variance in the data set. Of particular interest was in the timing of the injections. As the data was generated using a manual injector, the actual timing from injection to injection can range by several seconds, even though every attempt was made to ensure reproducible injection times while manually injecting the samples. Additional hidden variance was found as well; the method used to eliminate as much known variance as possible is described below.

Run-to-run retention time shifting has been seen and reported previously as a severe impediment for separation techniques.[138, 139] This problem arises due to fluctuating instrumental parameters such as flow rate and temperature. Fraga et al. had devised a complex alignment method using only the data matrices of overlapped peaks of interest to minimize the pseudorank of the two matrices and thereby perfectly align the data; however, this works only for replicates of the data. When two closely eluting peaks vary in amounts, the resultant single peak will shift in time due to the amounts in the two analytes and not due to fluctuating instrumental conditions or parameters. Aligning only the data containing the analyte peaks will cause an artificial shift in the data unrelated to experimental realities. To ensure this does not happen, a wider range of data needs to be examined, as discussed below.

A representative injection is shown below in Figure 2.2. The first analyte peak is the analyte peak of interest, either Epi or Dopa. The second peak is Dopamine, which was added to each sample prior to injection as an internal standard. Since each solution injected into the system contained 10^{-4} M hydrochloric acid, the physical properties of the injected standard solution are different from those of the mobile phase. Hydrochloric acid is not retained on the column, and this difference in physical properties causes a

deflection in signal as the unretained species flow across the electrode surface, as can be seen in Figure 4-5 below at between 0.6 minutes and 1.0 minutes. The deflection is reproducible for each of the four channels, giving the same pattern of peaks and valleys respective to the channels where the data was acquired, although not always as intensely as previous injections, and can be termed as the solvent front peaks. Upon examination of all of the injections for each channel, a feature common in the deflection to all the samples on all four channels was chosen as a point to start the time of the analytical run, and was denoted as t_0. It is from this point in time that the retention times of all peaks of analytical interest will be measured, and is the time whereupon the analysis is considered to start.

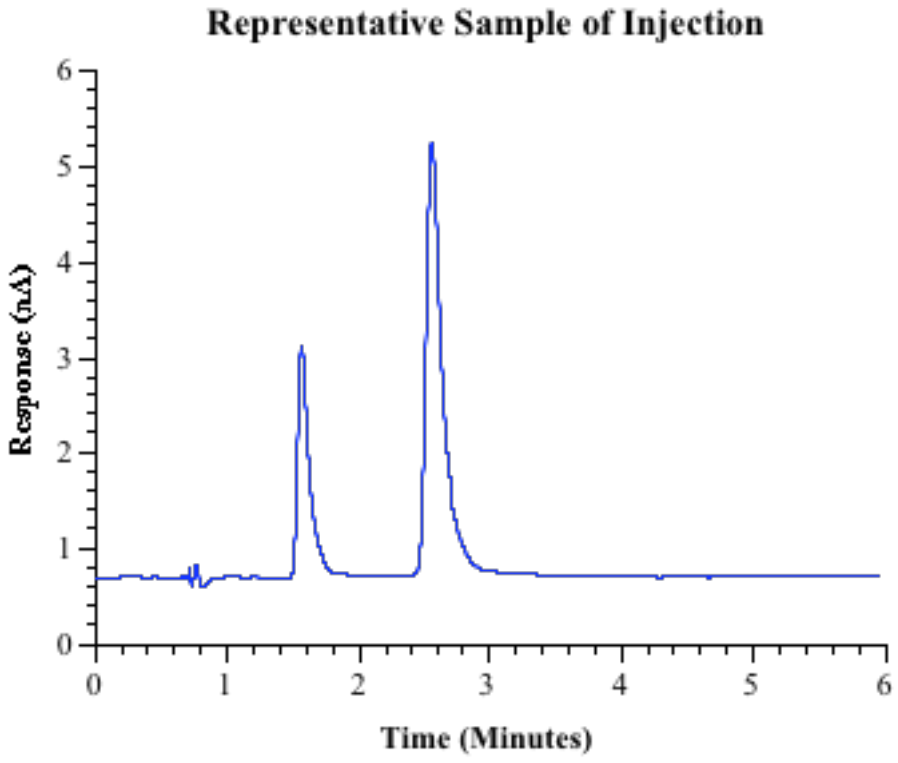


Figure 2-2: Representative Injection

Determination of time t_M. The solvent front pattern is complex, with several peaks, both positive and negative, as shown in Figure 2.3. This can be understood in terms of the conductivity change due to the ionic compounds in each sample solution passing across the electrode surface. Additionally, any interruption in the flow of the mobile phase will also be detected by the electrochemical cell by a slight deflection, and can be seen at approximately 0.22 minutes in Figure 2.3 below. This feature also consistently appeared in each of the injections, but the use of this slight bump in the baseline to signal the beginning of the chromatographic run did not result in consistent retention times of the compounds.

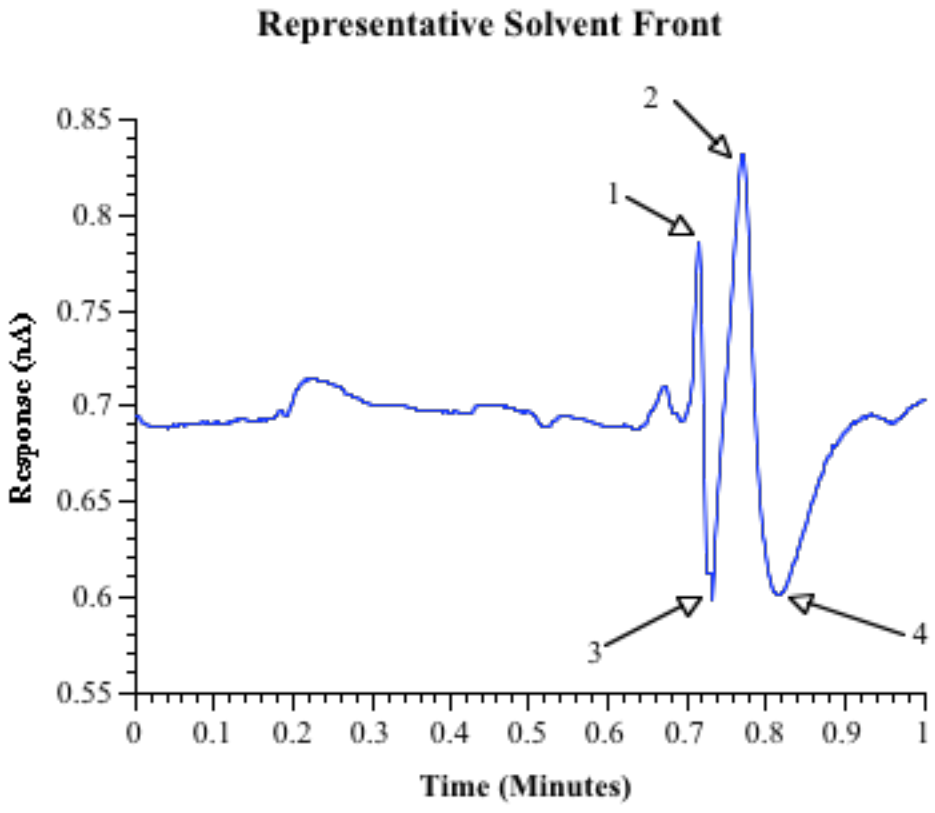


Figure 2-3: Solvent Front Pattern

An examination of all samples yielded a common solvent front pattern. This pattern, shown in Figure 2.3, are labelled numerically and are the pertinent regions of interest as discussed below. The first feature in the solvent front, labeled as 1, was the one initially chosen for the beginning of the solvent front, since it appeared to be the one most consistent from injection to injection. The data corresponding to this peak was first fitted using Excel with a minimization of a sum-of-squares of the residual approach, fitted using a binomial function. The number of data points used in the fitting became a concern, since Excel does not interpolate the function between the data points. The data was reanalyzed and fitted for a second order polynomial and a third order polynomial in DeltaGraph (DeltaPoint, Inc., Monterey, CA) to get a more exact fit to the equation than Excel could provide. The curves were then analyzed using the software package Maple (Waterloo Maple, Inc., Waterloo, Ontario Canada), where the first derivatives were determined allowing the exact retention times to be found analytically. The data was then compiled again in Excel, and the median time of occurrence for the peak was determined for each injection, the difference in the time of the appearance of feature 1 (Δt) and the median time (t_m) for all examples of feature 1 was calculated. Then each injection was adjusted by addition or subtraction of Δt from t_m, thereby forcing feature 1 to coincide for all the injections. When this was done, it was found that features 2 through 4 in the unretained solvent front also coincided to a very high degree with one another on all the injections. This entire pattern of peaks (feature 1 through feature 4) encompasses the solvent front. Other peaks within the solvent front could have reasonably been chosen as well. In this case, the other peaks were not always as easily defined as shown here; feature 1 was chosen because of it's consistency and ease of

measurement. Prior to the determination of time t_0, the raw retention times of the respective analytes were not reproducibly predictable.

Adjustment of the data for the non-retained species After adjusting the data to the non-retained species, however, the apparent randomness in the data only somewhat decreased, thus revealing itself to be due to differences arising not in the actual injections from the manual injection itself, nor from any property of the analyte.

After the retention times were corrected using the non-retained species, the following pattern emerged for Epi and Dopa in Figures 2.4 and 2.5, respectively. These figures show the retention time data prior to any corrections being made, after the solvent

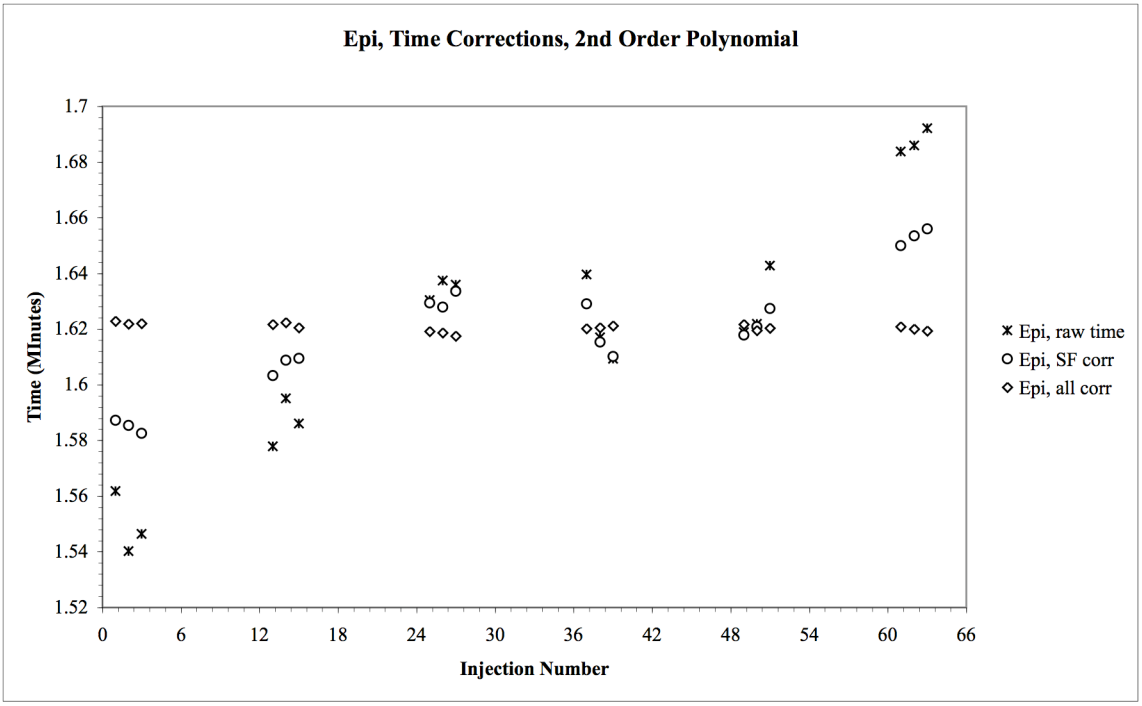


Figure 2-4: Retention time of Epi as a function of time corrections

front corrections were made, and finally, after the corrections made aligning the internal standard. The data points corresponding to the solvent front peak were fitted with both a second order and a third order polynomial fit in DeltaGraph, and then the resultant equation was used in Maple to determine the exact time for the unretained peak. As there is very little difference between the value of the retention times found using either the second or third order polynomial, the second order polynomial was selected for ease of use. The raw data shows a distinctive, almost S-shaped, curve as shown in Figure 2.4.

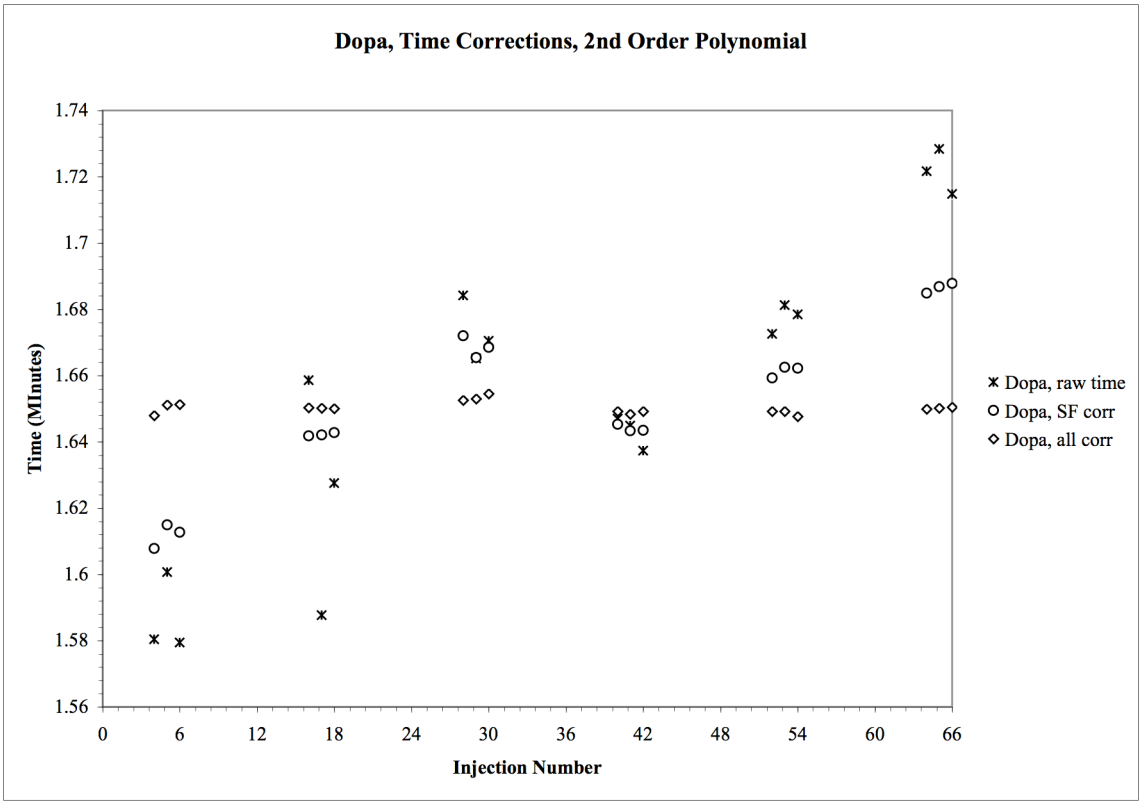


Figure 2-5: Retention time of Dopa as a function of retention time corrections

This is the data prior to any corrections. Each injection was done in triplicate, and the scatter in each triplicate set can be attributed to the variance introduced by the manual

injection. After the solvent front corrections, marked as SF corr in the figures, the S-shape begins to flatten out. Clearly, there is another component to this variance not addressed by the solvent front correction. After the correction for the internal standard, Dopamine, the S-shape to the curve flattens out and disappears. It is surmised that this variance is due to long-term pump oscillations.

Figure 2.5 shows the same analysis for Dopa. Again, the retention time for Dopa becomes much more reproducible after the solvent front corrections and the internal standard corrections are made. Figure 2-6 shows the analysis for the internal standard, Dopamine. The curvature in the peak is much pronounced here, simply because Dopamine is present in every injection. This phenomenon appears to begin at

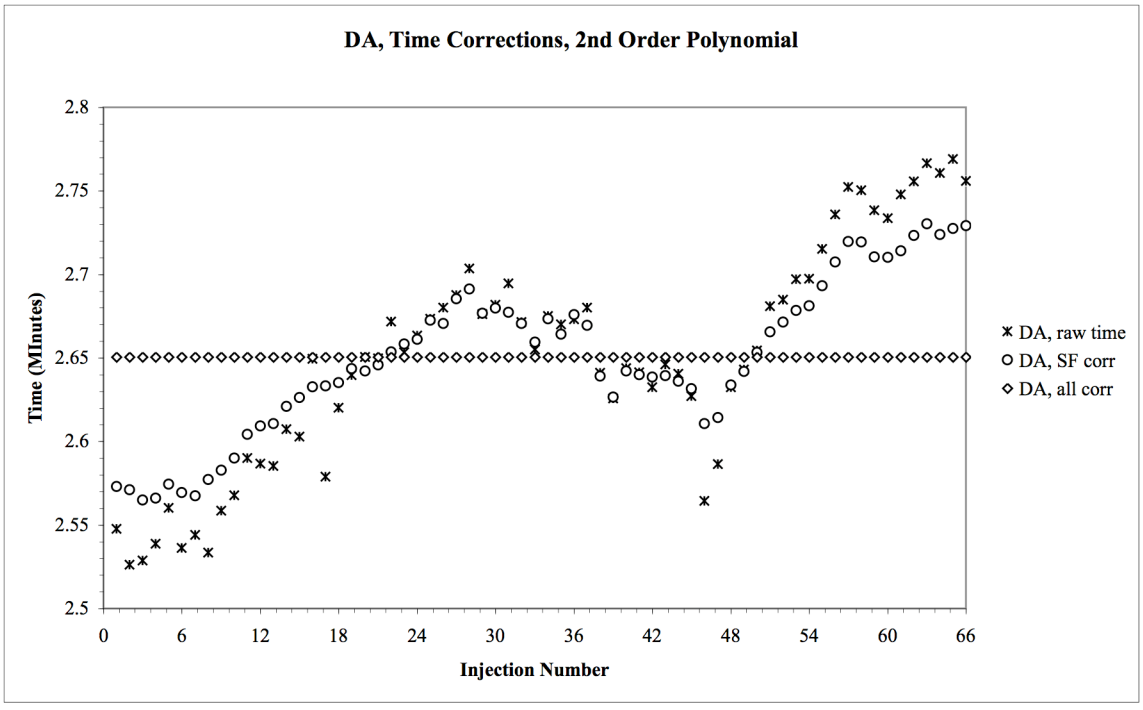


Figure 2-6: Retention time of Dopamine as a function of time corrections

approximately injection 29 and ends at injection 49. It is hypothesized that this behavior is due to pump oscillations. Since Dopamine is set to the average retention time found for the compound, the final data set will show Dopamine at a continuous time with no variance.

The linear pattern was expected for the raw data, either increasing or most likely decreasing. The internal standard was included to model for the expected decrease in the signal over the lengthy analysis time. As the time between injections was approximately 6.5 to 7 minutes, the total time for the experiment can be estimated to be between 7 hours and 9 minutes and 7 hours, 42 minutes. An internal standard was used in this work, and thus provides a method to adjust the data and thereby compensate for this oscillation.

Dopamine was used in this work as an internal standard as it is similar to the analytes and behaves similarly. The use of an internal standard allows for normalization of the peak should the amount injected be abnormally high or low. In this work, the Dopamine peak was fitted using the EMG function. The average time for the occurrence of the Dopamine peak in the solvent front corrected data was calculated. The data was then "zeroed" by subtracting the time the solvent front occurred in the solvent front corrected data from the data set, and the resultant time was used to create a ratio of the average Dopamine time to the actual Dopamine time. This ratio was then applied to each point in the data set, and the solvent front time added back in:

$$t_{corr,i} = \left(t_i - t_{\text{R, SF ave}}\right)\left[\frac{t_{\text{R, DA ave}} - t_{\text{R, SF ave}}}{\left(t_{\text{R, DA raw}} + t_{\text{SF, diff}}\right) - t_{\text{R, SF ave}}}\right] + t_{\text{R, SF ave}}$$

Equation 2.1

where $t_{corr,i}$ = the corrected time

t_i is the time corresponding to the data point

$t_{R,\ SF\ ave}$ = the average raw solvent front time for the injections made for the channel the data was collected on

$t_{R,\ DA\ ave}$ = the average raw retention time of Dopamine for the injections made for the channel the data was collected on

$t_{R,\ DA\ raw}$ = the raw retention time of Dopamine

$t_{SF,\ diff}$ = the difference between the average raw solvent front time for the injections made on the channel the data was collected on and the raw solvent front time

This procedure had the effect of keeping the solvent front peaks at zero and not shifting it due to the Dopamine time correction. The results of this work is shown in Table 2-4.

Table 2-4: Statistics for Channel 4 data

	Time of Raw Data (Minutes)	Time After SF Corr (Minutes)	Time After DA Corr (Minutes)
Solvent Front	0.7414	0.7414	0.7414
Std Dev	0.0223		
%RSD	3.0079		
Retention Time, Epi	1.6178	1.6192	1.6206
Std Dev	0.0448	0.0219	0.0015
%RSD	2.7676	1.3526	0.0898
Retention Time, Dopa	1.6542	1.6522	1.6500
Std Dev	0.0462	0.0246	0.0018
%RSD	2.7913	1.4908	0.1071
Retention Time, DA	2.6501	2.6501	2.6501
Std Dev	0.0666	0.0462	
%RSD	2.5115	1.7417	

The average retention times for the compounds, their standard deviations, and their relative standard deviations are shown above in Table 2-4 for Channel 4, for the raw

data, the data after the solvent front correction, and the data after the internal standard correction. As can be seen in this table, the percent RSD for Epi and Dopa both range about 2.8%, which amounts to an approximately 5.4 to 5.5 second window for Epi and Dopa, respectively, for these compounds prior to any correction to the data. After the solvent front correction, the window narrows to 2.6 seconds for Epi and 3.0 seconds for Dopa; finally, after all the corrections are made, the window narrows to 0.18 seconds for Epi and 0.22 seconds for Dopa. Given that the data is measured every 0.1 seconds, this precision was deemed acceptable. The retention times for the analytes Dopa and Epi prior to the time corrections were the same within a 95% confidence level, (Student's t-test); after the time corrections, the retention times for the two analytes are significantly different. The internal standard, Dopamine, as well as the mixtures were also treated in the same manner as for Epi and Dopa.

The data for the other channels was then treated in the same manner as for Channel 4, with the retention times for the non-retained species and the internal standard being set to the respective times found for Channel 4. Channel 4 data was the first data to be analyzed, and thus it was used as the template for time corrections in the remaining data sets.

Dopa and Epi in this study overlap one another to a very significant degree, with a resolution of less than 0.13. Figure 2-7 below shows the two analytes, Epi and Dopa, at peak heights that are roughly equivalent to one another at the potential being measured. As can be seen in this figure, the apparent peak shape is a fused peak, with no discernable shoulder that could be attributed to either of the two analytes. The slight bulge occurring

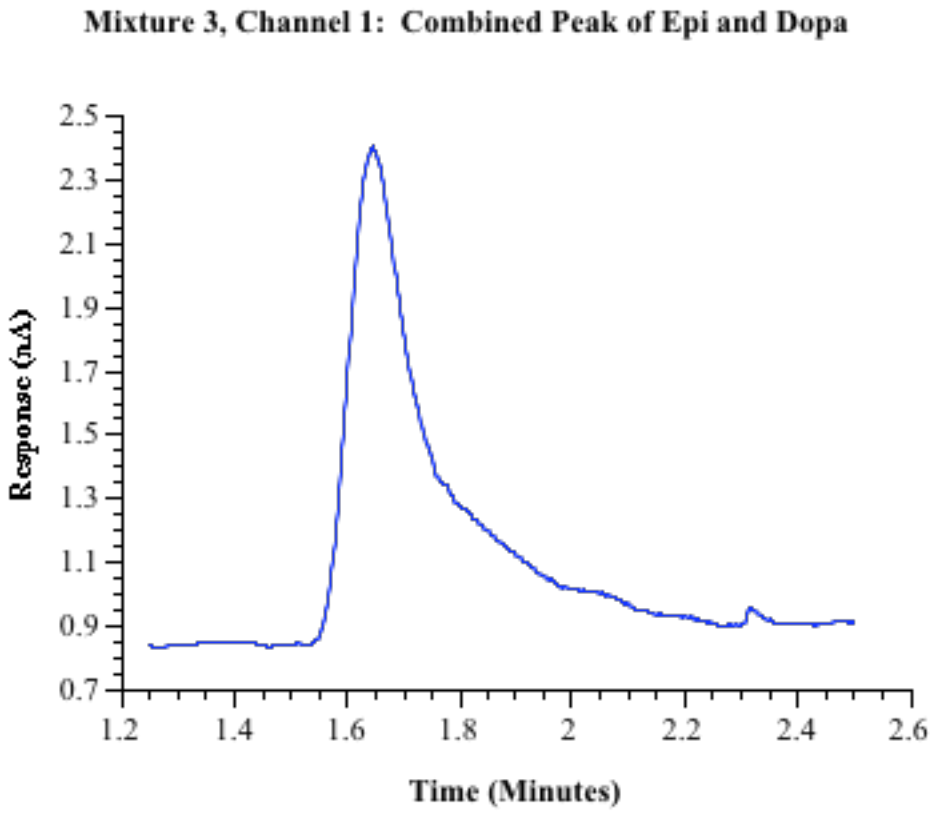


Figure 2-7: Overlap of Epi and Dopa

at approximately 2 minutes in this figure is not Epi nor is it Dopa, but appears to be a contaminant peak. The data presented here is prior to any corrections being made to the data. A total of 481 data point pairs (time and signal) in each data file were used in the subsequent deconvolution attempts. The time range of these data points encompassed both peaks, Epi and Dopa. The same number of data points were used for all data files, no matter the method attempted for deconvolution of the data. There were a total of 66 data files for each channel, one for each injection.

Curve Fitting (time-signal domain)

Determination of the chromatographic peak shape. The data for each standard injection was fitted to the twelve previously chosen chromatographic peak shapes using Excel (Microsoft, Redmond WA). Excel is chosen due both to its ease of use[140] and the ease of obtaining the software. The data for each standard was fitted to these equations using a residual-sum-of-squares (RSS) method in Excel, with Excel's Solver function. Solver uses the Levenberg-Marquardt method of iteration for convergence. The equations were restarted using differing parameters to minimize the possibility of ending at a local and not a global minimum.

Fitting the data to the model. The 4 or 5 equations exhibiting the lowest RSS values were then used in deconvoluting the mixture data. The model assumed that the mixture data could be represented by a function containing the addition of two such equations to total a "combination curve"; and the RSS minimization again using Excel enabled the best fit possible. The models were given the average values of the retention time and the peak width for the respective equations as found by fitting the standards. Bounding these parameters with upper and lower limits as dictated by the standard deviations of these parameters aided in the deconvolution of the mixture sets, and yielded results that were more in accord with the known values. Thus, each equation had two parameters with fixed ranges, and the remaining ones were allowed to change until the required level of convergence was achieved.

Application of the model and obtaining the analyte amounts. To obtain the peak heights, the respective parts of the combination curve were examined to determine the highest level the individual curves achieved. This was reported as the peak height. To

obtain peak areas, the summation of the signal for each time point was obtained and reported as the peak area for each individual curve. The total of the individual peak areas thus, under this model, is the area of the experimental curve.

To get the respective amounts of the analytes, the combination curve was applied to the mixture data. The curve fitting was repeated several times using slightly different parameters to ensure that a global minimum was achieved in the RSS determination.

Using Chemometric Methods.

The chemometric methods used in this work begin with PCA to identify the number of principal components in the data The regression methods PCR and PLS, using both NIPALS and SIMPLS, were used to determine the relative amounts of the analytes in the mixtures. PCR and PLS were then verified, and the results compared with each other and also with the other methods presented in this work. Since these methods require the data to be presented in a matrix form, Matlab (The MathWorks Inc., Natick NY) was used as the base application along with PLS Toolbox 2.1 (Eigenvector Research, Inc., Manson, WA). PLS Toolbox 2.1 is a module used with MatLab and contains the routines necessary for the chemometric regression methods used herein.

Cramer's Rule approach in the signal-potential domain

Fitting the Data to Electrochemical "Peak" Shape. The equation used was previously derived, giving current as a function of the potential the electrode:

$$i = \frac{-K_R C_R^*}{1 + 10^{-\left[\frac{(E - E_c)s}{0.05916}\right]}}$$

Equation 2.2

where the parameters shown above are now those found experimentally during the course of the analysis. E_c is the experimentally determined half-wave potential, s is overall experimentally determined number of electrons involved in the reaction, i is the current, E is the potential of the electrode, and C_R is the concentration of the reduced species. All are constant but for the current and the potential.

Determining the amounts of the analytes. The data for each standard was fitted to this equation using a residual-sum-of-squares method in Excel, using Excel's Solver function. Solver uses the Levenberg-Marquardt method of iteration for convergence. These parameters are then used to determine the constants to be used in the final simultaneous linear equation solutions, derived in Chapter 1 and reproduced here:

$$\begin{aligned}
i_{tot,Ch1} &= \left(K_O' C_O^*\right)_{Ch1,Cmpd1} + \left(K_O' C_O^*\right)_{Ch1,Cmpd2} + \cdots + \left(K_O' C_O^*\right)_{Ch1,Cmpd(n)} \\
i_{tot,Ch2} &= \left(K_O' C_O^*\right)_{Ch2,Cmpd1} + \left(K_O' C_O^*\right)_{Ch2,Cmpd2} + \cdots + \left(K_O' C_O^*\right)_{Ch2,Cmpd(n)} \\
i_{tot,Ch3} &= \left(K_O' C_O^*\right)_{Ch3,Cmpd1} + \left(K_O' C_O^*\right)_{Ch3,Cmpd2} + \cdots + \left(K_O' C_O^*\right)_{Ch3,Cmpd(n)} \\
i_{tot,Ch4} &= \left(K_O' C_O^*\right)_{Ch4,Cmpd1} + \left(K_O' C_O^*\right)_{Ch4,Cmpd2} + \cdots + \left(K_O' C_O^*\right)_{Ch4,Cmpd(n)}
\end{aligned}$$

Equation 1.87

for the reduced species in bulk. In this case, there are only two analytes and thus only two known sources of current. However, all four electrodes were in use. Once again,

Solver was employed to determine the best values for Epi and Dopa, using a sum-of-squares of the residual approach.

Chapter 3 — Curve Fitting Results and Discussion

Curve fitting is one approach to quantifying severely overlapped peaks. By modeling the observed chromatographic peak to an equation, the individual peaks that form the observed peak may be mathematically detected, thus allowing each contributor peak to be integrated. We evaluated twelve common curve-fitting equations (bi-Gaussian, exponentially modified Gaussian (EMG), Fraser-Suzuki, log-normal, Haarhoff-van der Linde, Cauchy-Gaussian, Chesler-Cram, Giddings, GMG, GEMG4, GEMG5, EMG+GMG) and the best performers (EMG, GEMG4, GEMG5, EMG+GMG and Cauchy-Gaussian) were used to deconvolute the observed chromatographic peak into two highly overlapped peaks by curve-fitting using the time vs. signal domain. Both area under the curve and peak heights were examined. Five mixture sets were made to encompass a range of peak height ratios, as described in Chapter 2.

An evaluation of the results must evaluate accuracy, precision, and both accuracy and precision. The results of the model may be accurate, but highly imprecise; the uncertainty in such a measurement is huge and thus the results the method yielded cannot be used. The answer may be very precise, but far from accurate, and thus the answer the method yielded may again be useless. Ideally, the model should produce results that are both accurate and precise. The results of the analyses were tested using Student's t-test to determine if the results agree with the known values of the standards to 95% certainty.

To evaluate the accuracy of the methods, the data is presented as a simple bar graph of the difference percentages, that is, the difference between the calculated and the

known value calculated as a percentage of the known value. The sign of the difference is an indication of whether the calculated value is higher or lower than the known value. The data is also presented as a simple bar graph of the %RSD values for both Epi and Dopa for each method. The methods yielding the lowest %RSD values for both analytes are more precise than the rest. An analysis of variance was also performed, thus giving an indication of the precision of the curve fitting methods as a whole. A final question arises as to whether or not some methods used in this work yield the same mean values; if this is the case, perhaps more than one method can be said to yield both precise and accurate answers, and thus either method could be used.

The results for each channel are presented below. For each channel, the equations which yielded the lowest residual sum-of-squares (RSS) was selected as possible deconvolution models. The value of the RSS could not be fixed, since each channel yielded varying "low RSS" values. Thus, the number of equations was arbitrarily chosen, based on an evaluation of the data. The lowest RSS values tended to cluster together, and the equations corresponding to these values were used as possible models. The results of the Student t-tests are presented first, followed by the accuracy and precision analysis. A discussion for the results follows.

Channel 1

The data from Channel 1 was considered the most likely channel for deconvolution of the mixture sets, as this channel is free of any interferences from upstream electrodes. Additionally, Mix 3 was designed to give approximately equal peak heights for the two analytes at the potential this channel was set, and thus was believed be

the "easiest" of the mixture sets and channels to deconvolute. The potential for this electrode was set for 600 mV, yielding data less susceptible to interferences from other higher oxidizing species.

Results

As shown in Figure 3-1, only 4 equations out of the 12 investigated met the lowest RSS value criteria as discussed above for both Epi and Dopa: the Cauchy-

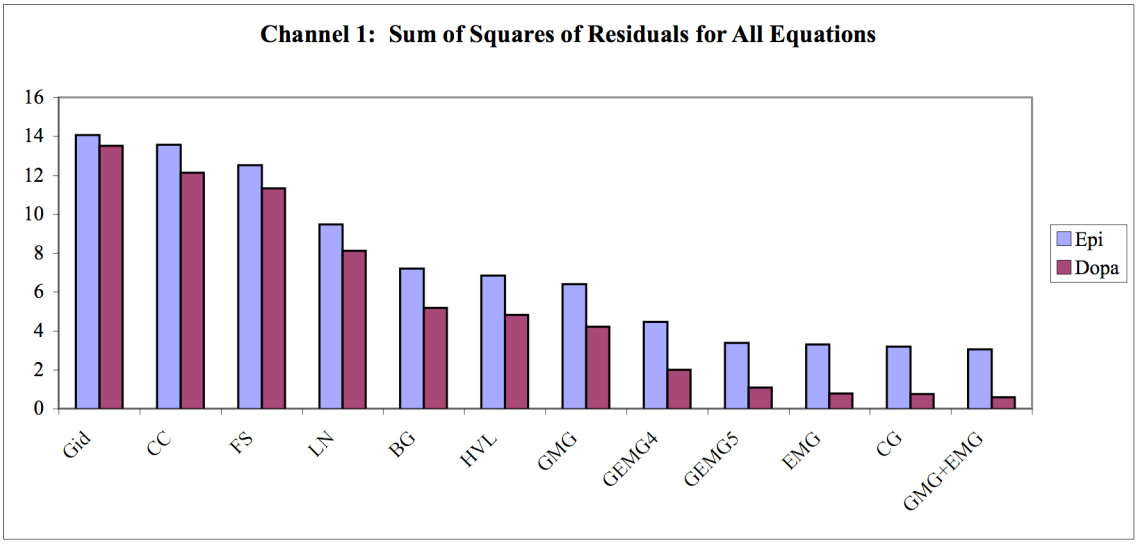


Figure 3-1: SSR results for Channel 1

Key: Gid=Giddings; CC=Chesler-Cram; FS=Frazier-Suzuki; LN=Log Normal; BG=Bi-Gaussian, HVL=Haarhoff-Van der Linde; GMG=Half Gaussian-Modified Gaussian; GEMG4=4 parameter Half Gaussian-Modified Gaussian; GEMG5=5 parameter Half Gaussian-Modified Gaussian; EMG=Exponentially Modified Gaussian; CG=Cauchy-Gaussian; GMG+EMG=Half Gaussian Modified Gaussian + Exponentially Modified Gaussian

Gaussian equation, the EMG equation, the EMG+GMG equation, and the GEMG5 equation.

Mix 1

None of these approaches produced the appropriate results for either analyte using the Student's t-test, 95% certainty using the time vs. signal domain. The results of the

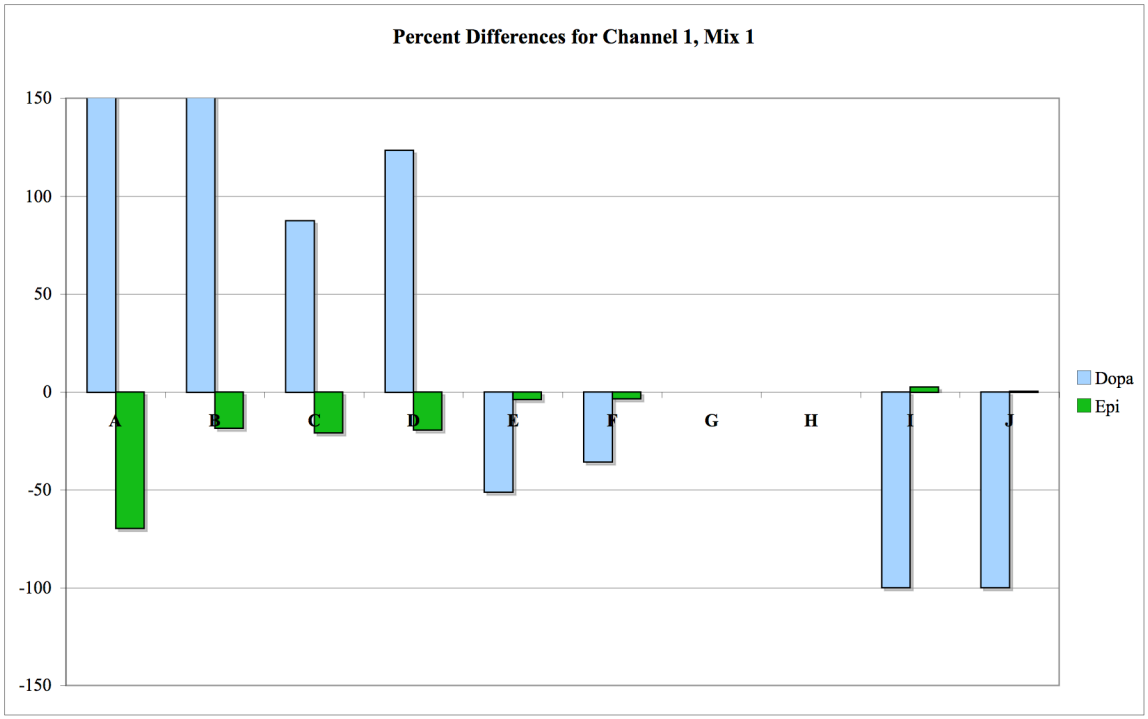


Figure 3-2: Percent Difference for Channel 1, Mix 1

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

percent differences in the returned values and the known values for Mix 1, Channel 1 are shown above in Figure 3-2. The method that consistently yields the lowest values would be the method deemed most accurate. The GEMG4 (both G and H) equation was not used in the formation of Channel 1 information, as the sum-of-squares found for the equation was outside the range considered acceptable for the residual sum of squares (RSS). Therefore, the table does not include these two methods, although the methods were performed for the remaining channels. According to the analysis, the EMG+GMG peak height method yielded the best results of all those presented, followed by the EMG+GMG Area. The percent differences found for Dopa for these methods was quite large, however, at over 30%.

The percent relative standard deviations for the results for all the methods for Channel 1, Mix 1, are shown below in Figure 3-3. Once again, the GEMG4 equation (shown as G and H) was not used. The methods producing the smallest %RSD for both Epi and Dopa are: Cauchy-Gaussian Area (A), Cauchy-Gaussian Peak Height (B), EMG Area (C) and EMG Peak Height (D). The lowest %RSD for both analytes was found using the Cauchy-Gaussian Peak Height (B). However, none of the methods produced acceptable results at the 95% confidence level. While these methods produce good precision, they do not produce good accuracy.

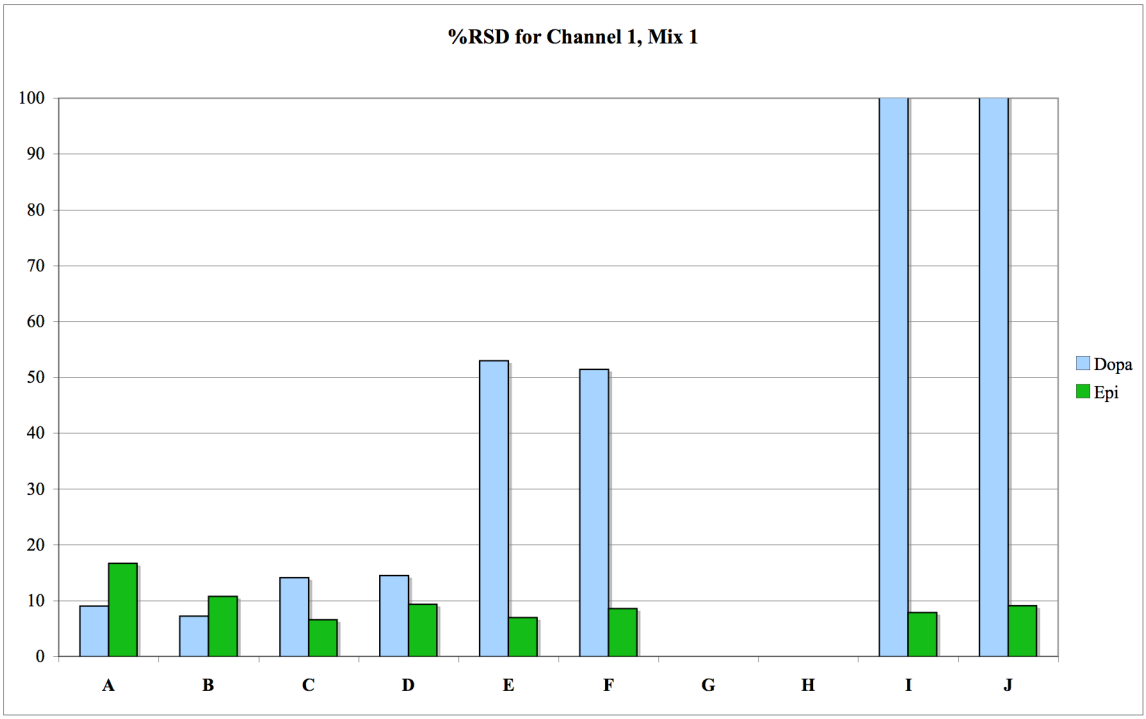


Figure 3-3: %RSD for Channel 1, Mix 1

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 2

For both Epi and Dopa, the following methods yielded accurate results with the required certainty of 95% using Student's t-test: the GEMG5 Area, the EMG Peak Height, and the GEMG5 Peak Height. The next question is which of these methods are the most accurate and/or precise.

The percent difference results for the analytes for Channel 1, Mix 2 are shown below in Figure 3-4. Once again, the GEMG4 equation was not used in the formation of Channel 1 information, as the sum-of-squares found for the equation was outside the

range considered acceptable, and thus should be ignored (G and H in Figure 3-4 above). t-Tests indicate that for both Epi and Dopa, the GEMG5 Area, the EMG Peak Height, and the GEMG5 Peak Height methods are acceptable to 95% certainty. Of these the EMG Peak Height (D) and GEMG5 Peak Height (J) methods are more accurate than the GEMG5 Area (I). In fact, the EMG Area (C) is closer to the known values than is the GEMG5 Area.

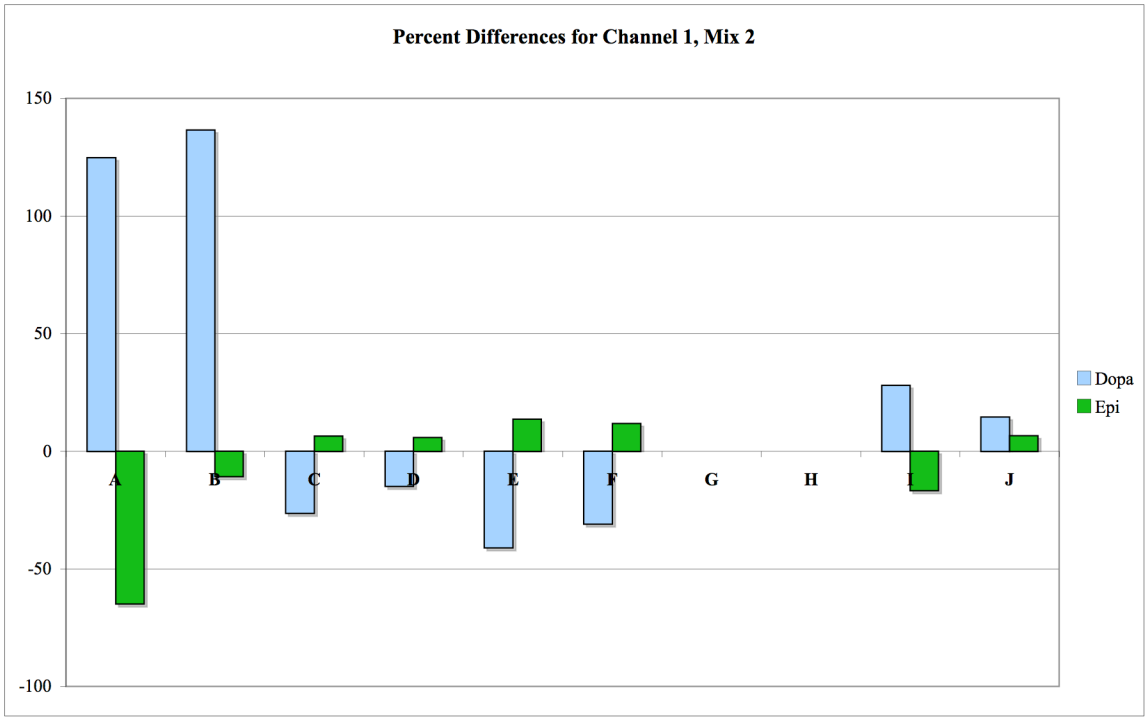


Figure 3-4: Percent Difference for Channel 1, Mix 2

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

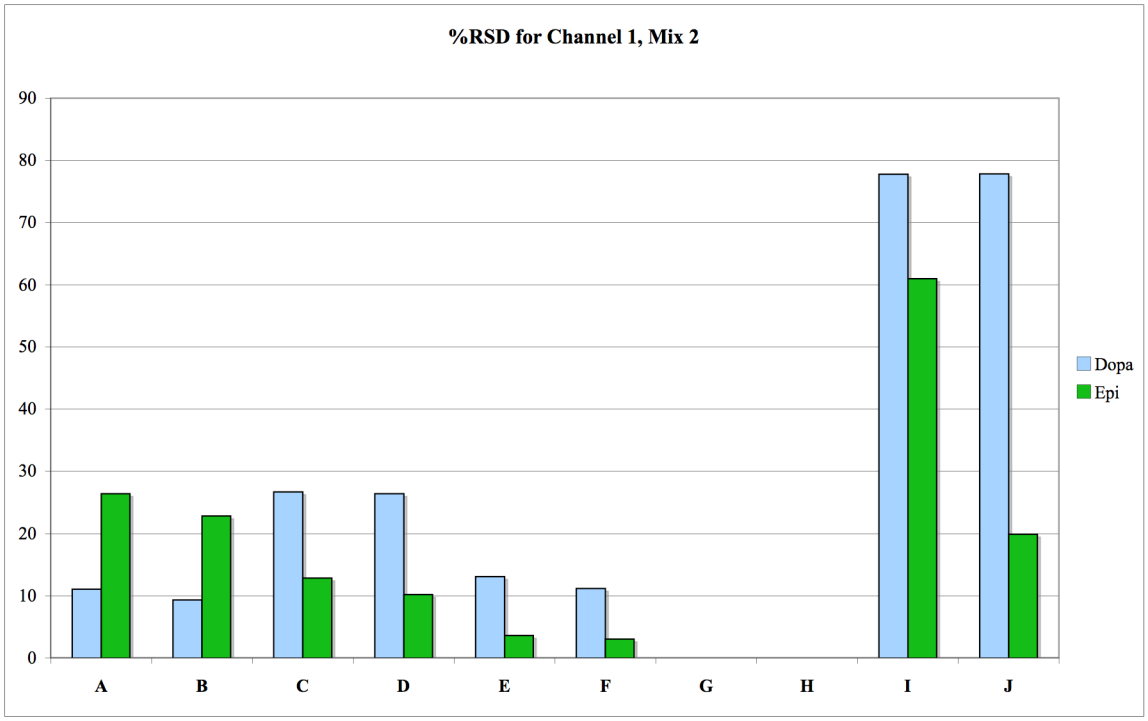


Figure 3-5: %RSD for Channel 1, Mix 2

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

The results of the %RSD study are shown above in Figure 3-5. The most precise methods shown here are the EMG+GMG Area method, (E), and the EMG+GMG Peak Height method, (F); neither of these two methods are accurate or precise as described by the Student's t-test results, which indicate the GEMG5 Area, the EMG Peak Height, and the GEMG5 Peak Height methods are acceptable to 95% certainty. The GEMG5 methods are the least precise of the methods examined for this mixture set with %RSD values for Dopa at about 78% RSD and high values for Epi. The EMG Peak Height method was much lower, with an approximately 28% RSD value for Dopa, and nearly

10% RSD for Epi. Thus, for the methods found acceptable by Student's t-test, the EMG Peak Height is the most accurate and precise.

Mix 3

The results of the Student's t-test for Mix 3 Channel 1 for both Epi and Dopa revealed that only the EMG Peak Height method yields the correct answer to 95% confidence. The percent difference analyses for Dopa and Epi are shown below in Figure 3-6 below, which indicates the EMG Peak Height method (D) has one of the lowest

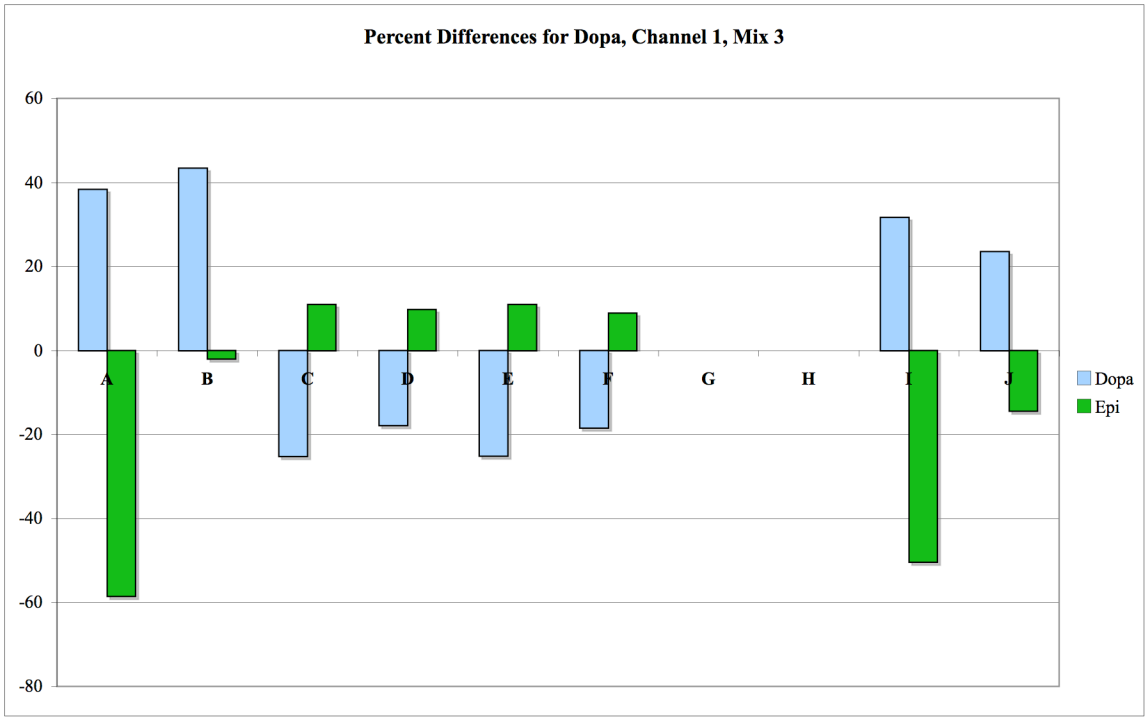


Figure 3-6: Percent Difference for Channel 1, Mix 3

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

percent differences for both analytes of the methods investigated. Once again, the GEMG4 equation was not used for Channel 1 because of the higher sum-of-squares found for this equation.

The %RSD results for this mixture set are shown below in Figure 3-7. The lowest %RSD values were found for the methods other than the EMG Peak Height method (D): the Cauchy-Gaussian Peak Height (B) and the Cauchy-Gaussian Area (A) are both considerably lower than the EMG Peak Height, which returned values for Dopa and Epi at about 21% and 15%, respectively.

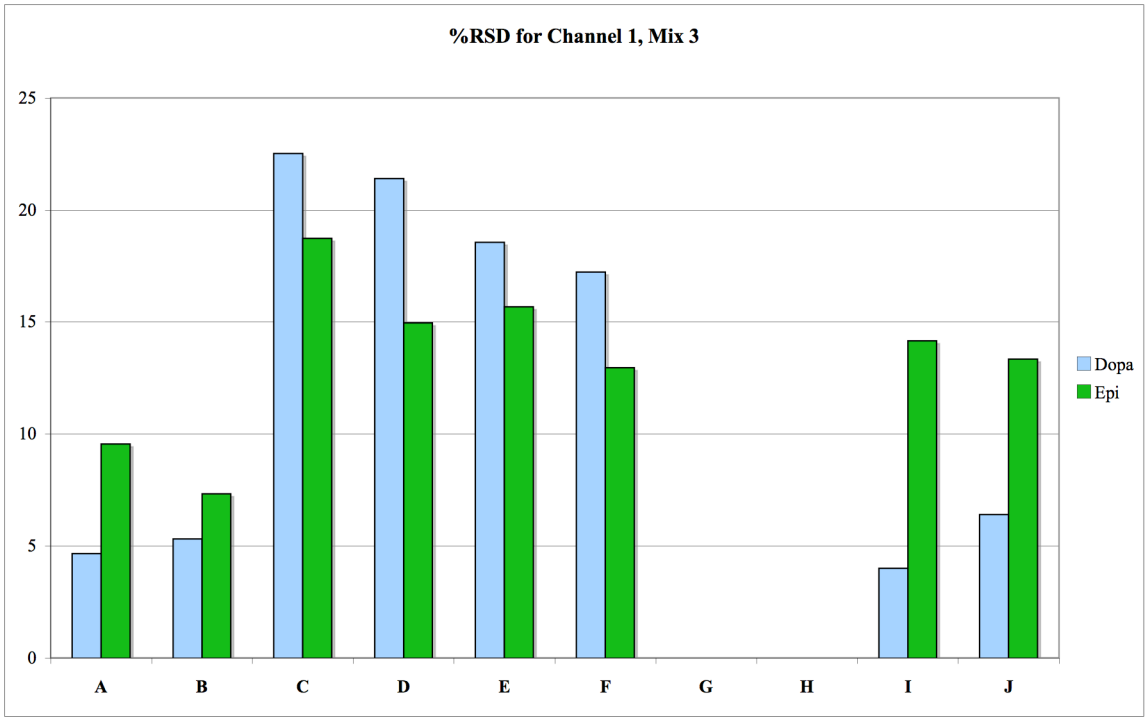


Figure 3-7: %RSD for Channel 1, Mix 3

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 4

With both analytes, the Cauchy-Gaussian Area and Cauchy-Gaussian Peak Height both yields the correct answer to with a confidence of 95%. As seen in Figure 3-8, the Cauchy-Gaussian Peak Height method, (B), gives results closer to the known values than does the Cauchy-Gaussian Area method, (A). Once again, the GEMG4 methods were not used in this analysis.

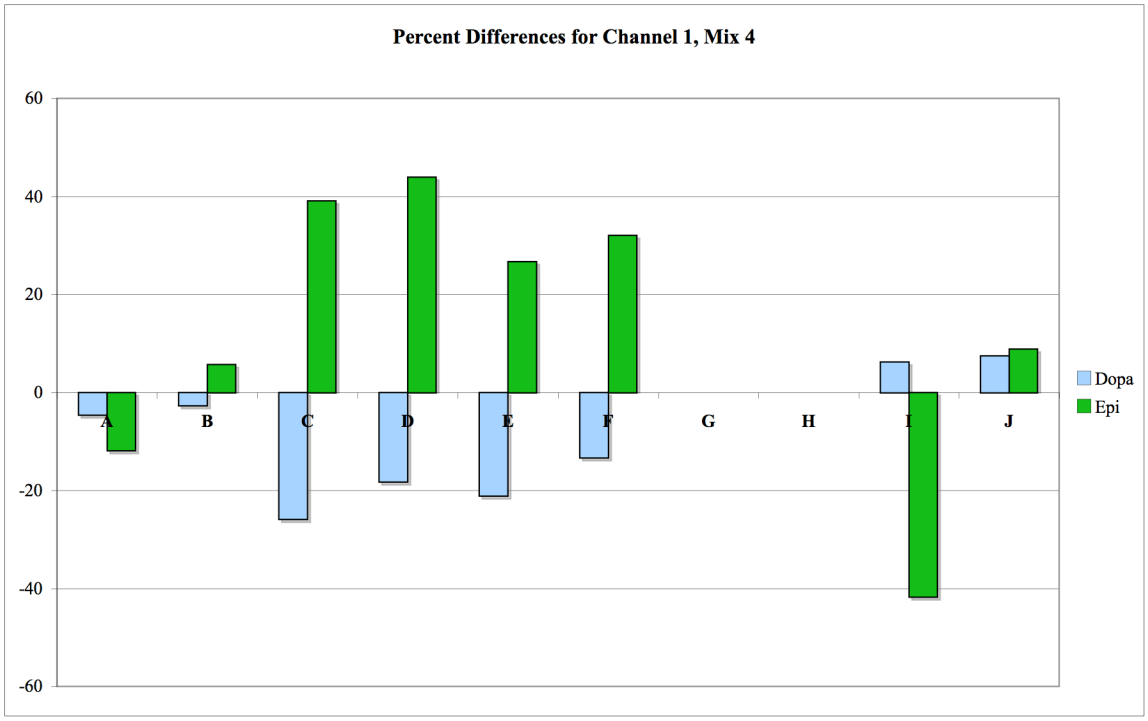


Figure 3-8: Percent Difference for Channel 1, Mix 4

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

The %RSD results are shown below in Figure 3-9. As the GEMG4 equation was not used for Channel 1, these results indicate that the GEMG5 Area (I) and the GEMG5 Peak Height (J) methods yield the best %RSD values. The two methods found to be acceptable, according to the Student's t-test, are the two methods that yield the worst %RSD values. Thus, the use of either of these two methods to determine the values of Epi and Dopa in Mix 4 would yield values that, while accurate, are not precise. Thus, for Mix 4, no method can be recommended for deconvolution of the data.

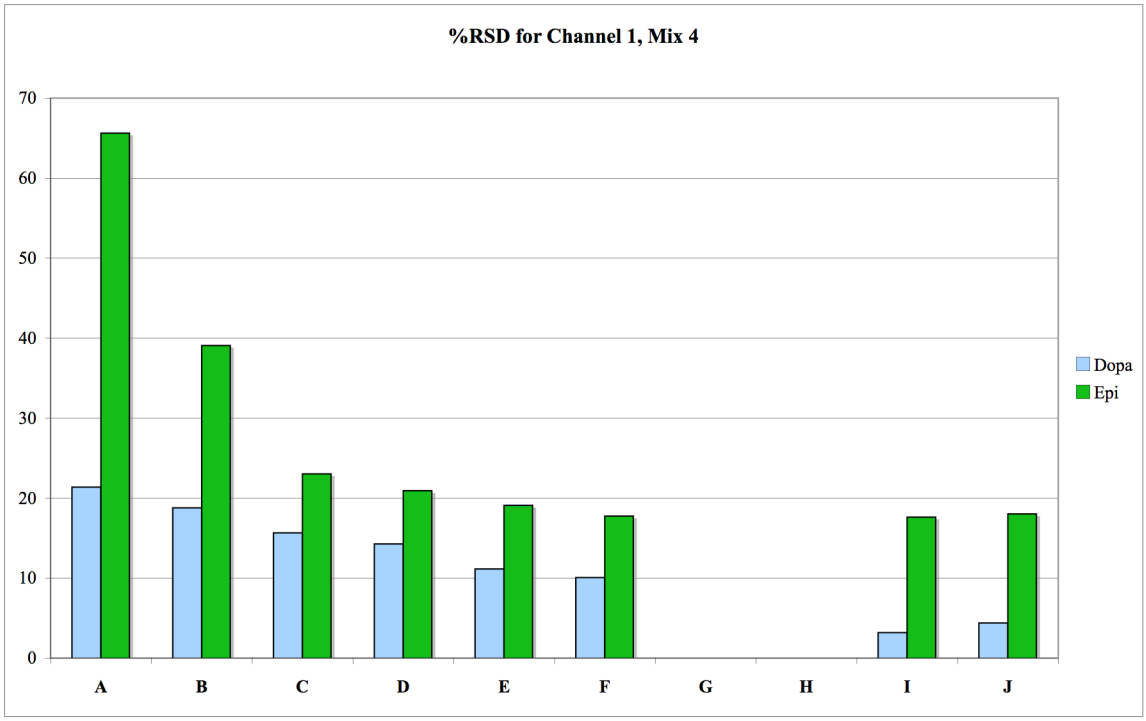


Figure 3-9: %RSD for Channel 1, Mix 4

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 5

For this mixture set, none of the method used give acceptable results to 95% confidence according to the Student's t-test. Figure 3-10 shows that the only method that can be considered accurate is the GEMG5 Area method.

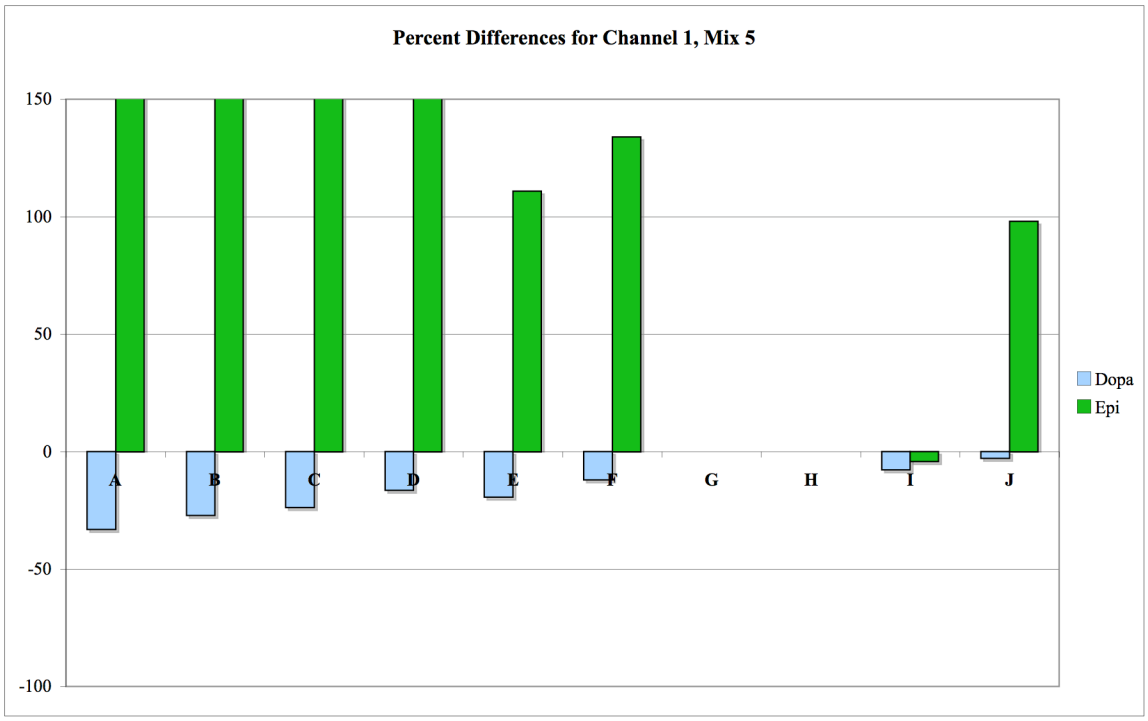


Figure 3-10: Percent Difference for Channel 1, Mix 5

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Figure 3-11 below shows the %RSD values for Dopa and Epi. These values are not high for Dopa for any of the methods used; however, the values found for Epi range from a reasonable 9% RSD to an unreasonable 21% RSD. The one method that appeared

to be accurate from the percent difference results was shown to be imprecise for the determination of Epi, at some 18% RSD.

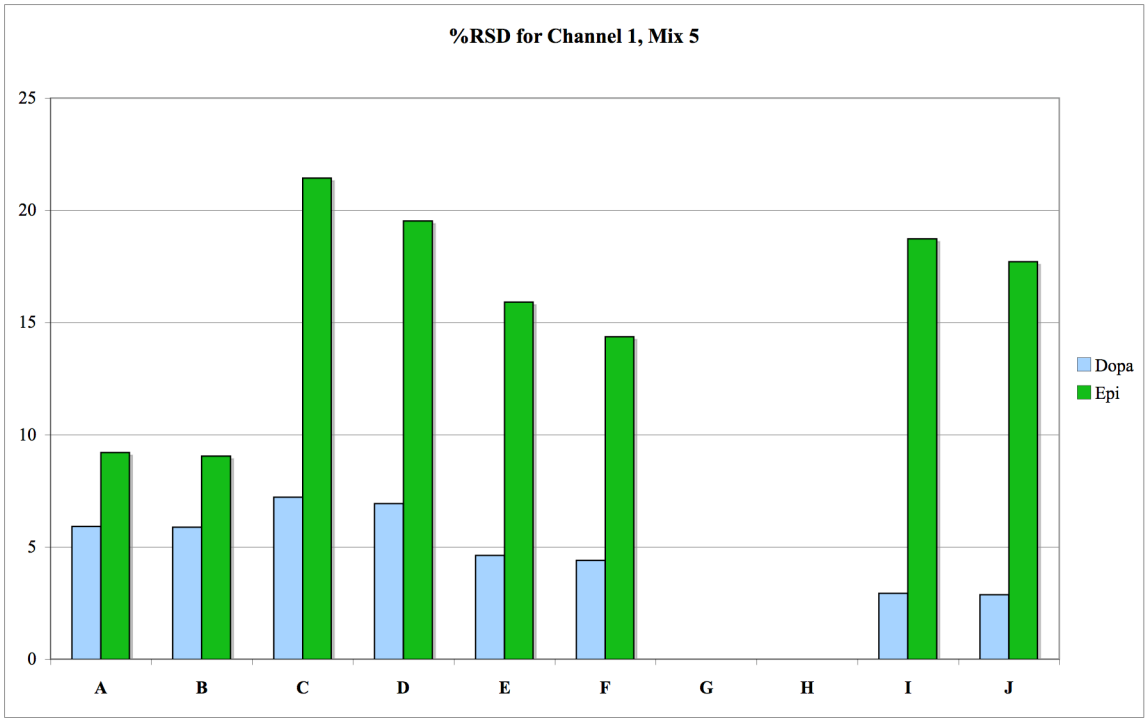


Figure 3-11: %RSD for Dopa, Channel 1, Mix 5

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Discussion

No single method yielded the proper values at the 95% confidence levels for all five mixture sets. Mix 1 contains the largest amount of Epi present at approximately 90% of the peak height, and the smallest amount of Dopa at approximately 10% of the peak height. For Mix 5, these numbers are reversed. The mixtures only examined ratios of

10:90, 30:70, 50:50, 70:30 and 90:10 Dopa:Epi of the peak heights for two analytes present in a single merged peak. For the Channel 1 data, no method presented herein using the time-response domain was able to deconvolute the merged peak throughout the range of amounts used in this study. This is significant, as this domain is the one most commonly and traditionally utilized.

Mixes 1 and 5 could not be accurately deconvoluted by any method investigated herein. The remaining mixtures (Mix 2, 3, and 4) were unable to be deconvoluted as a group by any of the methods. The argument that the mixtures with the lowest proportion of the "contaminating" peak could not be deconvoluted because the amounts were too low cannot be used here, as the remaining mixture sets would have been deconvoluted with a consistent method. Mix 2 was able to be deconvoluted by the GEMG5 Area, the EMG Peak Height, and the GEMG5 Peak Height. Of these three, only the EMG Peak Height method could be accurate, but with unacceptable precision (28% RSD) for Dopa. For Mix 3, only the EMG Peak Height method was able to deconvolute the mixture set, albeit with low precision (21% for Dopa and 15% for Epi). Mix 3 was expected to be the easiest to deconvolute, as the individual component peak are approximately equal in height. For Mix 3, the Cauchy-Gaussian methods were within 95% confidence per the Student's t-test. Thus, none of the methods presented for Channel 1 could be used for the range of molar responses seen in this work.

Additionally, the data obtained from the first channel is free from interferences arising from the other channels. If the sample is to be analyzed via LCEC, the use of a single detector in the electrochemical cell is a reasonable assumption. As this work has

shown, the use of a single detector set at a low potential of 600mV did not yield data that could be deconvoluted using curve fitting methods.

Channel 2

Channel 2 was set to oxidize the analytes at a voltage of 700mV. As it physically located behind the electrode used for Channel 1, some differences in the peak shape were expected to occur. Some of the analyte had already been oxidized; however, this amount is usually considered minimal given this method of detection. If any cross-talk were to take place between electrodes 1 and 2, it would be expected to be seen here at the second electrode due to the flow of the mobile phase across the electrode surfaces. The longer path that the analyte plug had to travel within the electrode cell would be expected to contribute to some slight broadening of the peak shape. Additionally, the larger voltage applied to the electrode ensured that the peak height ratios would change as well, although the amount presented to the electrode would be considered to be the same.

Results

The following five out of the original 12 equations yielded the lowest RSS values: the GMG+EMG, the EMG, the GEMG5, the Cauchy-Gaussian, and the GEMG4 equations.

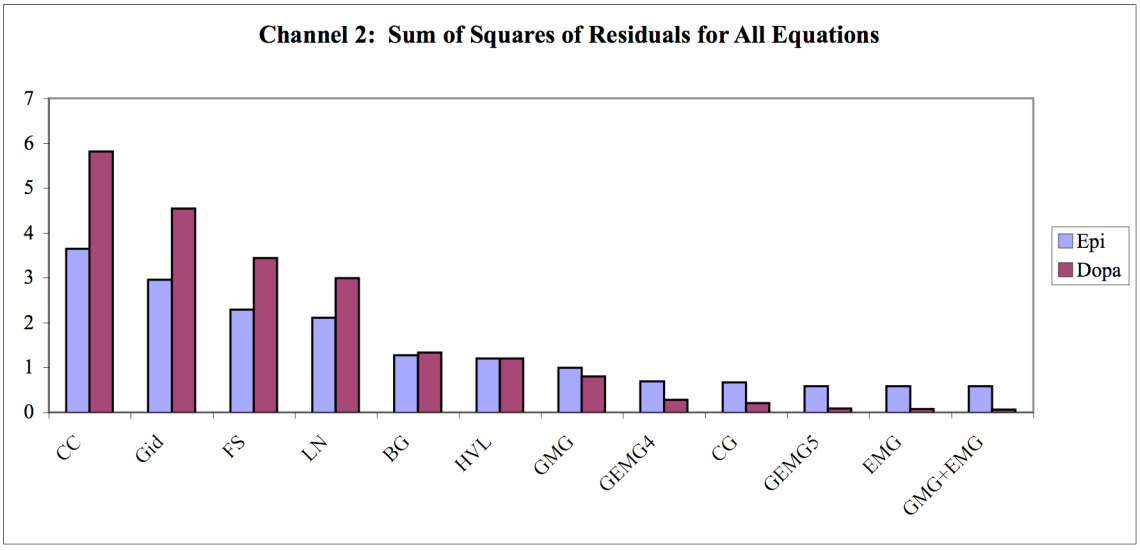


Figure 3-12: SSR results for Channel 2

Key: Gid=Giddings; CC=Chesler-Cram; FS=Frazier-Suzuki; LN=Log Normal; BG=Bi-Gaussian, HVL=Haarhoff-Van der Linde; GMG=Half Gaussian-Modified Gaussian; GEMG4=4 parameter Half Gaussian-Modified Gaussian; GEMG5=5 parameter Half Gaussian-Modified Gaussian; EMG=Exponentially Modified Gaussian; CG=Cauchy-Gaussian; GMG+EMG=Half Gaussian Modified Gaussian + Exponentially Modified Gaussian

Mix 1

An evaluation of the results for Mix 1 found none of the methodologies studied herein are able to achieve the goals set forth, using the Student's t-test at 95% confidence.

The percent differences for Dopa and Epi are shown in Figure 3-13. The smallest percent differences for for both Epi and Dopa appear to be the GEMG5 Area method (I) followed by the GEMG4 Area method (G). While there appear to be small values due to the scale of the y-axis, the percent differences for the GEMG5 Area method are above

10% to 15%. Therefore, none of the methods examined here can be considered acceptable.

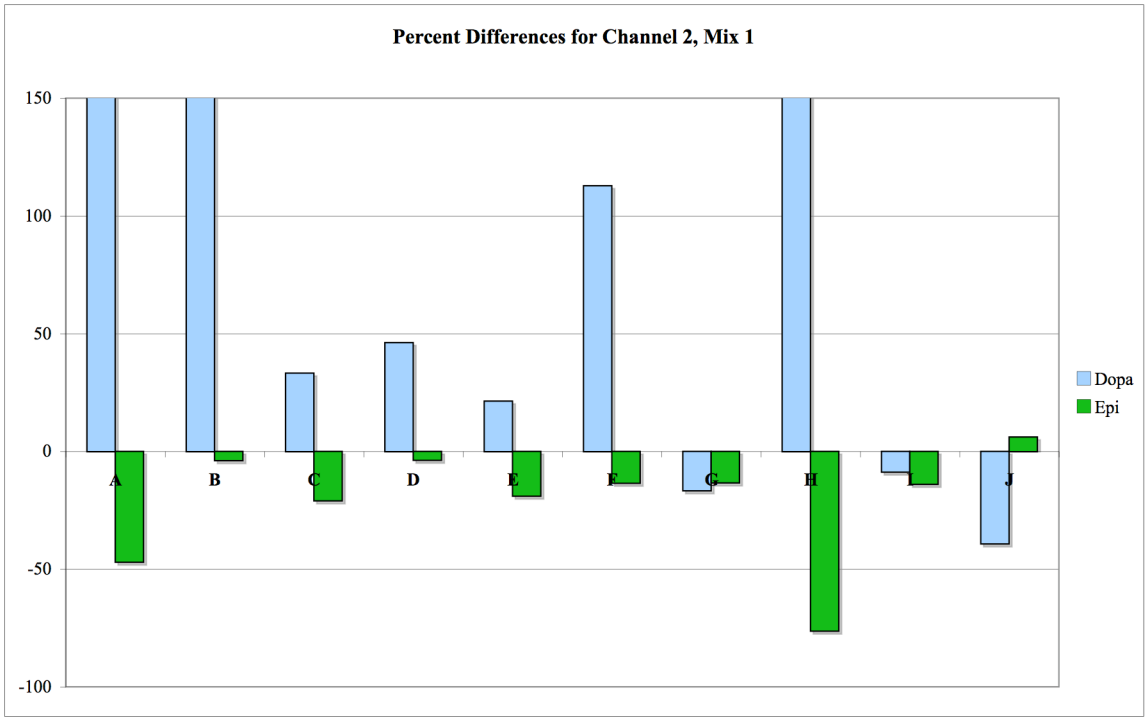


Figure 3-13: Percent Difference for Channel 2, Mix 1

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Figure 3-14 examined the question of precision in this mixture set. The GEMG4 Peak Height method (H), the GEMG4 Area method (G), and the Cauchy-Gaussian Peak Height method (B) all are precise, returning values less than approximately 5% for both Dopa and Epi.

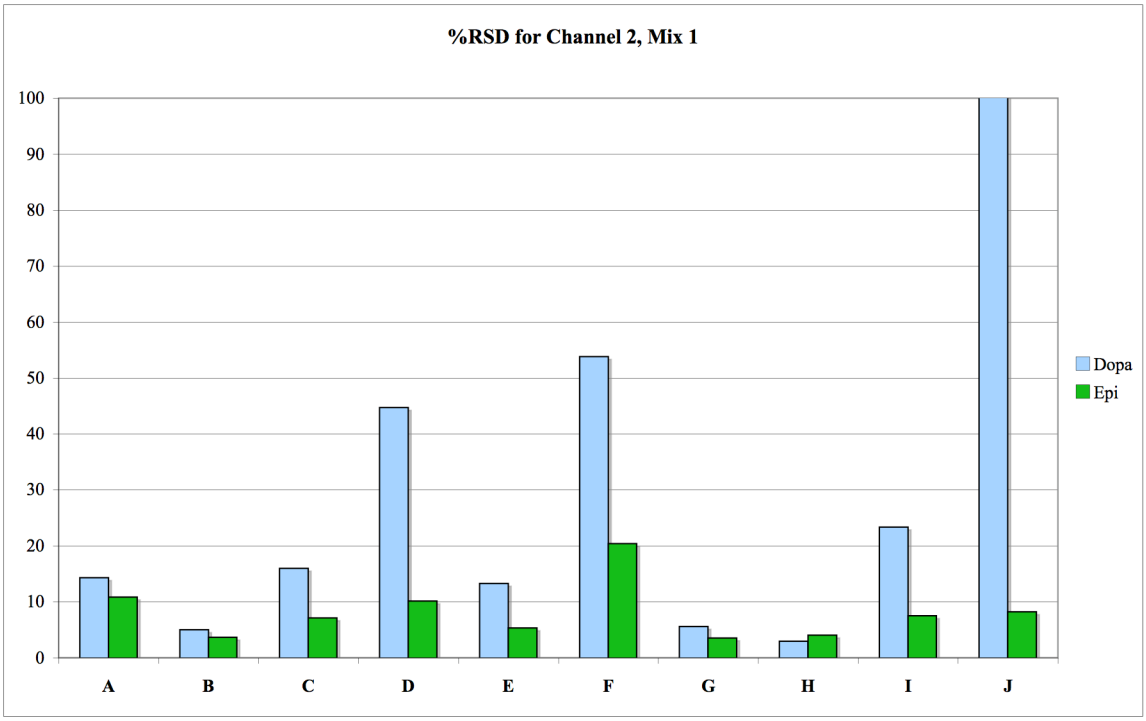


Figure 3-14: %RSD for Channel 2, Mix 1

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 2

When both analytes are considered for this mixture set, none of the curve fitting methodologies yielded the appropriate values within 95% confidence using the Student's t-test. The accuracy and precision of these methods is discussed below.

The percent difference results for this mixture set are shown below in Figure 3-15. The most accurate method for both Dopa and Epi is the EMG Area (C) method, where Epi is under-reported by nearly 10%, followed by the EMG+GMG Area (E) method, where Dopa is under-reported, again by nearly 10%.

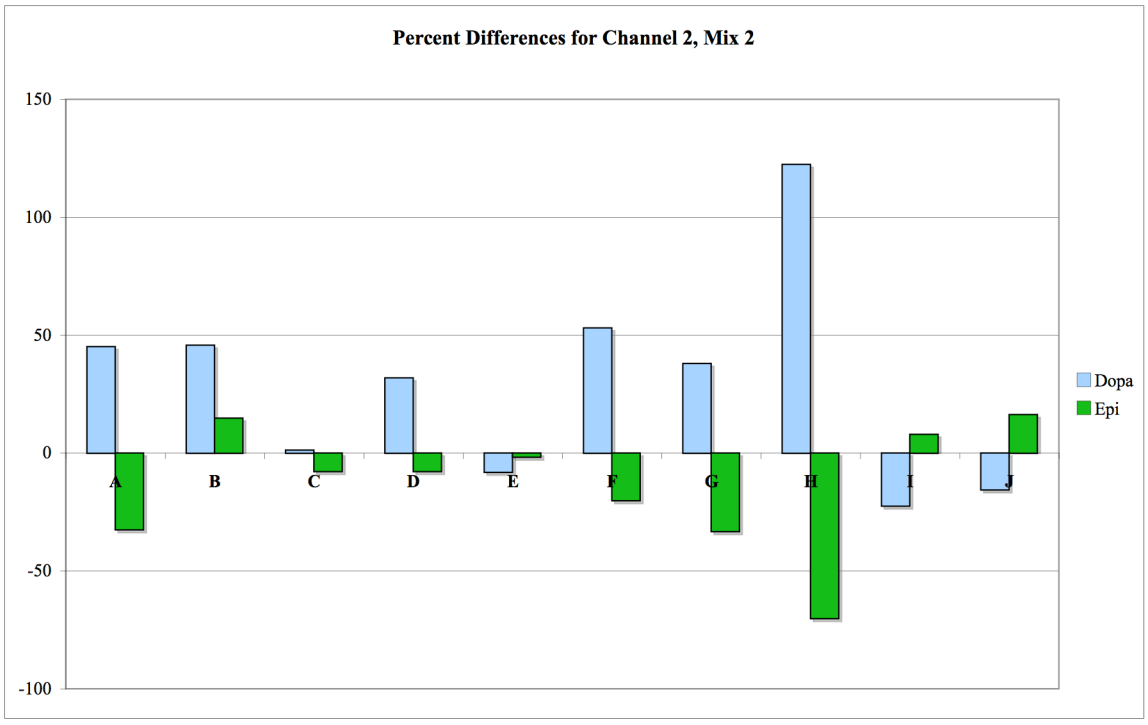


Figure 3-15: Percent Difference Channel 2, Mix 2

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Examination of Figure 3-16 shows the precision of the methods for this mixture set. The most precise methods for both Dopa and Epi appear to be the GEMG4 Peak Area (K), the Cauchy-Gaussian Area (A), the EMG+GMG Area (E), the GEMG5 Area (I) method. These methods consistently produced %RSD values near 5%.

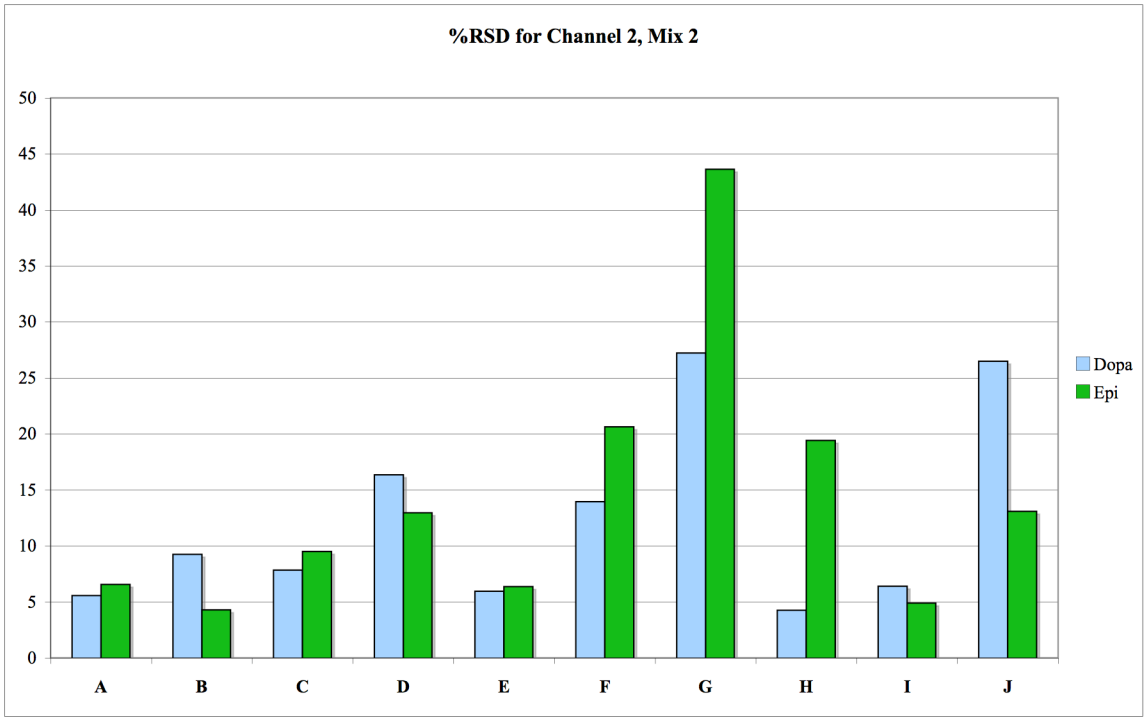


Figure 3-16: %RSD for Channel 2, Mix 2

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 3

When the two analytes are considered together, only the EMG Peak Height and EMG+GMG Peak Height yields the required values at 95% confidence for the levels of analytes present in this mixture set.

The accuracy of the methods for this mixture set can be determined from Figure 3-17. The methods yielding the lowest percent differences for both Dopa and Epi are the EMG Peak Height method (D), the EMG+GMG Area method (E), and the EMG+GMG Peak Height method (F). These all yielded less than a 10% difference from the known

values and the calculated values, with the EMG+GMG Area under-reporting both analytes.

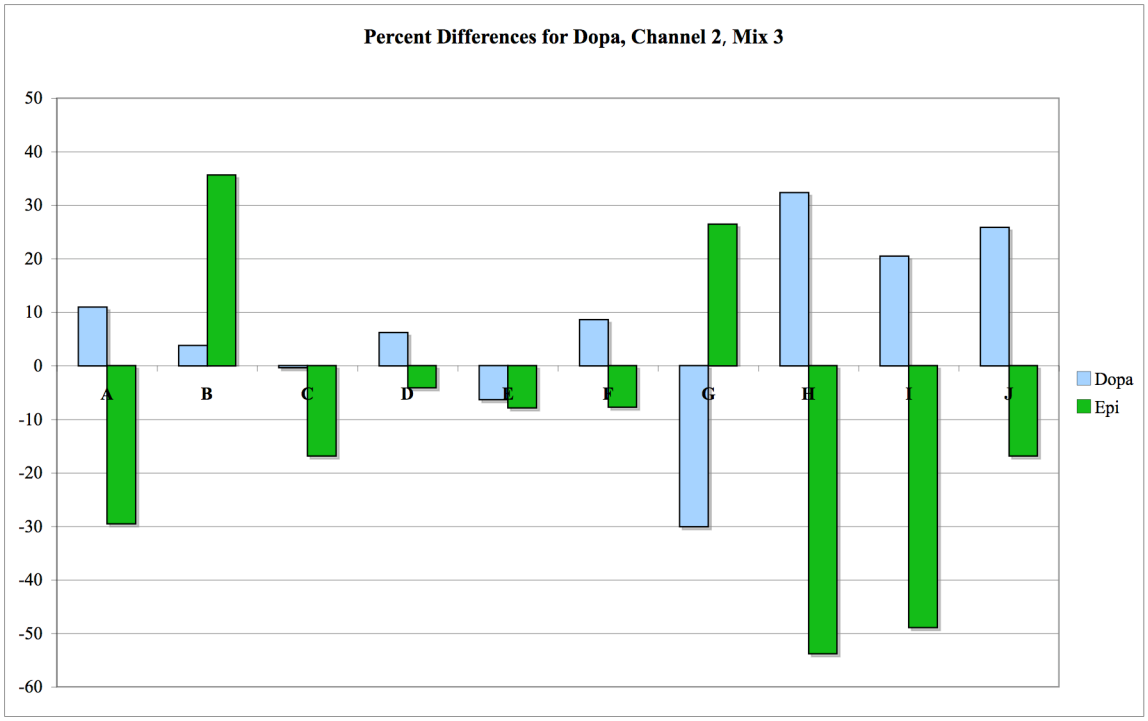


Figure 3-17: Percent Difference for Dopa, Channel 2, Mix 3

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

In Figure 3-18, the methods that yield the best precision are the Cauchy-Gaussian Area method (A), the Cauchy-Gaussian Peak Height method (B), the GEMG4 Peak Height method (H), and the GEMG4 Area method (G), which produced values for both analytes that are under 10% RSD. The Cauchy-Gaussian Peak Height method produced

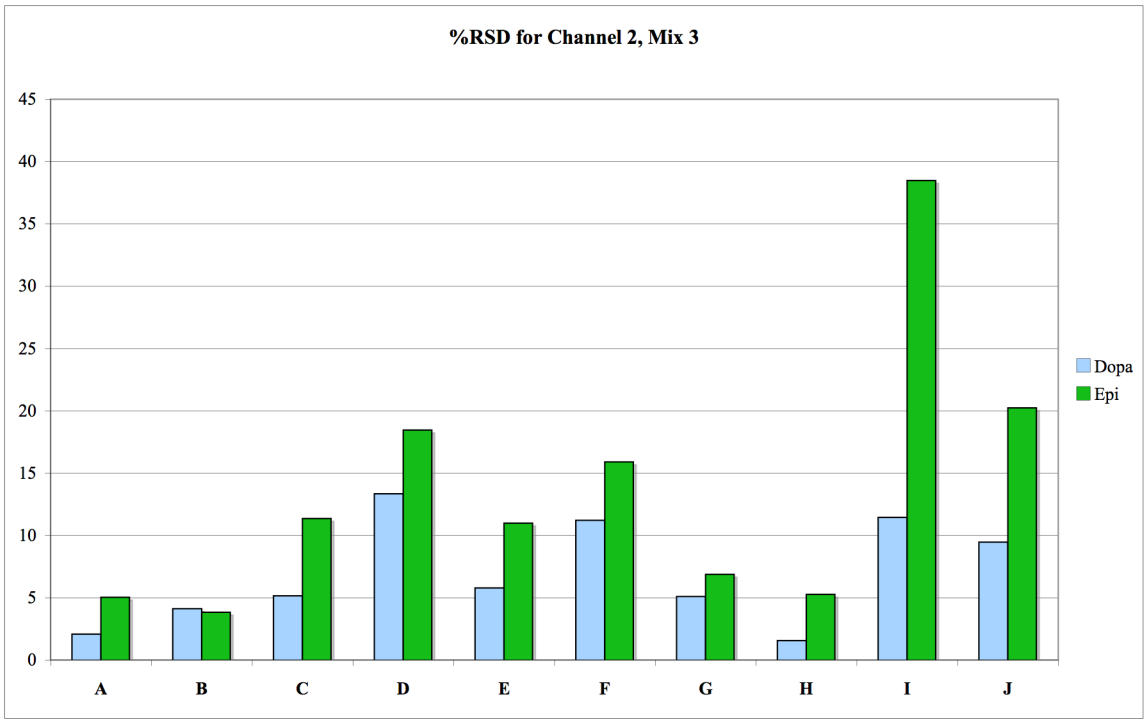


Figure 3-18: %RSD for Dopa, Channel 2, Mix 3

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

values for both analytes that are under 5%, and hence can be said to be the most precise of the curve fitting methods attempted. The %RSD reported by the EMG Peak Height (C) method is twice that found for the Cauchy-Gaussian Area, with about 12% for Epi and 5% for Dopa; the EMG+GMG Peak Height method is higher still. Of the two methods that passed the Student's t-test, the EMG Peak Height is the more precise.

Mix 4

None of the curve-fitting methodologies are able to yield the correct answer to the required levels of accuracy and precision as indicated by the Student's t-test at 95% confidence. The results are discussed below.

The accuracy of the methods attempted in this work is examined below in Figure 3-19. For both Dopa and Epi, the EMG+GMG Area method (E) yields the smallest percent differences between the calculated and known values for the analytes. This

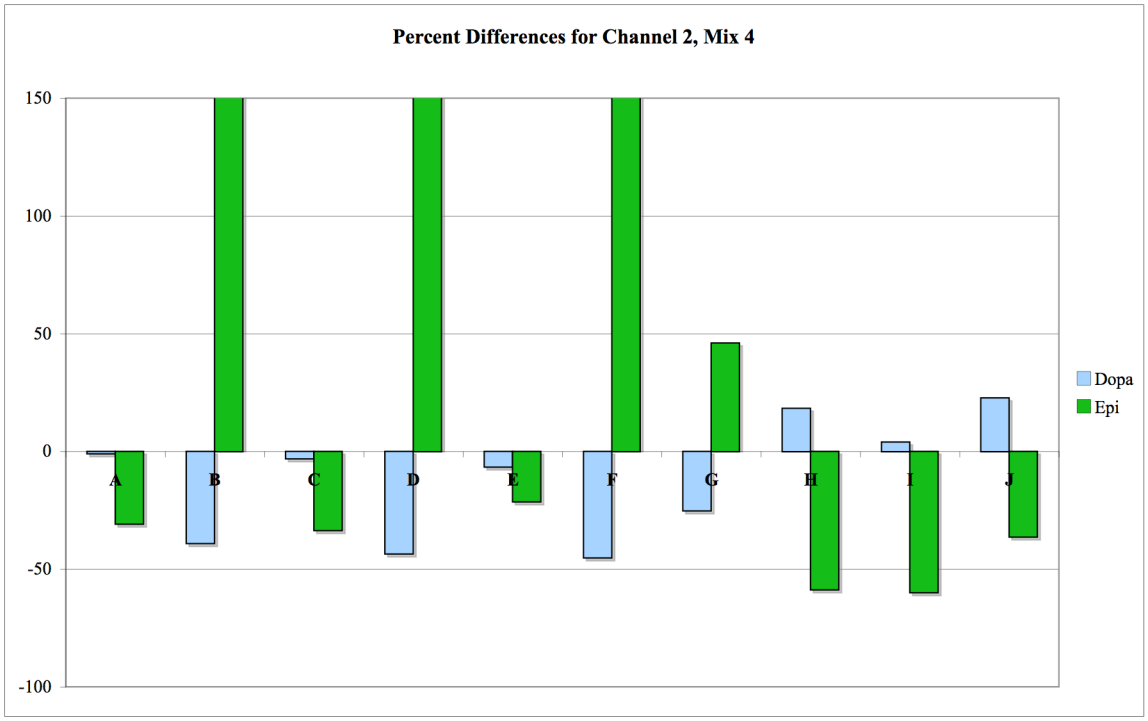


Figure 3-19: Percent Difference for Dopa, Channel 2, Mix 4

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

method under-reported both Dopa and Epi, with Epi unacceptably underestimated by 25%. Thus, these methods cannot be considered to be very accurate for this mixture set.

The results found for Mix 4, Channel 2, are shown in Figure 3-20. For both Dopa and Epi the the GEMG4 Peak Height (H) method gives the best estimate of the values in the mixture set. This was followed closely by the GEMG4 Area method (G). Both these methods gave %RSD values less than 5% for Dopa and less than 8% for Epi.

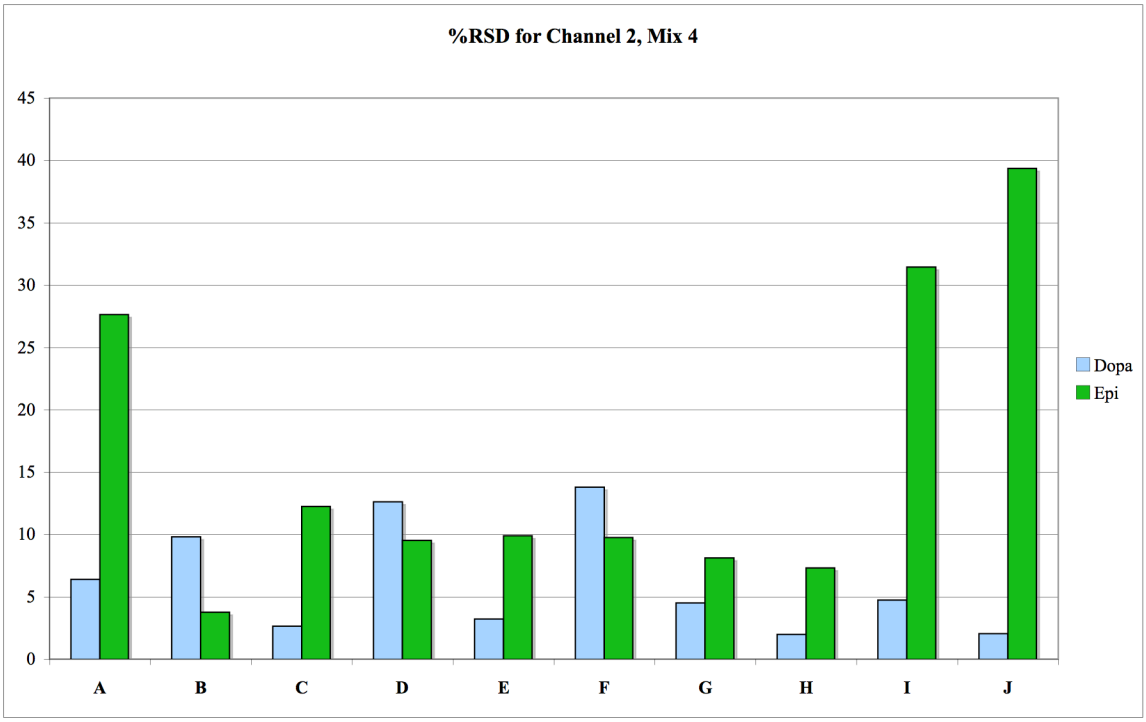


Figure 3-20: %RSD for Dopa, Channel 2, Mix 4

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 5

No curve-fitting methodology was able to deconvolute to the peaks and yield the correct values to 95% confidence, using the Student's t-test. Even with this result, an examination of the percent differences and %RSD values was performed.

The most accurate method can be discerned using Figure 3-21, where the Cauchy-Gaussian Area method (A) is shown. This method yields nearly a 30-35% difference for Epi. This value is not completely unexpected, as Epi is the minor component in this

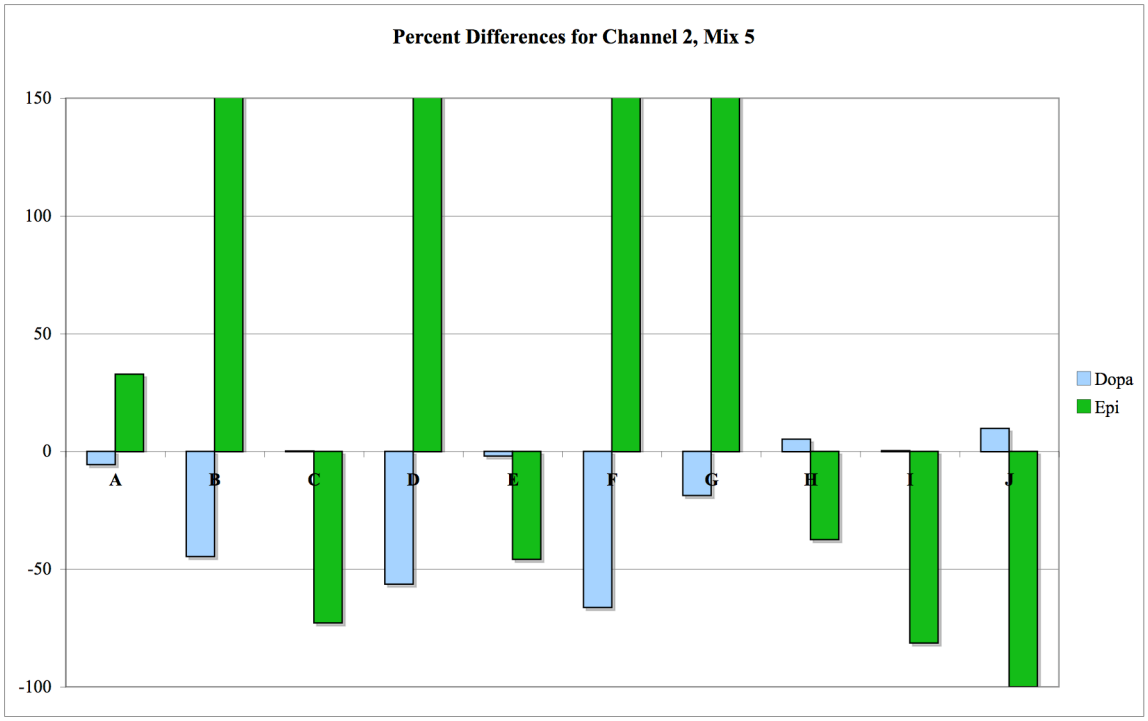


Figure 3-21: Percent Difference for Dopa, Channel 2, Mix 5

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

mixture set. The next most accurate method appears to be the GEMG4 Peak Height method (H), which yielded nearly 40% difference for Epi.

Figure 3-22 show the %RSD values for the two analytes. The methods indicated to be the most precise are the Cauchy-Gaussian Peak Height method (B) and the GEMG4 Peak Height metho (H); these are the only ones to yield results for both Epi and Dopa under 10% RSD.

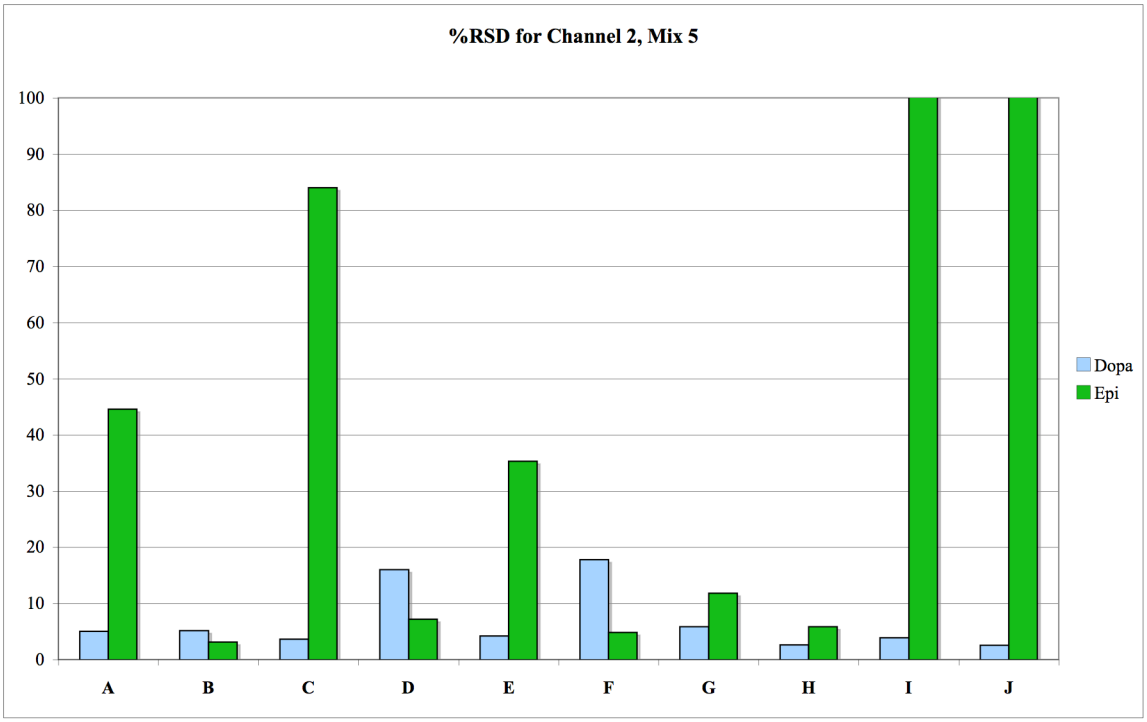


Figure 3-22: %RSD for Dopa, Channel 2, Mix 5

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Discussion

Overall, none of the curve fitting methods employed is able to deconvolute the entire range of molar response values attempted in this work for Mix 2. In fact, the only mixture set that was successfully deconvoluted was Mix 3, utilizing the EMG Peak Height and the EMG+GMG Peak Height methods. Although the EMG Peak Height method was found to be as accurate as the EMG+GMG Peak Height method, it was more precise, and thus is recommended for possible deconvolution of such a mixture. This recommendation, however, must take into account the rather high %RSD values found for Epi (12%).

Channel 3

Channel 3 was set to oxidize the analytes at a voltage of 800mV. At this potential, more compounds will oxidize than at the voltages used for Channels 1 and 2. Thus, more interferences to be expected on this channel than for the previous ones.

The electrode used for Channel 3 is physically located after those used for Channels 1 and 2, some additional differences in the peak shape were expected to occur. Some of the analyte had already been oxidized, but this amount is usually considered minimal given this method of detection. If any cross-talk were to take place between these three electrodes, it would be expected to be seen at the second or even the third electrode due to the flow of the mobile phase across the electrode surfaces. The lengthened flow path that the analyte plug must travel within the electrode cell to the third electrode would be expected contribute to some additional slight broadening of the peak shape. Additionally, the larger voltage applied to the electrode ensured that the

peak height ratios would change as well, although the amount of analyte exposed to the electrode is considered unchanged.

Results

The equations yielding the lowest SSR values for Channel 3 are the GEMG5, the GMG+EMG, the EMG, the GEMG4, and the Cauchy-Gaussaian, as shown below in Figure 3-23.

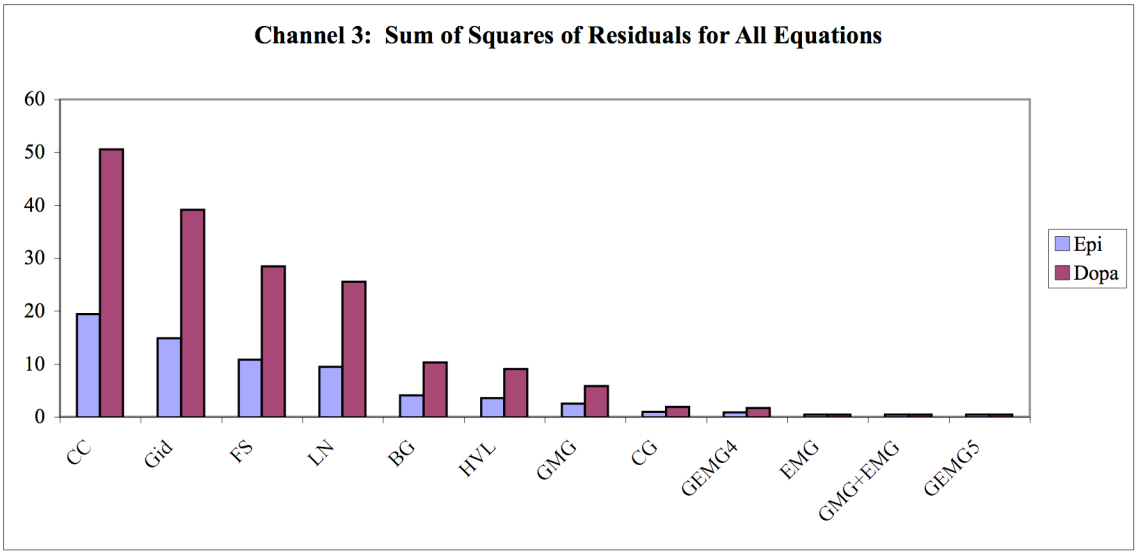


Figure 3-23: SSR results for Channel 3

Key: Gid=Giddings; CC=Chesler-Cram; FS=Frazier-Suzuki; LN=Log Normal; BG=Bi-Gaussian, HVL=Haarhoff-Van der Linde; GMG=Half Gaussian-Modified Gaussian; GEMG4=4 parameter Half Gaussian-Modified Gaussian; GEMG5=5 parameter Half Gaussian-Modified Gaussian; EMG=Exponentially Modified Gaussian; CG=Cauchy-Gaussian; GMG+EMG=Half Gaussian Modified Gaussian + Exponentially Modified Gaussian

Mix 1

For Mix 1, Channel 3, only the GEMG5 Peak Height methodology appears to yield the correct values per the Student's t-test to 95% confidence.

Figure 3-24 shows the most accurate methods of those used in this study to be the EMG+GMG Area method (E), which yields a 20-25% overestimated difference for

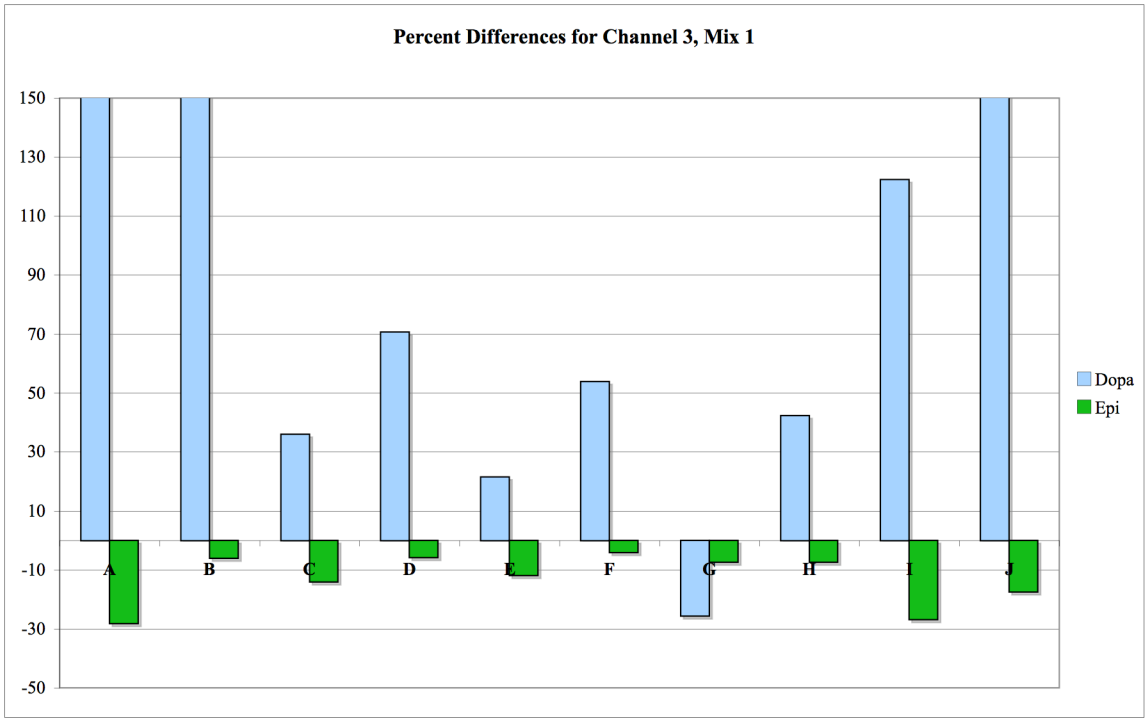


Figure 3-24: Percent Difference for Dopa, Channel 3, Mix 1

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Dopa, and underestimates Epi by some 10%. The next most accurate method was the GEMG4 Area method (G), which underestimates both analytes to some 28% for Dopa and about 8% for Epi. The GEMG5 Peak Height method, which passes the Student's t-test, gives extremely high percent differences for Dopa (over 150%), and thus does not have acceptable accuracy.

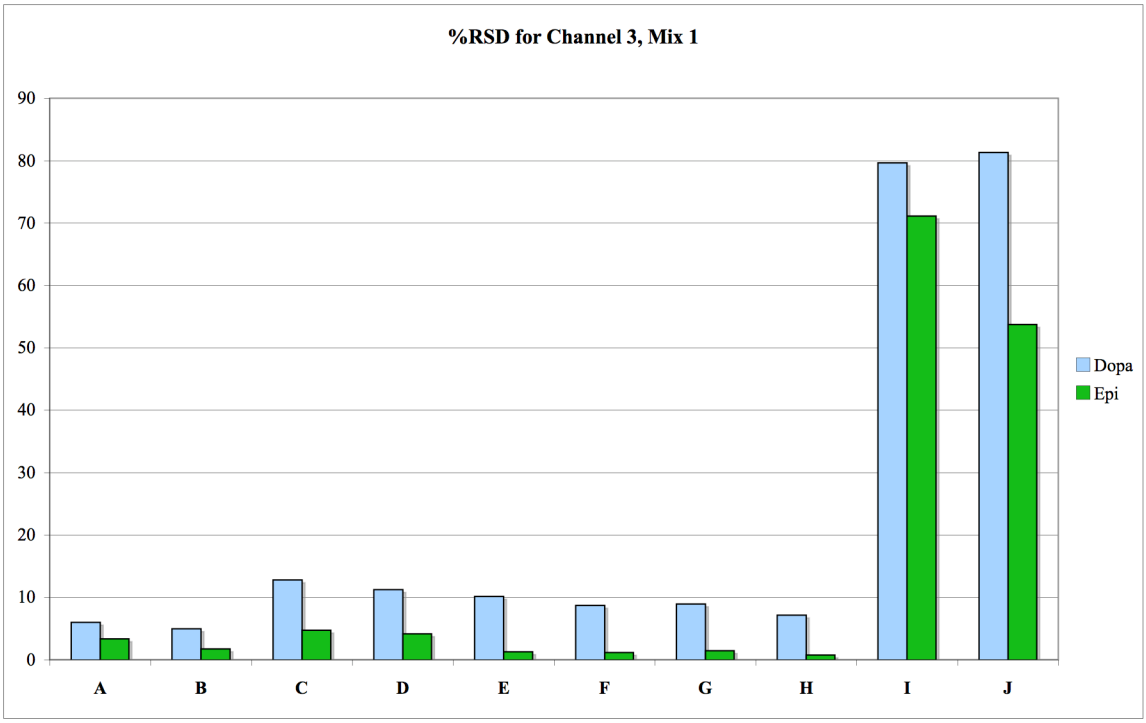


Figure 3-25: %RSD for Dopa, Channel 3, Mix 1

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

The results of the percent difference for Epi shows a negative bias for all the methods; if these differences are due to random error, the bias would be random in nature

and would be expected to be randomly positive and negative. The results here are biased by an as yet undetermined process.

When Figure 3-25 is examined, it becomes apparent why the GEMG5 Peak Height method passed Student's t-test while the other methods did not. The %RSD for this method is over 80% for Dopa and is approximately 53% for Epi. With the exception of the GEMG5 Area (L), all the other methods yield less than 15% RSD values for both Epi and Dopa. Thus, these methods can be considered to be precise, but are not accurate, and thus fail the Student's t-test as the 95% confidence level does not encompass the true values. The large percent difference for the GEMG4 Peak Height coupled with the large %RSD values enables the true value for the analytes to be encompassed within this large range of uncertainty, and thus the GEMG4 Peak Height method, while it passed the Student's t-test, cannot be considered to be either accurate or precise.

Mix 2

Only the GEMG4 Area and the GEMG4 Peak Height methodologies are able to deconvolute the peak as indicated by the Student's t-test at 95% confidence. As was seen for Mix 1 above, the fact that the method passed the Student's t-test does not ensure that it is either accurate or precise.

Figure 3-26 below addresses the question of accuracy. The most accurate methods are the EMG+GMG Area method (E) and the GEMG4 Peak Height method (H), which yields results below 10% for both analytes. The two methods that passed the Student's t-test are the GEMG4 Area (G) and the GEMG4 Peak Height (H). Of these two, the GEMG4 Peak Height is clearly more accurate than the GEMG4 Area.

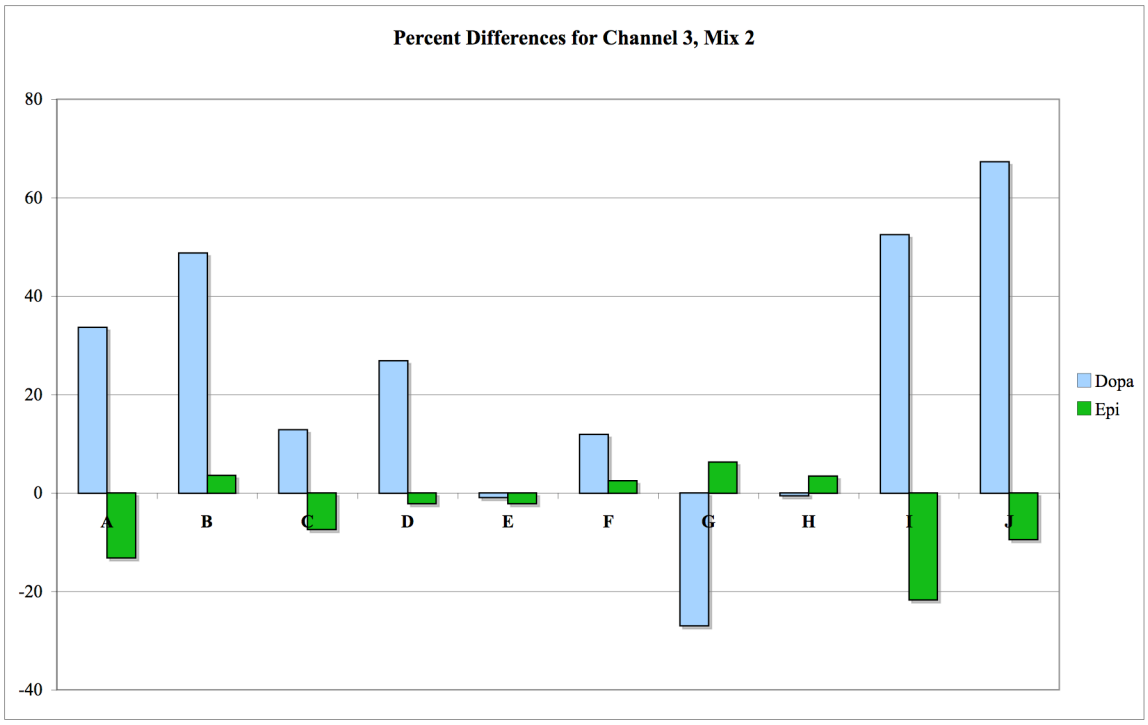


Figure 3-26: Percent Difference for Dopa, Channel 3, Mix 2

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Precision is addressed in Figure 3-27 below. The methods with the lowest %RSD values for both analytes are the EMG+GMG Peak Height (F) which has less than 5% RSD, then the EMG+GMG Area (E), the EMG Peak Height (D), the EMG Area (C), and finally the Cauchy-Gaussian Peak Height (B), all of which were below 10% RSD for both analytes. The two methods that passed the Student's t-test were the GEMG4 Area (G) and the GEMG4 Peak Height (H), both of which are clearly not very precise. Thus, the use of the GEMG4 Peak Height method, which was more accurate than the GEMG4

Area, cannot be considered acceptably precise as it produces %RSD values of approximately 41% for Dopa and about a 21% for Epi, the major component in the mix.

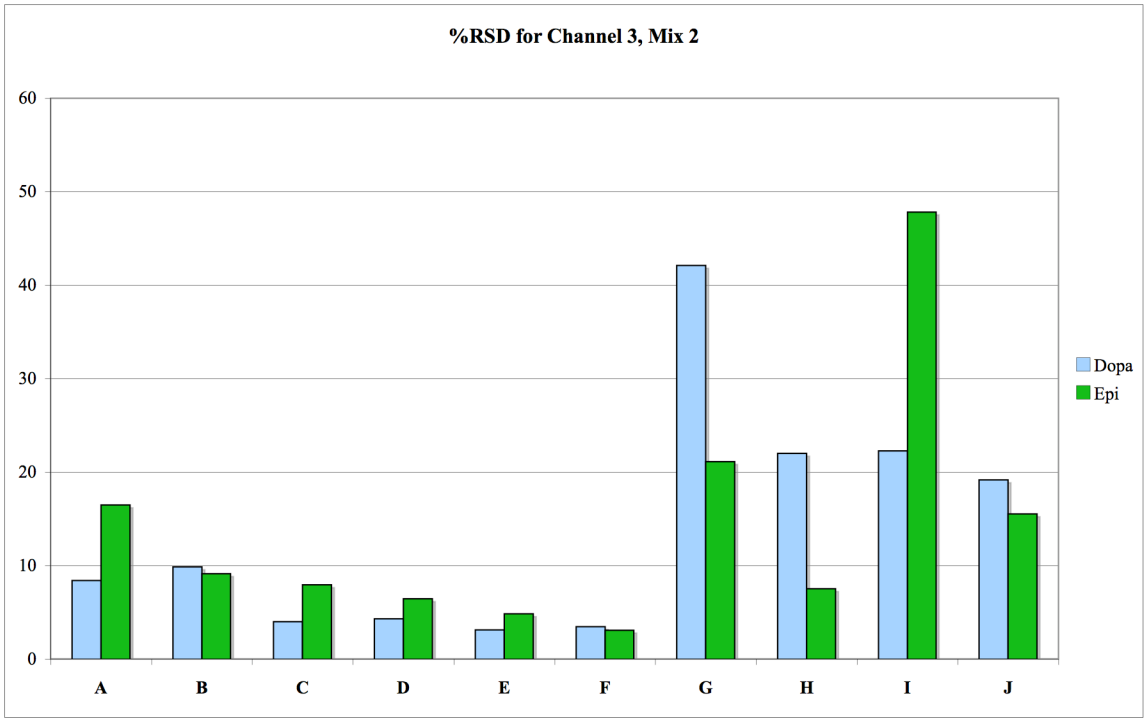


Figure 3-27: %RSD for Dopa, Channel 3, Mix 2

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 3

For both analytes, only the EMG Peak Height and the GEMG5 Area methodologies are able to deconvolute the peak, as shown by Student's t-test at 95% confidence. As shown in Figure 3-28, the EMG Peak Height (D) method and the GEMG5 Area method (I) yield acceptable results, but are not the most accurate of the

methods investigated here. The EMG Area (C), the EMG+GMG Area (E), and EMG+GMG Peak Height (F) methods all yields percent differences for both analytes that are under 5%.

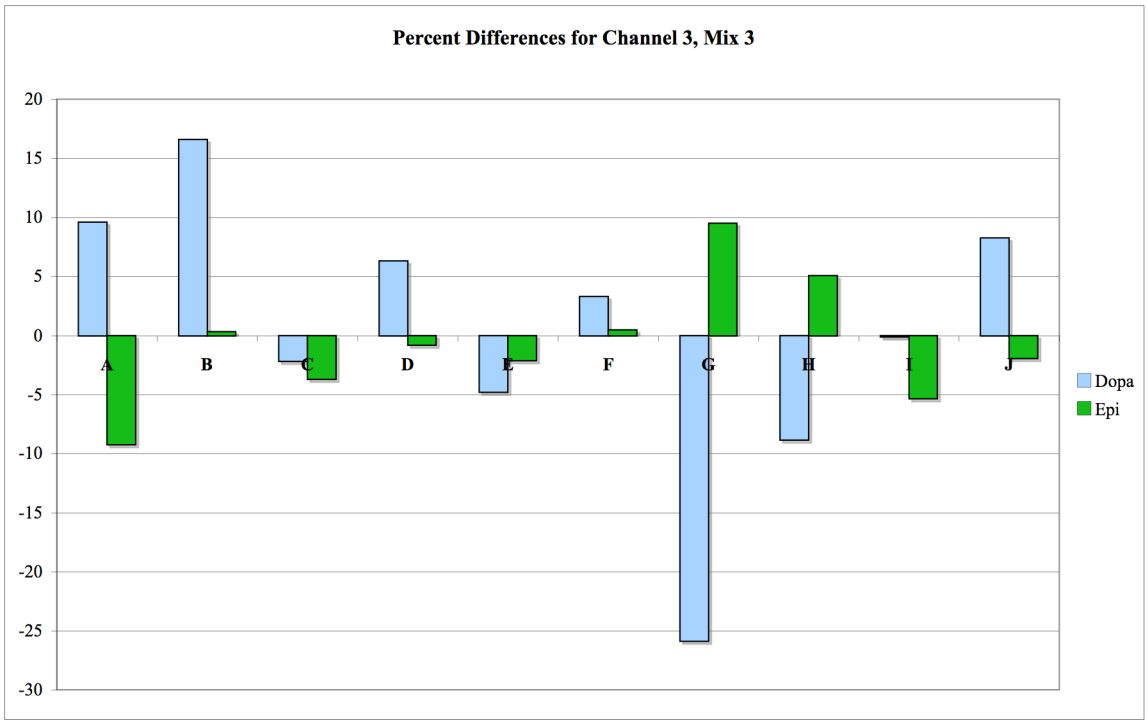


Figure 3-28: Percent Difference for Channel 3, Mix 3

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

The precision of the methods for Channel 3, Mix 3, are shown in Figure 3-29 below. Both the EMG Peak Height (D) and the GEMG5 Area (I) methodologies, which the Student's t-test and yield acceptably accurate results; however, when the precision of the results is examined, it is clear that the EMG Peak Height method is superior to the

GEMG5 Area. The EMG Peak Height method yielded %RSD values under 10% for both analytes, while the GEMG5 Area method yielded %RSD values of just over 10% for Dopa and about 19% for Epi. The EMG Area (C), the EMG+GMG Area (E), and EMG+GMG Peak Height (F) methods also are accurate.

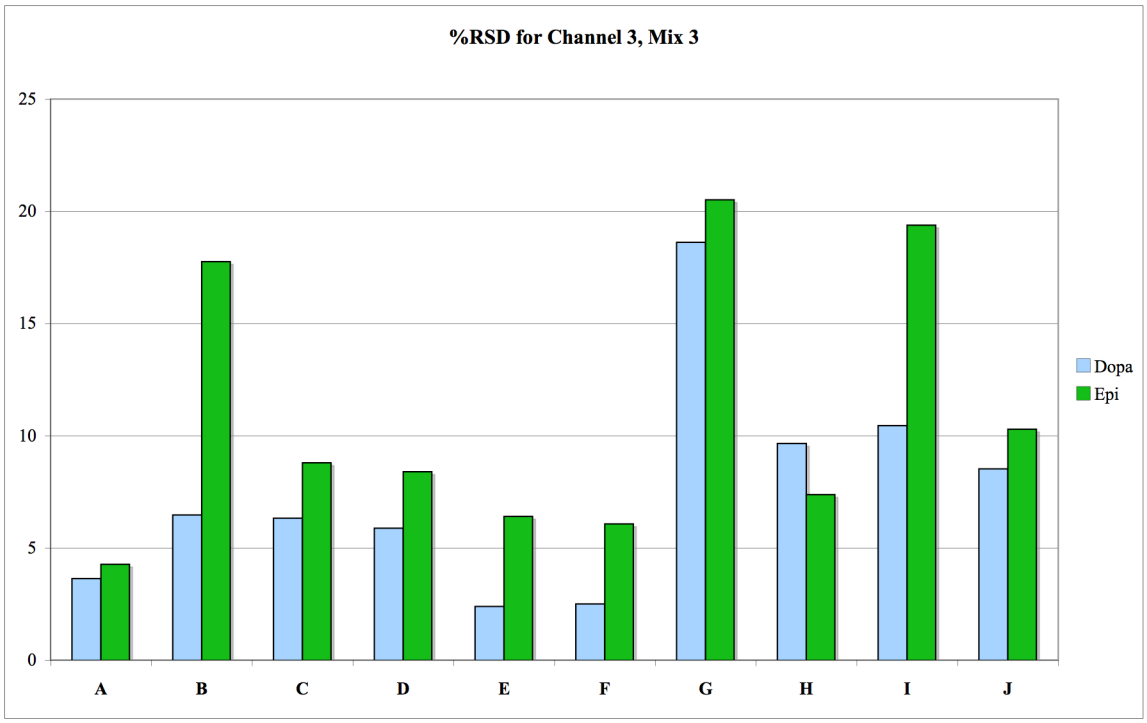


Figure 3-29: %RSD for Channel 3, Mix 3

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 4

The Student's t-test showed that the GEMG5 Area, GEMG5 Peak Height, and the EMG+GMG Peak Height methodologies all yield results that are within the 95% confidence level. The accuracies of these methods are shown in Figure 3-30. The lowest percent differences for both Dopa and Epi are the EMG+GMG Peak Height method (F) and the EMG Area method (C), both of which yielded values less than 5% different from the known values for both analytes. The GEMG5 Area method yields results that are

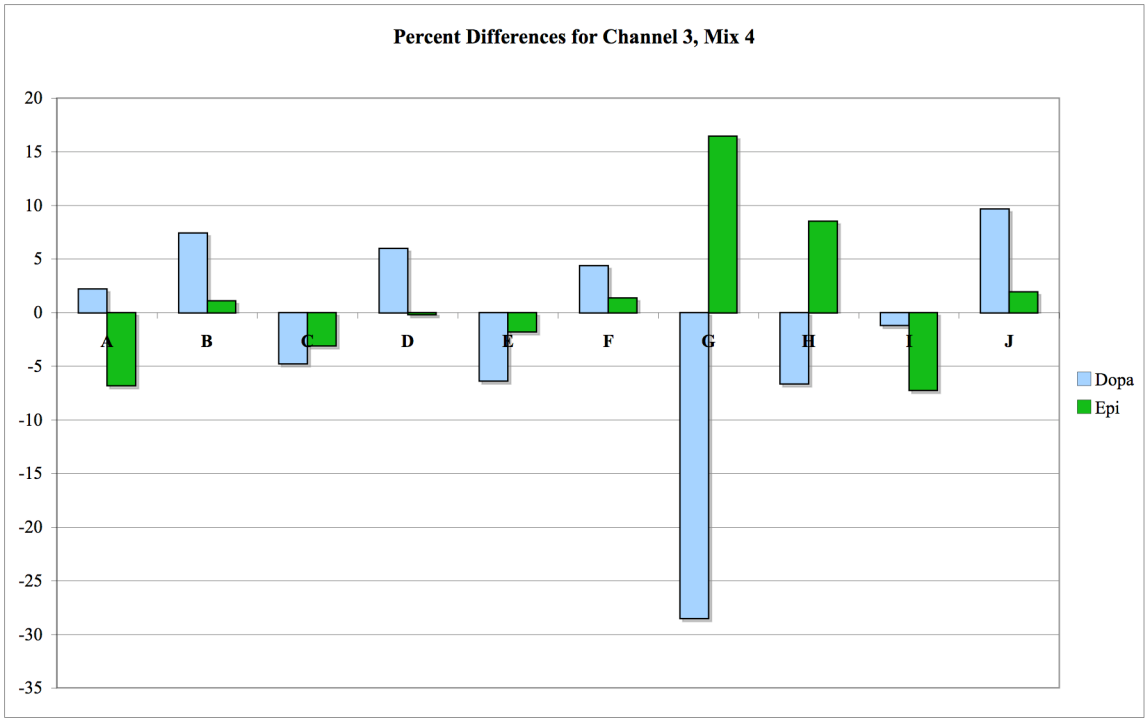


Figure 3-30: Percent Difference for Channel 3, Mix 4

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

negatively biased, and the GEMG5 Peak height method is positively biased; both these methods yielded results under 10%.

The %RSD values for Mix 4 are shown in Figure 3-31. The GEMG5 Area (I), GEMG5 Peak Height (J), and the EMG+GMG Peak Height (F) all yield results that are within the 95% confidence level for the Student's t-test. No significant difference can be found between the GEMG5 Area and the GEMG5 Peak Height methods; both yielded large %RSD values (nearly 60%) for Epi. The EMG+GMG Peak Height method yielded

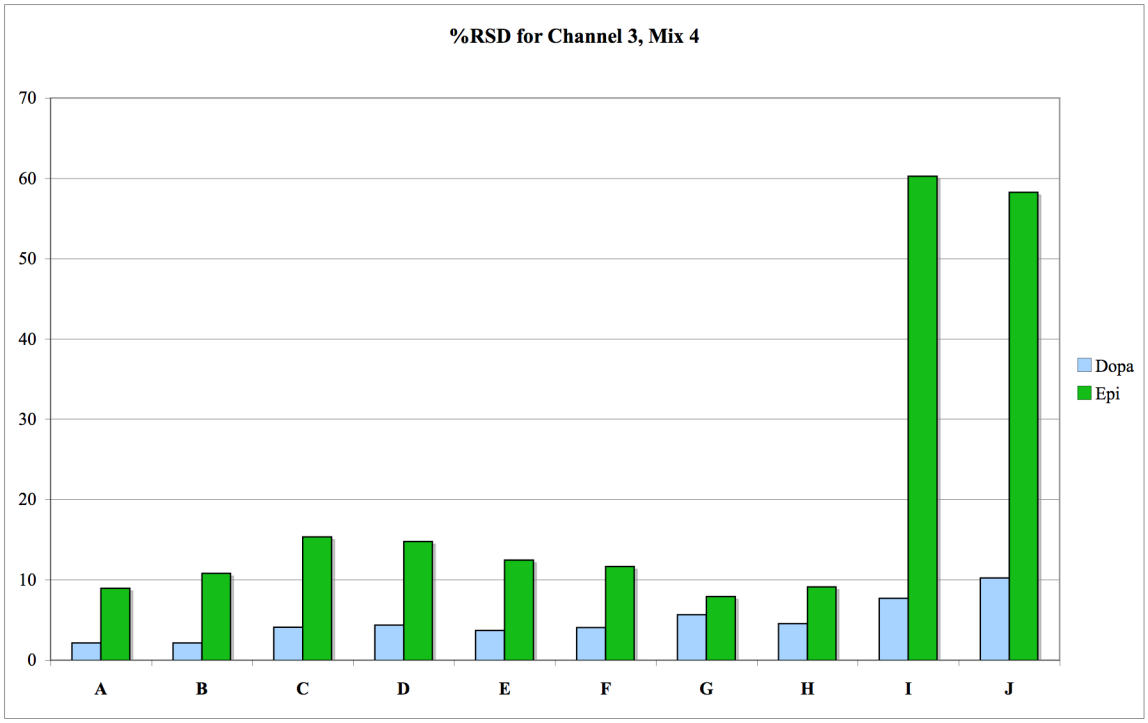


Figure 3-31: %RSD for Channel 3, Mix 4

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

some 5% RSD for Dopa and some 12% for Epi, and thus was much more accurate than the other two. Thus, the EMG+GMG Peak Height is both accurate and precise when used for this mixture set.

Mix 5

For both analytes, none of the curve-fittting methodologies are able to produce the correct results at 95% confidence, using the Student's t-test. However, the question of accuracy and precision in the methods was still investigated.

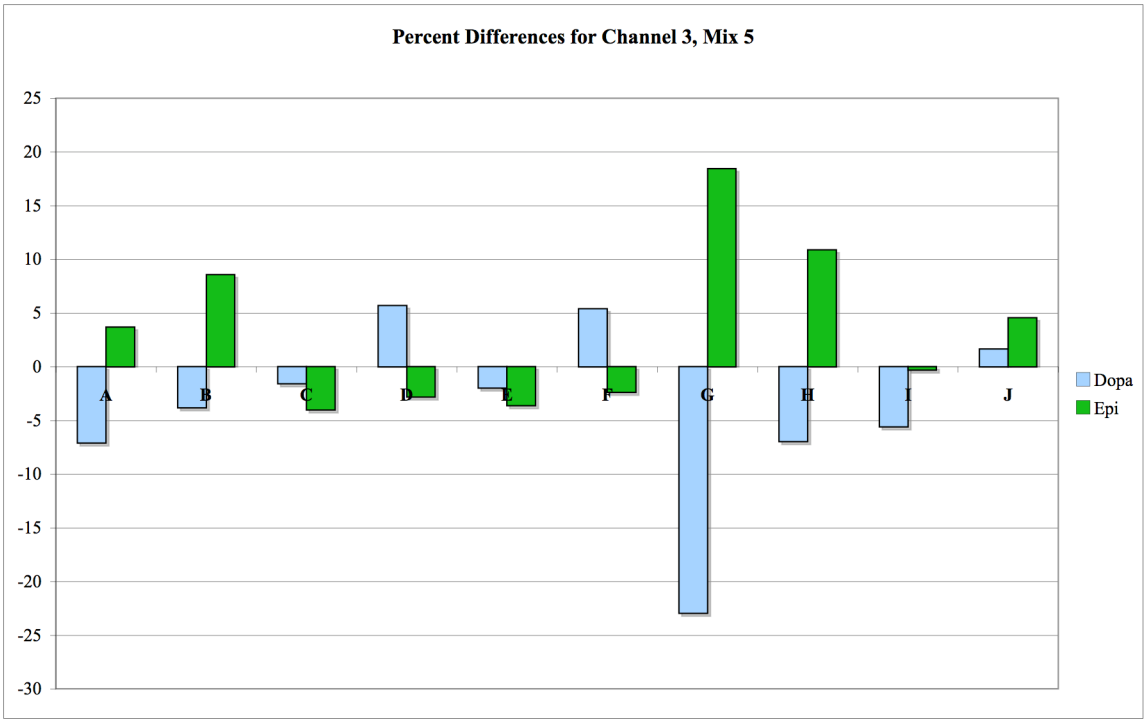


Figure 3-32: Percent Difference for Channel 3, Mix 5

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

The percent difference figure, Figure 3-32 above, shows that the most accurate methods are the EMG Area (C), the EMG+GMG Area (E), and the GEMG5 Peak Height (J). These methods return percent difference values for both analytes under 5%, and hence are considered to be accurate.

The precision of the methods is shown in Figure 3-33 below. The methods yielding the lowest %RSD for the mixture is the GEMG4 Area (G), and then the GEMG4 Peak height (H). The methods considered accurate in the analysis above are the EMG

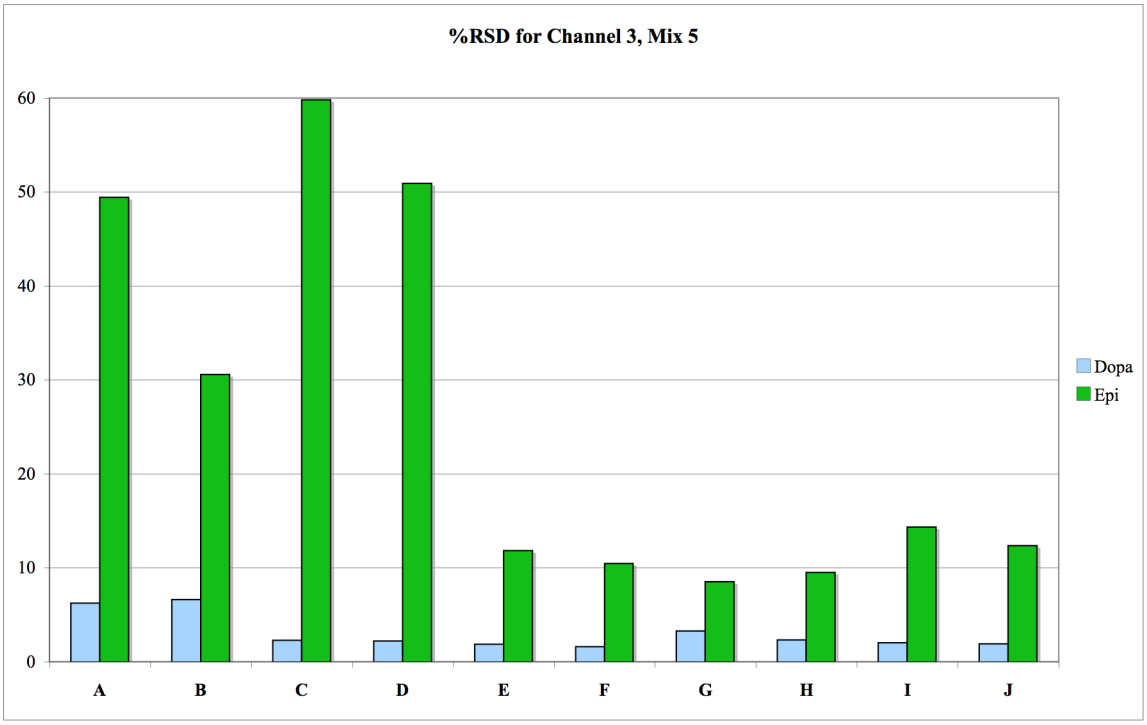


Figure 3-33: %RSD for Channel 3, Mix 5

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Area (C), the EMG+GMG Area (E), and the GEMG5 Peak Height (J), all of which returned %RSD values for Epi that are above 10%. The EMG Area was particularly imprecise for this analyte, returning a %RSD value of nearly 60%. This is not very surprising, considering that Epi is the minor component in the peak.

Discussion

For Channel 3, no single method was able to deconvolute the entire range of molar responses attempted in this work. For Mix 1, the GEMG5 Peak Height passes the Student's t-test, but has very poor accuracy and precision. For Mix 2, the GEMG4 Area and GEMG4 Peak Heights pass the Student's t-test. Of these two, the GEMG4 Peak Height has good accuracy, but poor precision. For Mix 3, two methods passes the t-test: the EMG Peak Height and the GEMG5 Area. The EMG Peak Height method is as accurate as the GEMG5 Area, but has better precision. For Mix 4, the EMG+GMG Peak Height method has good accuracy and good precision, and the GEMG5 Area and the GEMG5 Peak Height methods have good accuracy but very poor precision. Finally, no method was able to deconvolute the last mixture set, Mix 5, to 95% confidence by the Student's t-test. As no single method was able to consistently deconvolute the range of molar responses, no single curve-fit approach can be recommended for the data on Channel 3.

Channel 4

Channel 4 was set to oxidize the analytes at a voltage of 900mV, and is the last electrode in the series. Again, some differences in the peak shape are expected to occur, as some of the analyte has already been oxidized; however, this amount is usually considered minimal given this method of detection. If any cross-talk were to take place between the other electrodes, it would be expected to be seen here as well due to the flow of the mobile phase across the electrode surfaces. The longer path the analyte plug must travel within the electrode cell contributed to some slight broadening of the peak shape. Additionally, the larger voltage applied to the electrode ensured that the peak height ratios would change as well, although the amount presented to the electrode would be considered to be the same. It is on this channel that the maximum amount of interference from other compounds would be expected.

Results

The equations yielding the lowest SSR values are the GEM5, the GMG+EMG, the EMG, the GEMG4, and the Cauchy-Gaussian equations, as shown below in Figure 3-34, with the EMG and the GMG+EMG equations fitting particularly well.

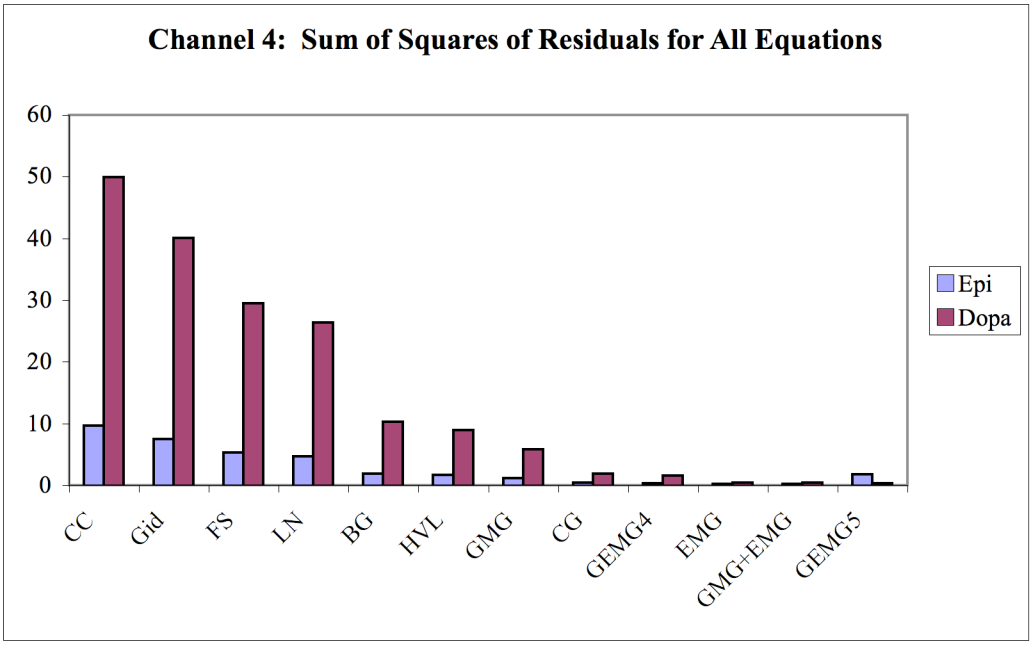


Figure 3-34: SSR results for Channel 4

Key: Gid=Giddings; CC=Chesler-Cram; FS=Frazier-Suzuki; LN=Log Normal; BG=Bi-Gaussian, HVL=Haarhoff-Van der Linde; GMG=Half Gaussian-Modified Gaussian; GEMG4=4 parameter Half Gaussian-Modified Gaussian; GEMG5=5 parameter Half Gaussian-Modified Gaussian; EMG=Exponentially Modified Gaussian; CG=Cauchy-Gaussian; GMG+EMG=Half Gaussian Modified Gaussian + Exponentially Modified Gaussian

Mix 1

For Mix 1, Channel 4, unfortunately, none of the methods used in this study are able to return the correct value within 95% confidence for either Epi or Dopa. To answer the question of accuracy, Figure 3-35 above shows the percent differences for Dopa and Epi. The results for Epi showed a definite negative bias in the results, indicating that the errors associated with the data are not random. None of these methods can be considered

as accurate; the method yielding the lowest percent difference for Dopa produced a value of about 20%, and the lowest for Epi about 5%. The high percent differences for Dopa is not surprising, as Dopa is a minor component in the mixture peak.

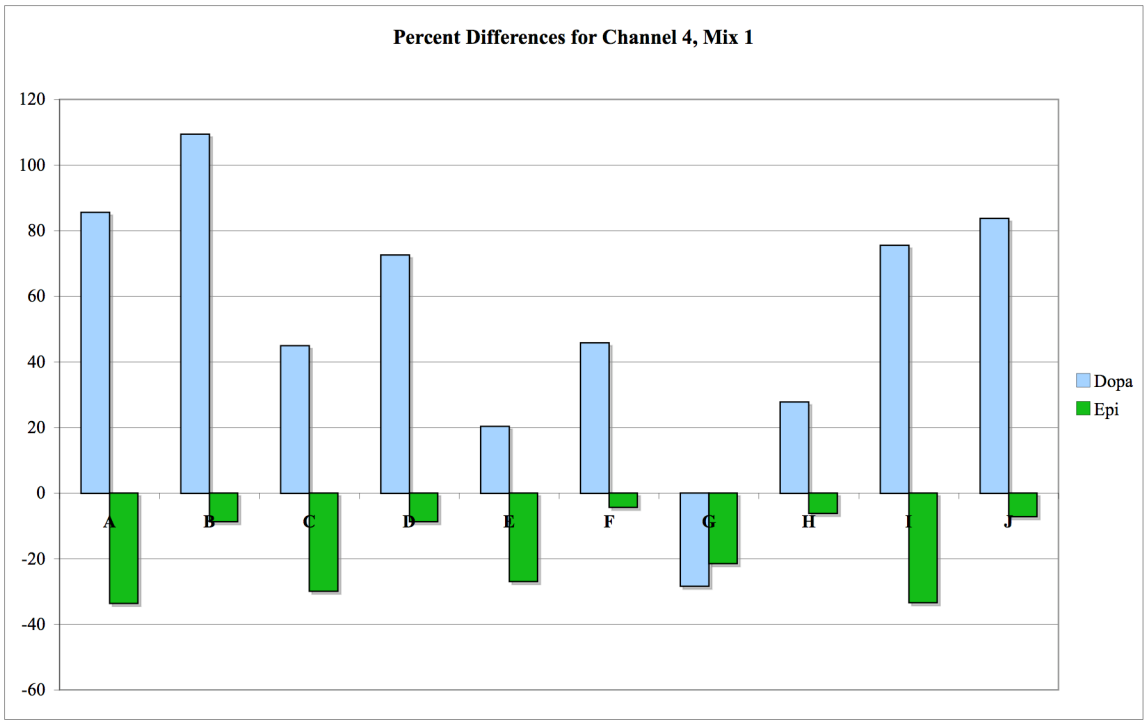


Figure 3-35: Percent Difference for Channel 4, Mix 1

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

The precision found for the methods is shown below in Figure 3-36 below. All of the methods returned a %RSD value of less than 12% for Dopa, and values of less than 9% for Epi. Overall, the precision found for the methods for Mix 1, Channel 4 are good,

with the EMG Peak Height (D) considered to be the most precise of the methods for this mixture set.

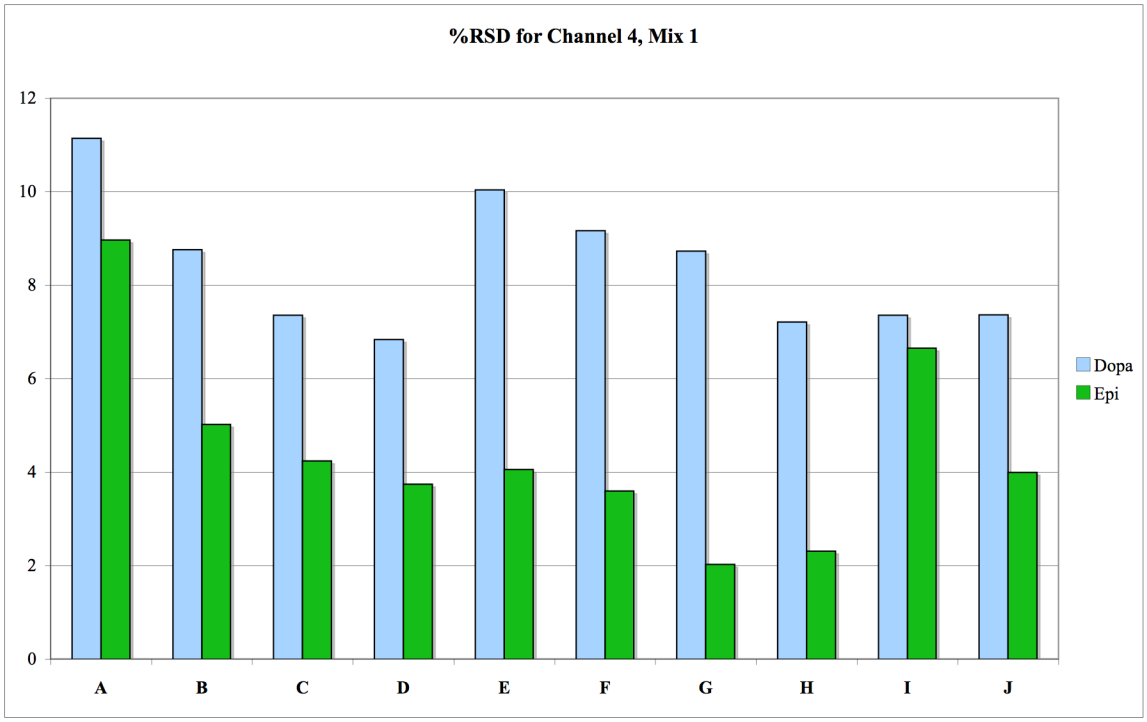


Figure 3-36: %RSD for Channel 4, Mix 1

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 2

No one method was able to deconvolute the correct values for both the analytes as shown by the Student's t-test analysis at 95% confidence. However, the question will arise whether these methods failed from imprecision in accuracy, in precision, or in both. Figure 3-37 below shows the percent differences found for Dopa and Epi. The GEMG4

Peak Height method (H) produces the best overall percent differences for the mixture set, at about 6% for Epi and some 8-9% for Dopa. The next most accurate method is found to be the EMG+GMG Peak Height method, which yields low values for Epi, and high values for Dopa.

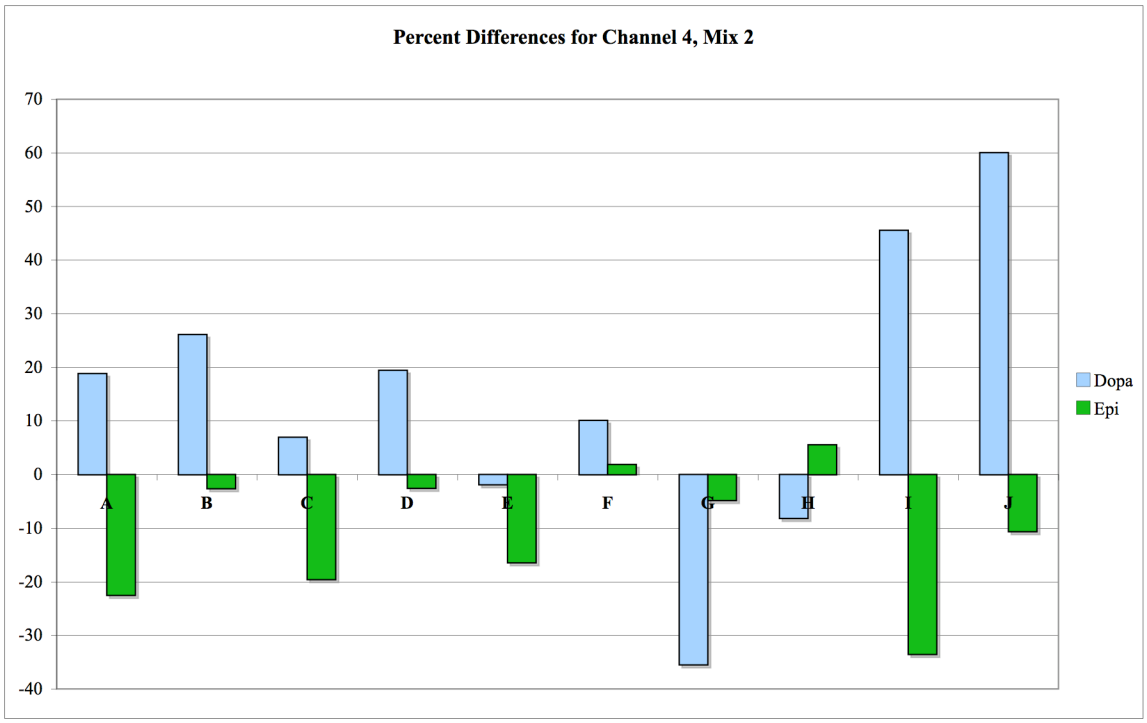


Figure 3-37: Percent Difference for Channel 4, Mix 2

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

The precision of the methods can be seen in Figure 3-38. The methods yielding the lowest %RSD values for the mixture peak are: the EMG Peak Height (D), the GEMG4 Area (G), the GEMG4 Peak Height (H), and the Cauchy-Gaussian Peak Height

(B). The GEMG4 Peak Height method (H) returns the best overall percent differences for the mixture set, followed by the EMG+GMG Peak Height method. Of these two, the GEMG4 Peak Height method appears to be more accurate and more precise; however, it still does not meet the Student's t-test criteria.

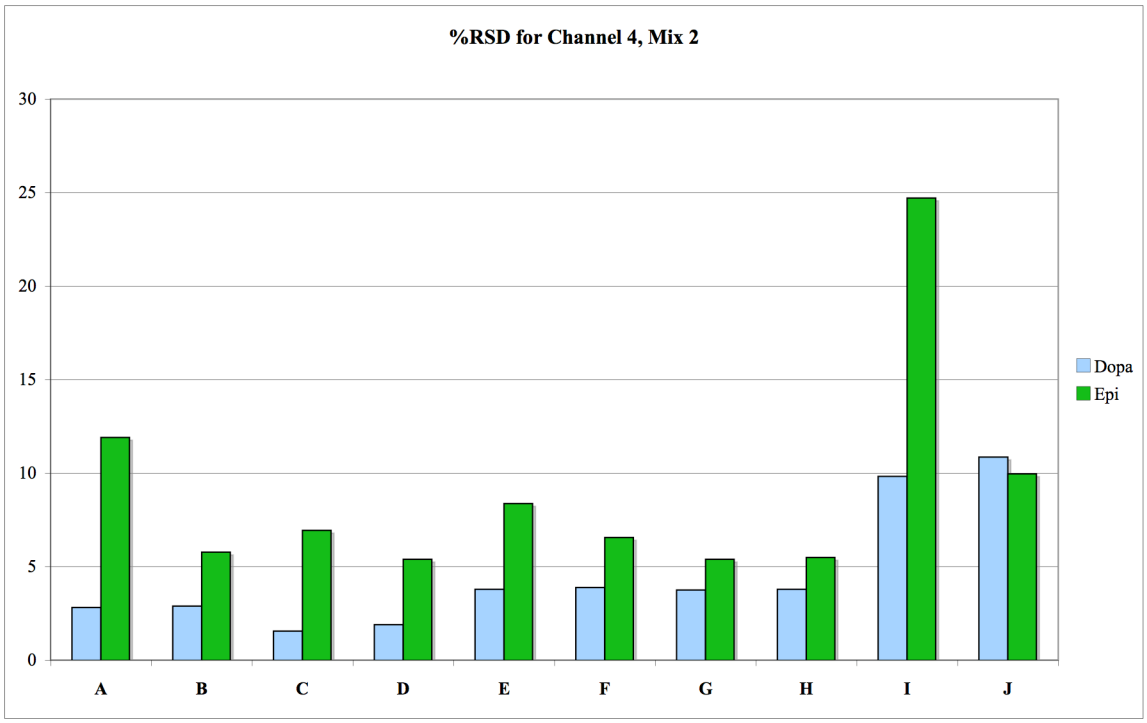


Figure 3-38: %RSD for Channel 4, Mix 2

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 3

The EMG Peak Height method is the only method that is able to successfully deconvolute the amounts of Epi and Dopa to 95% certainty, as indicated by the Student's

t-test. The percent differences for all the methods for Dopa and Epi in Mix 3 are shown below in Figure 3-39. The EMG Peak Height method is clearly more accurate than any of the other methods, yielding percent differences less than 4% for each analyte.

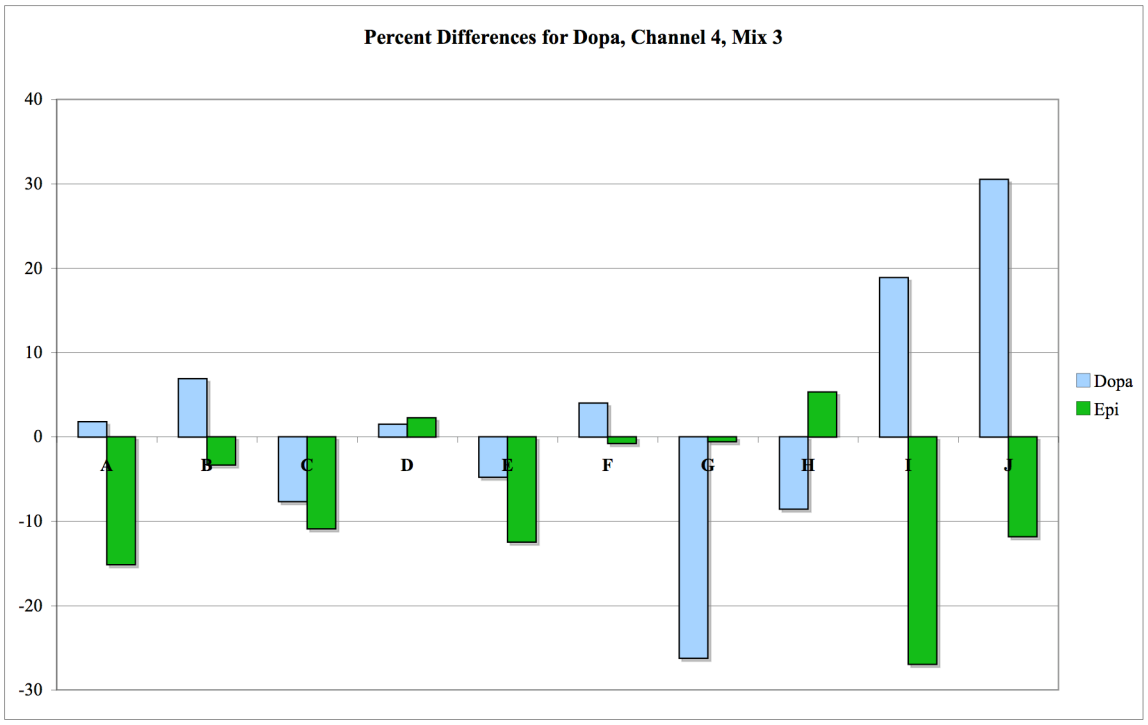


Figure 3-39: Percent Difference for Channel 4, Mix 3

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

The %RSD values returned by the methods are presented in Figure 3-40. The methods with the lowest overall %RSD values for the two analytes is the Cauchy-Gaussian Area (A), followed by the Cauchy-Gaussian Peak Height (B), the GEMG4 Peak Height (H), the GEMG4 Area (G), and then the EMG Peak Height (D). The EMG Peak

Height method is not the most precise of the methods, and thus can be considered accurate and relatively imprecise, as the %RSD values it returns are an acceptable value of about 3.5% for Dopa and an unacceptable value of 10% for Epi.

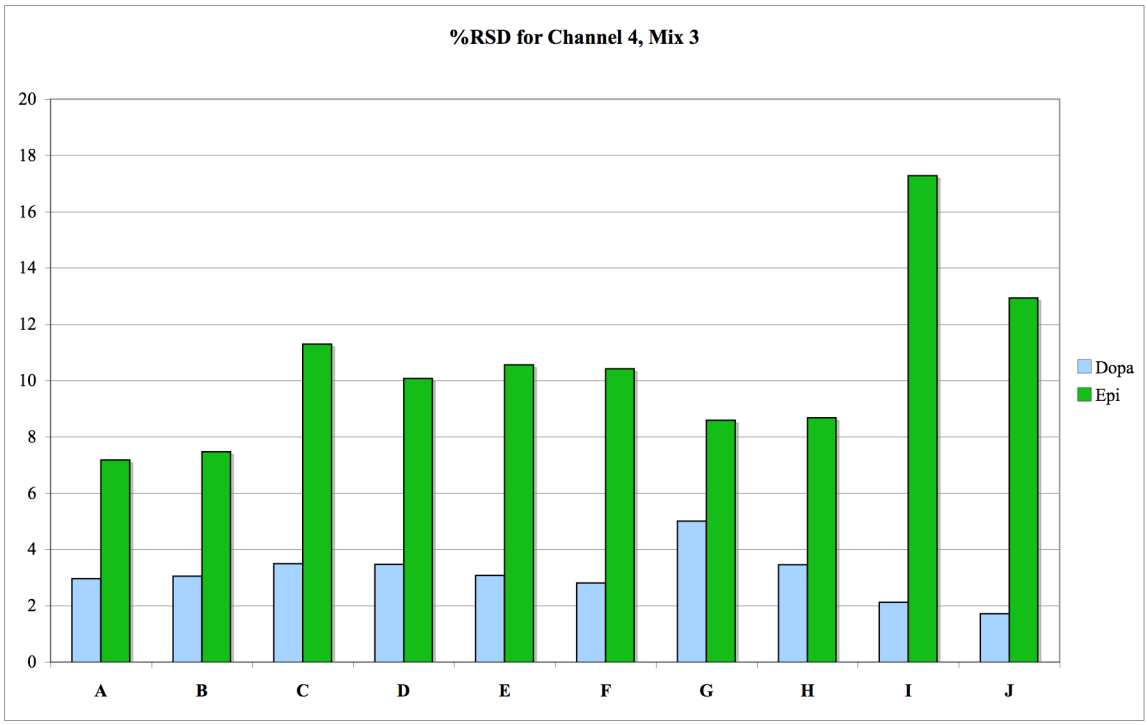


Figure 3-40: %RSD for Channel 4, Mix 3

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Mix 4

For this mixture set, no curve-fitting method was able to deconvolute the peak and give the correct values for Epi and Dopa to 95% confidence as per the Student's t-test.

To examine accuracy and precision, the percent differences for Mix 4 are shown below in Figure 3-41. In this mixture set, the most accurate methods are the EMG+GMG Peak Height (F), the EMG Peak Height (D), and the Cauchy-Gaussian Peak Height (B) all of which return values less than 5% for both analytes.

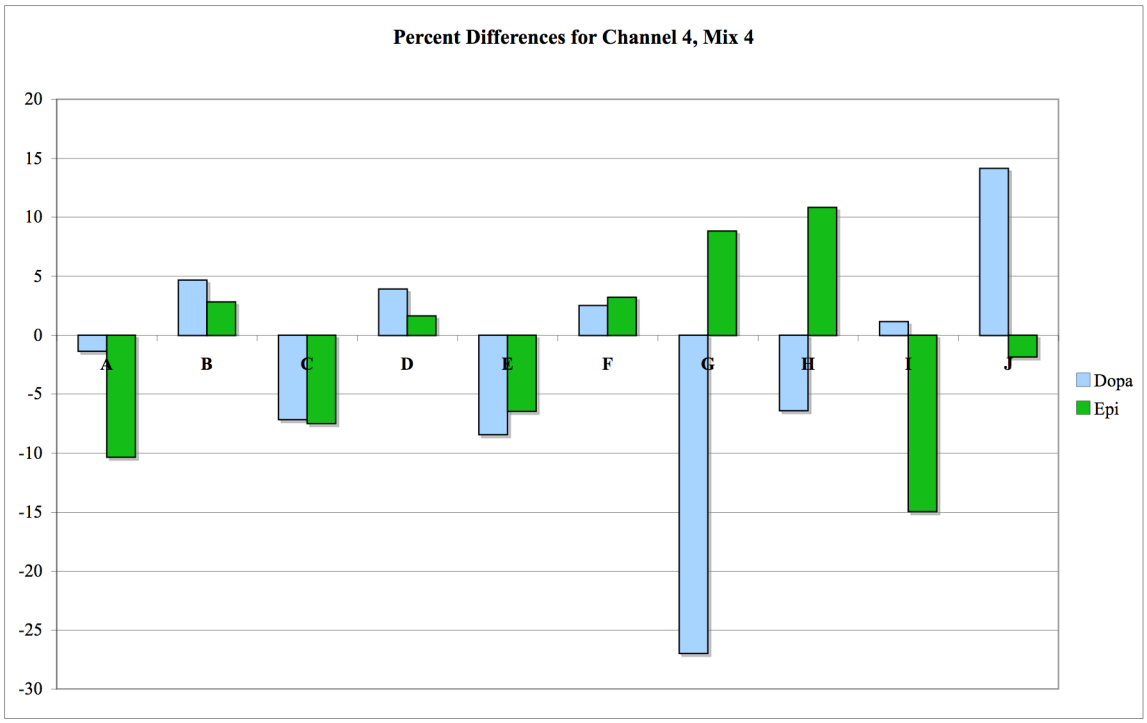


Figure 3-41: Percent Difference for Channel 4, Mix 4

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

An examination of Figure 3-42 shows that the precision of the GEMG4 Area method (G) is best for Epi (approximately 11%) but, however, it is the worst performer for Dopa (approximately 5.5%). A fair compromise for precision for both analytes would

be the EMG Area method (C) or the EMG Peak Height method (D), where the %RSD for Epi is about 12 to 13% and the %RSD for Dopa is about 3%.

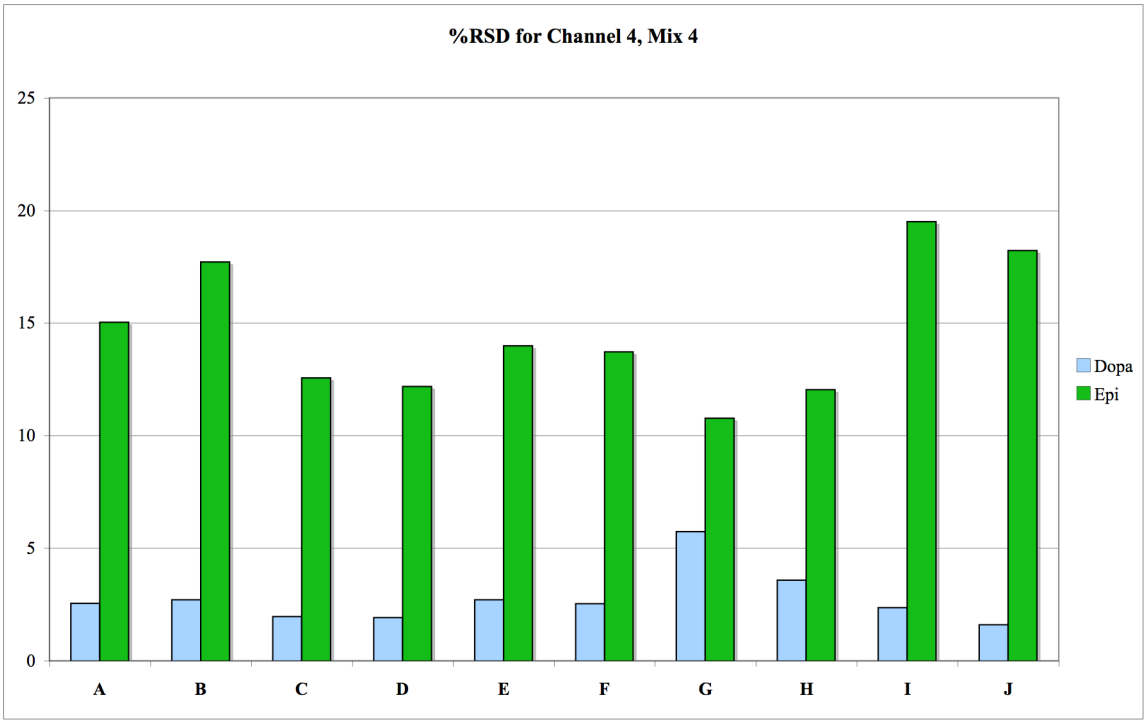


Figure 3-42: %RSD for Channel 4, Mix 4

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Overall, the EMG Area (C), the EMG Peak Height (D), the EMG+GMG Area, (E), and the EMG+GMG Peak Height (F) methods are fairly accurate and precise, but not so accurate or so precise that they are able to satisfy the Student's t-test at 95% confidence. Therefore, none of the methods attempted for this mixture set can be recommended.

Mix 5

For Mix 5, Channel 4, the results of the Student's t-test indicated that none of the methods used in this study are able to deconvolute the peak into its component analytes. The question of accuracy and precision still remains.

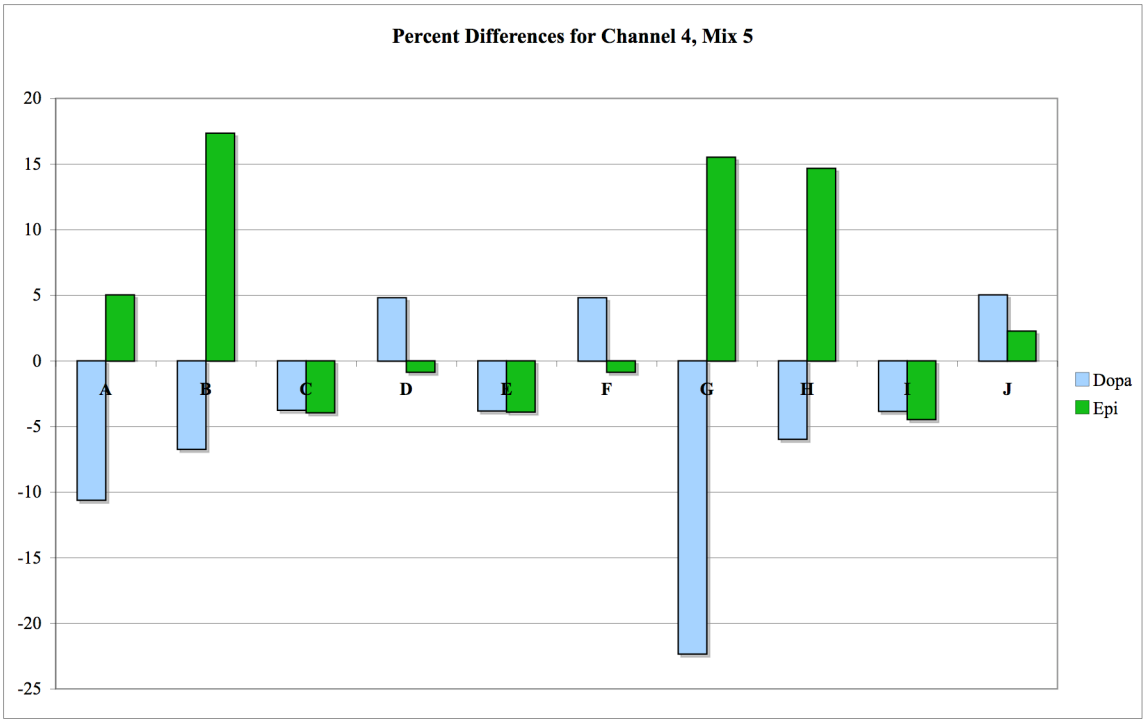


Figure 3-43: Percent Difference for Dopa, Channel 4, Mix 5

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

As can be seen in Figure 3-43 above, for this mixture containing both Epi and Dopa, the EMG Area (C), the EMG Peak Height (D) , the EMG+GMG Area (E), the

EMG+GMG Peak Height (F), the GEMG5 Area (I), and GEMG5 Peak Height (J) would be all good choices based on accuracy. All these methods yield percent differences that are within ± 5%.

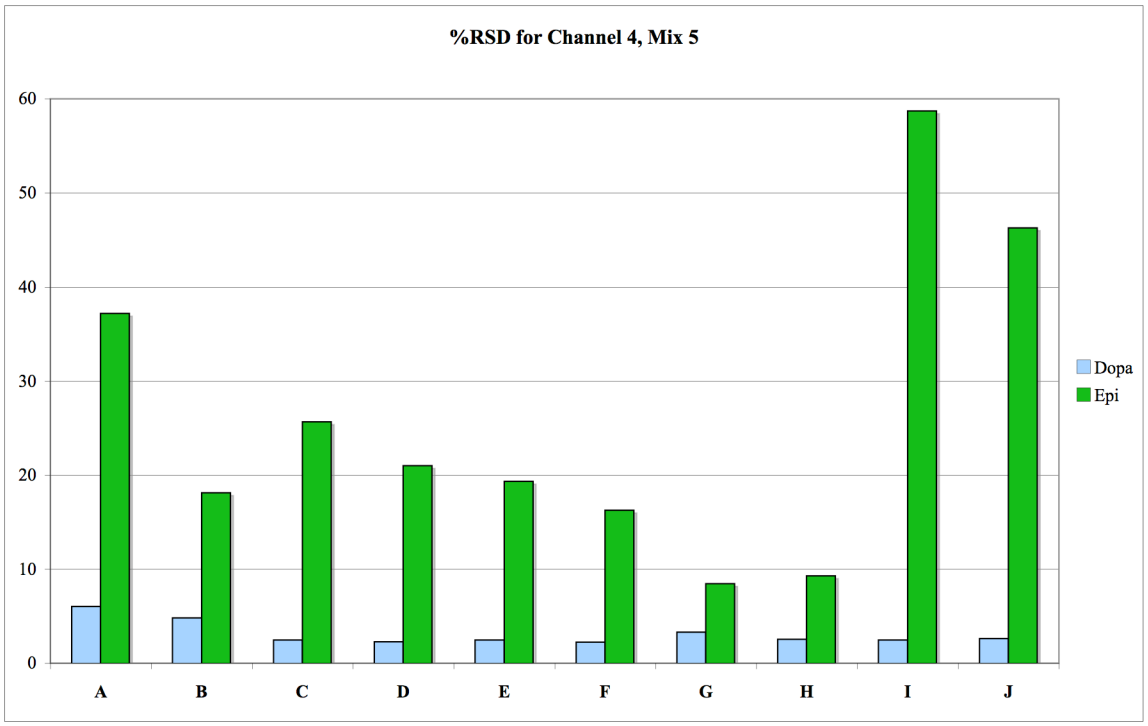


Figure 3-44: %RSD for Channel 4, Mix 5

Key: A: Cauchy-Gaussian Area; B: Cauchy-Gaussian Peak Height; C: EMG Area; D: EMG Peak Height; E: EMG+GMG Area; F: EMG+GMG Peak Height; G: GEMG4 Area; H: GEMG4 Peak Height; I: GEMG5 Area; J: GEMG5 Peak Height

Figure 3-44 above shows the %RSD values for both Dopa and Epi. For Dopa, all of these are reasonable, with the highest %RSD being returned using the Cauchy-Gaussian Area method (A). The rest of the methods return values of less than 5% RSD. High %RSD values are shown for Epi, with the lowest being returned from the GEMG4

Area (J) at about 8% and the GEMG4 Peak Height (K) methods at almost 10%. The worst performers are the GEMG5 Area (L) and the GEMG5 Peak Height (M) methods at about 58% and 45%, respectively. Neither of the GEMG5 methods are able to acceptably deconvolute the peak. The most precise method is the GEMG4 Area, followed by the GEMG4 Peak Height method.

Discussion

For Channel 4, no single method was able to deconvolute the entire range of molar response ratios attempted. In fact, the only mixture set that was able to be deconvoluted was Mix 3. The EMG Peak Height method deconvoluted this mixture set with excellent accuracy and good precision. However, one successful instance of deconvolution out of five attempts does not make an endorsement of a method, and so no method can be recommended.

Curve-Fitting Methods: Summary and Conclusions

The data is summarized in Table 3-1. This table summarizes the charts shown earlier in the chapter, giving the data in a numerical fashion rather than visually. The methods that passed the Student's t-test are bolded. As can be seen in this table, for all the mixture sets on all four channels, no single curve fitting method was able to properly deconvolute the peaks.

Table 3-1: Summary of the data

n = 6, t_{Table} = 2.57		Mix 1					Mix 2					Mix 3					Mix 4					Mix 5				
Channel	Technique	Epi Ave	Epi SD	t_{Cal}	Dopa Ave	Dopa SD	t_{Cal}	Epi Ave	Epi SD	t_{Cal}	Dopa Ave	Dopa SD	t_{Cal}	Epi Ave	Epi SD	t_{Cal}	Dopa Ave	Dopa SD	t_{Cal}	Epi Ave	Epi SD	t_{Cal}	Dopa Ave	Dopa SD	t_{Cal}	Epi Ave	Epi SD	t_{Cal}	Dopa Ave	Dopa SD	t_{Cal}
1	EMG Area	48.9	3.2	9.76	25.1	3.5	8.11	51.1	6.5	**1.17**	29.6	25.1	3.27	38.0	7.1	**1.29**	50.04	25.1	3.66	28.6	6.6	2.99	69.5	25.1	5.45	17.5	3.8	6.95	92.0	25.1	10.5
	EMG Peak Height	49.8	4.6	19.22	29.9	4.3	9.36	50.8	5.2	**1.33**	34.2	9.0	**1.61**	37.6	5.6	**1.45**	54.97	11.8	**2.48**	29.6	6.2	3.57	76.6	10.9	3.83	19.1	3.7	8.05	101	7	6.89
	GEMG4 Area	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
	GEMG4 Peak Height	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
	GEMG5 Area	63.2	5.0	**0.73**	0.00	0.00	6.08E+08	40.0	24.4	**0.81**	51.4	40.0	**0.69**	17.0	2.4	17.5	88.07	3.52	14.7	12.0	2.1	9.96	99.4	3.1	4.51	6.6	1.2	**0.57**	111	3	6.87
	GEMG5 Peak Height	62.0	5.6	**0.10**	0.00	0.00	1.25E+08	51.2	10.2	**0.76**	46.0	35.8	**0.40**	29.4	3.9	3.09	82.62	5.28	7.29	22.4	4.0	**1.11**	100.6	4.4	3.88	13.6	2.4	6.85	117	3	**2.42**
	EMG+GMG Area	59.4	4.1	**1.39**	6.55	3.47	4.82	54.5	2.0	8.19	23.7	3.1	13.1	38.0	6.0	**1.54**	50.10	9.29	4.43	26.1	5.0	2.70	73.9	8.2	5.88	14.5	2.3	8.09	97.3	4.5	12.6
	EMG+GMG Peak Height	59.6	5.1	22.17	8.60	4.42	2.65	53.7	1.6	8.58	27.7	3.1	9.84	37.3	4.8	**1.54**	54.57	9.40	3.21	27.2	4.8	3.35	81.2	8.2	3.74	16.1	2.3	9.77	106	5	7.59
	Cauchy-Gaussian Area	18.7	3.1	33.82	77.7	7.0	22.5	16.9	4.5	17.1	90.2	10.0	12.3	14.2	1.4	36.2	92.54	4.30	14.6	18.1	11.9	**0.50**	89.4	19.1	**0.55**	24.8	2.3	19.2	80.7	4.8	20.4
	Cauchy-Gaussian Peak Height	50.3	5.4	16.69	83.5	6.0	28.4	42.9	9.8	**1.29**	95.0	8.8	15.2	33.6	2.5	0.69	95.95	5.10	13.9	21.8	8.5	**0.34**	91.2	17.1	**0.35**	24.9	2.3	19.6	87.7	5.2	15.5
2	EMG Area	48.9	3.5	9.09	17.8	2.9	3.82	44.3	4.2	**2.15**	40.6	17.8	**0.36**	28.5	3.2	4.37	66.7	17.8	**0.14**	13.7	1.7	10.1	90.8	17.8	2.89	1.87	1.57	7.76	121	18	**0.11**
	EMG Peak Height	59.5	6.0	18.8	19.6	8.7	**1.73**	44.3	5.7	**1.61**	52.9	8.6	3.62	32.9	6.1	**0.56**	71.0	9.5	**1.07**	55.7	5.3	16.2	52.9	6.7	15.0	66.3	4.7	30.7	52.7	8.5	19.6
	GEMG4 Area	53.5	1.9	10.6	11.2	0.6	8.67	32.1	14.0	2.79	55.4	11.2	**2.48**	43.4	3.0	7.47	46.8	11.2	20.6	30.1	2.4	9.5	70.1	11.2	18.4	19.7	2.3	13.5	98.0	11.2	9.60
	GEMG4 Peak Height	14.6	0.6	5.3	75.0	2.2	67.7	14.3	2.8	29.6	89.3	3.8	31.6	15.9	0.8	54.2	88.6	1.4	38.3	8.5	0.6	47.6	111	2	19.2	4.3	0.3	25.0	127	3	4.64
	GEMG5 Area	53.2	4.0	5.26	12.2	2.9	**1.01**	51.8	2.5	3.67	31.1	2.0	11.1	17.5	6.7	6.09	80.6	9.2	3.63	8.3	2.6	11.6	97.4	4.6	**1.98**	1.28	1.42	9.64	121	5	**0.16**
	GEMG5 Peak Height	65.5	5.4	**1.71**	8.14	9.80	13.4	55.9	7.3	2.63	33.9	9.0	1.70	28.5	5.8	**2.45**	84.2	8.0	5.32	13.1	5.2	3.54	115	2	22.3	0.0	0.0	481	132	3	8.54
	EMG+GMG Area	50.1	2.7	10.6	16.2	2.2	3.24	47.2	3.0	**0.65**	36.9	2.2	3.61	31.6	3.5	**1.90**	62.7	3.6	2.84	16.2	1.6	6.69	87.5	2.8	5.36	3.73	1.31	5.84	118	5	**1.12**
	EMG+GMG Peak Height	53.5	10.9	9.00	28.5	15.3	**2.41**	38.4	7.9	2.98	61.4	8.6	6.09	31.7	5.0	**1.28**	72.7	8.1	**1.74**	56.8	5.5	16.0	51.4	7.1	14.6	75.0	3.6	46.0	40.8	7.3	26.9
	Cauchy-Gaussian Area	32.7	3.5	20.1	40.6	5.8	11.5	32.4	2.1	18.0	58.2	3.2	13.7	24.2	1.2	20.4	74.2	1.5	11.8	14.2	3.9	3.94	92.8	5.9	**0.37**	9.11	4.06	**1.4**	114	6	2.80
	Cauchy-Gaussian Peak Height	59.4	2.2	52.2	41.3	2.1	33.2	55.1	2.4	7.32	58.5	5.4	8.32	46.5	1.8	16.8	69.4	2.9	**2.18**	68.9	2.6	45.7	57.2	5.6	16.0	72.0	2.3	70.5	66.9	3.4	38.1
3	EMG Area	47.7	2.2	15.3	18.2	2.3	5.07	40.6	3.2	5.61	45.3	18.2	6.98	30.6	2.7	3.37	65.5	18.2	**0.86**	17.5	2.7	2.81	89.2	18.2	3.00	2.86	1.71	5.74	119	18	**1.69**
	EMG Peak Height	56.0	2.3	44.7	22.8	2.6	9.06	45.9	3.0	**1.73**	50.9	2.2	12.1	33.5	2.8	**0.69**	71.1	4.2	**2.47**	20.4	3.0	**0.13**	99.3	4.3	3.19	4.07	2.07	3.30	127	3	6.00
	GEMG4 Area	54.5	0.8	22.8	10.0	0.9	9.46	54.3	11.5	**1.34**	29.3	10.0	**2.14**	43.8	9.0	2.59	49.6	10.0	4.59	37.0	2.9	13.8	67.0	10.0	17.3	25.3	2.2	21.0	92.8	10.0	22.3
	GEMG4 Peak Height	54.4	0.4	247	19.1	1.4	10.3	51.4	3.9	**2.18**	39.9	8.8	**0.05**	39.4	2.9	4.27	61.0	5.9	**2.46**	29.1	2.7	7.88	87.5	4.0	3.83	17.7	1.7	15.8	112	3	7.87
	GEMG5 Area	35.0	24.9	2.63	29.8	23.7	**1.69**	26.3	12.6	4.22	61.2	13.6	3.79	29.0	5.6	**2.33**	66.8	7.0	**0.02**	13.4	8.0	**2.20**	92.6	7.1	**0.37**	6.55	0.94	**0.80**	114	2	7.24
	GEMG5 Peak Height	44.4	23.9	**1.78**	40.2	32.7	**1.62**	38.6	6.0	3.86	67.1	12.9	5.14	32.4	3.3	**1.41**	72.4	6.2	**2.19**	22.5	13.1	**0.36**	103	10	**2.11**	11.4	1.4	7.90	122	2	**2.09**
	EMG+GMG Area	49.9	0.6	45.7	16.3	1.7	4.27	45.9	2.2	**2.37**	39.8	1.2	**0.72**	32.2	2.1	**2.50**	63.7	1.5	5.17	18.8	2.3	**1.88**	87.7	3.2	4.49	3.26	0.39	22.9	118	2	2.64
	EMG+GMG Peak Height	57.7	0.7	166	20.6	1.8	9.90	50.5	1.6	3.90	44.9	1.5	7.59	34.8	2.1	**0.57**	69.1	1.7	3.15	22.0	2.6	**1.31**	97.8	3.9	**2.54**	4.51	0.47	12.2	127	2	7.86
	Cauchy-Gaussian Area	33.6	1.1	61.8	35.7	2.1	25.6	34.9	5.7	5.60	53.6	4.5	7.34	25.1	1.1	21.1	73.3	2.7	5.92	13.8	1.2	13.5	95.7	2.0	**2.48**	10.6	5.2	**1.73**	112	7	2.99
	Cauchy-Gaussian Peak Height	55.8	1.0	108	41.7	2.1	33.6	51.6	4.7	**1.86**	59.7	5.9	8.16	34.6	6.1	**0.13**	78.0	5.0	5.39	21.7	2.3	**1.16**	101	2	7.93	15.4	4.7	4.45	116	8	**1.46**
4	EMG Area	31.9	1.4	54.1	19.4	1.4	10.3	28.5	2.0	24.3	42.9	19.4	10.3	23.4	2.6	10.0	61.8	19.4	5.80	13.1	1.6	11.1	87.0	19.4	9.63	2.92	0.75	12.8	116	19	3.84
	EMG Peak Height	53.1	2.0	49.0	23.1	1.6	15.1	45.5	2.5	**2.50**	47.9	0.9	21.0	36.5	3.7	**1.49**	67.9	2.4	**1.03**	22.2	2.7	**1.48**	97.3	1.9	4.81	6.01	1.26	**1.64**	126	3	4.94
	GEMG4 Area	40.3	0.8	64.7	9.59	0.84	11.1	43.2	2.3	5.06	25.9	9.6	35.9	33.8	2.9	**0.45**	49.4	9.6	17.4	29.4	3.2	6.82	68.4	9.6	15.8	22.4	1.9	20.1	93.5	9.6	21.2
	GEMG4 Peak Height	55.6	1.3	80.8	17.1	1.2	7.39	53.5	2.9	4.61	36.9	1.4	5.75	39.6	3.4	3.78	61.2	2.1	6.60	31.4	3.8	7.02	87.7	3.1	4.64	21.5	2.0	17.9	113	3	6.12
	GEMG5 Area	28.4	1.9	43.3	23.5	1.7	14.3	14.5	3.6	22.9	58.4	5.7	7.80	7.36	1.27	51.9	79.5	1.7	18.3	5.63	1.10	33.3	94.7	2.2	**1.18**	2.41	1.41	7.72	116	3	3.94
	GEMG5 Peak Height	54.6	2.2	7.98	24.6	1.8	50.3	37.4	3.7	6.96	64.2	7.0	8.47	22.5	2.9	9.93	87.3	1.5	33.5	18.7	3.4	**1.32**	107	2	18.9	9.14	4.23	**1.32**	126	3	4.49
	EMG+GMG Area	34.8	1.4	46.7	16.1	1.6	4.13	31.6	2.6	15.2	39.4	1.5	**1.21**	21.8	2.3	13.2	63.7	2.0	3.99	14.1	2.0	7.99	85.8	2.3	8.30	2.99	0.58	16.4	116	3	3.94
	EMG+GMG Peak Height	57.5	2.1	52.2	19.5	1.8	8.40	49.9	3.3	**1.42**	44.2	1.7	5.79	33.5	3.5	**0.53**	69.6	2.0	3.34	23.8	3.3	**2.41**	96.0	2.4	**2.36**	6.02	0.98	**2.11**	126	3	5.05
	Cauchy-Gaussian Area	28.2	2.5	32.5	24.8	2.8	10.1	25.5	3.0	18.2	47.7	1.3	13.8	19.2	1.4	26.9	68.1	2.0	**1.47**	10.3	1.5	16.4	92.4	2.4	**1.32**	11.9	4.4	2.78	108	7	4.81
	Cauchy-Gaussian Peak Height	53.1	2.7	36.5	28.0	2.5	14.6	45.4	2.6	**2.43**	50.6	1.5	17.6	31.0	2.3	3.49	71.5	2.2	5.18	23.4	4.1	**1.66**	98.0	2.7	4.02	24.2	4.4	9.69	112	5	3.66

Key: NR = Not run for that Channel. Bolded values are those that pass the Student's t-test to 95% certainty. t_{Cal} values are the calculated t-values for the t-test, were n=6, t_{Table} = 2.57

The use of curve fitting methods fail to properly deconvolute the data as presented for all four data channels for the range of molar response ratios studied. When the curve fitting methods were able to properly deconvolute the data for an individual mixture set, it was rare that it was able to do so with both good accuracy and good precision. Within the channel itself, several curve fit equations were found to be useful in properly deconvoluting the data, however the equations were not consistent from one mix set to another, with one equation passing the t-test for one mixture, but failing for the next mixture.

An argument could be made that the mixtures on the far extremes, i.e., Mixes 1 and 5, both of which had low amounts of the minor species, are more difficult to properly deconvolute and thus could be expected to fail. If this was the case, then consistency in the remaining mixture sets would be expected. This was not found.

The methods failed because of the nature of the surface to be fitted. The function to be fitted to deconvolute the data is in actuality two identical functions added together, and then the parameters are allowed to change until a minimum in the sum-of-squares of the residual (SSR) is found. The minimum in the SSR is shallow, and hence a large number of possible combinations of the parameters for the equations will satisfy the SSR requirements, even with constricting the values to something reasonably or physically meaningful. Some of these combination of parameters, when put back into the equations, yield answers that are not in accord with what is actually in the mixture set. The found parameters yielded equations that when added together, resulted in a very low SSR value. Of course, one does not know if it is correct without knowing the actual amounts!

The use of curve fitting methods may work well when the peaks are separated by a distance sufficient to generate a valley between them, but once the peaks overlap to the extent that valley disappears there is nothing for the equation to fit but the peak as a whole. As the two peaks merge into one, the single merged peak takes on characteristics of a peak containing a single analyte, and the curve fitting methods cannot distinguish between the merged double peak and a single peak.

Chapter 4 — 3-Dimensional Chemometric Methods: Results and Discussion

The difficulties encountered using the data from the individual channel data in the previous chapter is not unexpected. The objective of performing the analysis with multiple electrochemical channels is to utilize the information provided by the additional channels. The data was thus examined using the common chemometric methods PCR, PLS, and SIMPLS.

One may initially consider all the data collected in the analysis to contain information concerning five separate, pertinent items in addition to various levels of noise in each. The five practical or "real" dimensions of initial concern include (1) time measured relative to an initial injection event for each individual injected chromatographic sample (the injected sample), (2) the applied electrochemical potential, which is different for each of the four electrodes spaced sequentially at the output of the chromatographic column, (3) the signal measured as the instantaneous current at each of the four electrodes at the specified time intervals following injection, (4) the concentration of Epi in the injected sample, and (5) the concentration of Dopa in the injected sample. Practically, one typically sees this data collection readily broken into the two obvious types of mathematical variables, i.e. independent and dependent. In the current situation, most would typically visualize both time and potential as being independent variables selected by the analyst. Subsequent measurements of signal at each of the electrodes for each of the specified times would contain the concentration information which we hope to glean concerning both Epi and Dopa. Of course, we would

have to employ standards of Epi and Dopa to determine their responses individually. We further typically assume that the contributions of these components to the signal are the same when encountered individually as it is in a mixture of the two components. Stated another way, we have collected signal data at specific times following injection of various samples on electrodes maintained at four different potentials in hopes that the result would be able to predict concentrations of both Epi an Dopa when appropriately unraveled. The times (J), the signals or current responses on each electrode (I) for each injected sample (each row of which are now termed "sample row", i.e., the row in the matrix), and the potentials (K) can be viewed as a three-dimensional matrix as indicated in Figure 4-1.

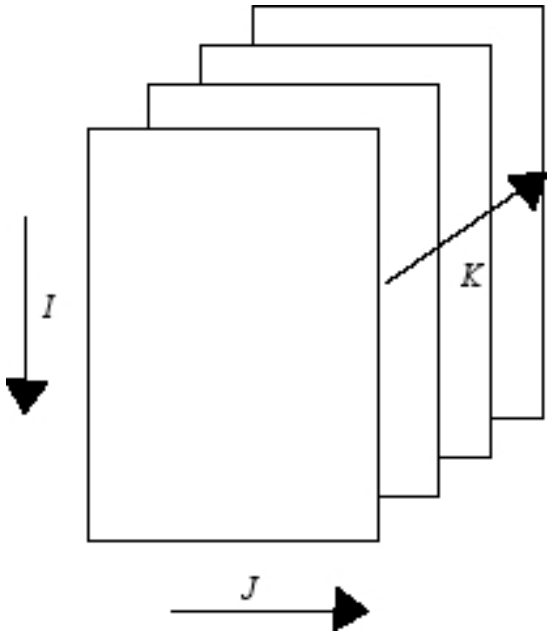


Figure 4-1: A three-dimensional matrix

The chemometic approach to this sort of problem is a bit more obtuse than the use of the simple "real" dimensional approach described above entails. In the chemometric approach, we must consider all the variables simultaneously and, essentially, let the data and associated outcomes from the entire data set create new dimensions or axes to more accurately define the data set(s) and subsequently be used to unravel information like concentrations of Epi and Dopa in novel injected samples. The new dimensions, known as principal components for the method, are obviously related to the original parameters, since they are derived from these; however, they no longer have direct and immediate practical correlations to the physical parameters originally submitted.

A primary step in any chemometric approach is to decide which of the initial pieces of information will be considered to be fixed (the independent variables) and which will be allowed to vary (the dependent variables). This decision will initially be submitted to the Principal Component Analysis (PCA) technique to determine the number of principal components in the new n-dimensional space. The identified principal components can subsequently be employed to predict the parameters which were originally allowed to vary; this allows for the determination of these independent variables in subsequent analysis of novel injected samples. In the current situation, we selected the time and signal to be fixed or independent variables in our studies. All signal vs. time data was submitted to the PCA fitting procedure for the entire data set. For each injected sample, this entailed four separate time vs. signal data sets, each with an associated applied potential for the electrode at which it was collected. We also submitted the Epi concentration and the Dopa concentration for each of these injected samples. The entire data set consisted of eighteen injections of an Epi standard, eighteen injections of a

Dopa standard, and six injections each of five different mixtures containing different amounts of both Epi and Dopa. This constitutes a total of 18 + 18 + 30 = 66 individual sample injections.

The submission of the entire data set to the PCA procedure yields information concerning the inherent dimensionality of the data set. This fitting procedure provides a scree plot and a table of cumulative captured variance, indicating the number of principal components necessary to encompass the fundamental underlying informational content of the data. The PCA also provides associated loadings plots and scores plots, both of which are used to confirm that the number of new principal components chosen is appropriate for inclusion of a majority of the information contained in the data sets. The subsequently selected number of principal components are then utilized by each of the PCR, PLS, and SIMPLS chemometric methods in attempts to fit the data. For each of these three chemometric fitting procedures, we initially provided the program with a training or basis set selected to comprise exactly half, or 33, of the entire number of injected samples. This basis set included the signal vs. time data for the four potentials for each of nine injections of an Epi standard, nine injections of a Dopa standard, and three injections each of the five different mixtures containing different amounts of both Epi and Dopa. The remaining half of the injected samples were subsequently employed by each of these three procedures as the "unknown" data to be analyzed by the fitting approach.

The number of principal components to employ is selected to include all the pertinent informational content while excluding extraneous information (noise). This selection is verified, in part, by the loadings and scores plots. Then the scores and loadings plots for the individual principal components are further analyzed to identify

potentially smaller regions within the original fundamental dimension (time) to examine. The loadings and scores plots identify such regions as containing the most substantial fraction(s) of the information content linking us to predictions of the independent variables including, most notably, the concentrations of Epi and Dopa.

Once the initially employed PCA yielded feedback information concerning the region of time which contained the majority of the essential information necessary to deconvolute the concentrations of Epi and Dopa, we narrowed our time window and repeated what is described above. Thus, with a smaller data set in time, we produced a scree plot, selected of the pertinent number of principal components, investigated fitting capabilities with the three chemometric procedures, and investigated of the associated loading and scores plots to possibly even further narrow our search. The original data set contained 481 times at which data was collected at the four electrochemical detectors. These 481 time points were reduced to 121 points (the reduced data set) and, finally, to a smaller 21 points (the very reduced data set). Each data set was successively treated by the same PCA, chemometric, and loading/scores analysis procedures. Along the way, examination of the chemometric fitting outcomes to see how well they could predict the test injected samples revealed, unfortunately, that our models were not providing as good a predictive capability for the Epi and Dopa concentrations as we had hoped. This iterative procedure and its associated results are described sequentially below.

After all attempts undertaken with the chemometrics procedures provided less than satisfactory predictive capabilities for the concentrations of Epi and Dopa in the test mixtures, we decided to examine the very simple Cramer's Rule possibility for deconvolution using the signal collected on each of the four electrochemical potentials at

single time points with careful selection of the time points to be used. This approach is described at the end of this chapter.

A. Principal Component Analysis and Associated Chemometric Analyses for the 481 Time Data Points

The entire data set for the 66 individual injections were used as the initial data set submitted for principal component analysis (PCA) to determine the inherent dimensionality of the data. Since there are 481 time measurements, each of which are related to four potentials, there are 481 dimensions in the data, each containing four measurements. The results of the PCA are shown below.

The first step in determining the dimensionality of the data is an examination of the scree plot, as shown in Figure 4-2. The term "scree" comes from the appearance of the plot itself, where the downward slope resembles the downward slope of a mountain and the bottom resembles the scree found at the bottom of such a mountain.

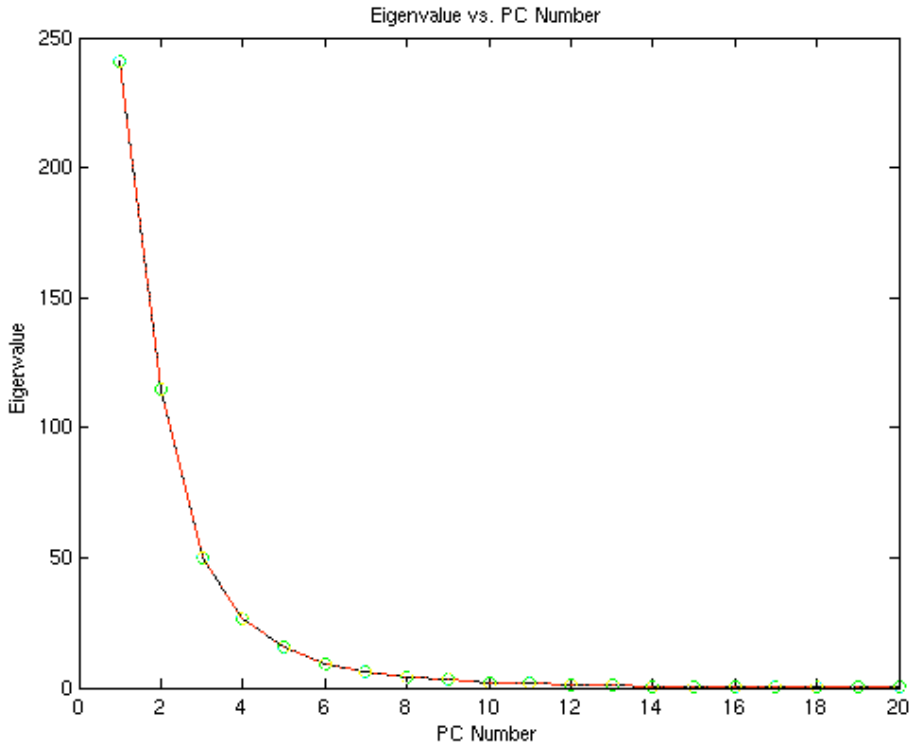


Figure 4-2: Scree plot for the original data set

The scree plot is, of course, subjectively interpreted to provide the number of principal components. In the present case, it would appear from Figure 4-2 that we have approximately eight principal components for our data. Since the plot may appear skewed to the eye, or may gradually decrease as seen in the figure above, an examination of the cumulative variance table can be of further assistance in assessing the number of principal components. The variance associated with Figure 4-2 is shown below in Table 4-1.

Table 4-1: Percent variance captured by PCA model for the original data set

Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC	% Variance Captured Total
1	1.02e+02	87.87	87.87
2	1.18e+01	10.15	98.03
3	1.88e+00	1.63	99.65
4	1.99e-01	0.17	99.82
5	1.12e-01	0.10	99.92
6	2.51e-02	0.02	99.94
7	1.77e-02	0.02	99.96
8	1.38e-02	0.01	99.97
9	5.49e-03	0.00	99.97
10	4.94e-03	0.00	99.98

The percent variance captured can be thought of as a measure of the amount of signal information found for each principal component (PC) after the contributions from the previous PCs have been eliminated. When the amount of variance found with the next PC is approximately equal to the amount of noise in the data set, one can be assured that the signal part of the data has been determined and can then use this information as a starting point to determine the number of PCs to retain for the model. According to the

table above, two PCs would capture 98% of the variance in the data set, even though there are three items under examination: the concentration of Epi, the concentration of Dopa, and the applied potential.

Although the scree plot and the percent variance table directs us concerning the number of principal components to retain, additional information can often be found by an examination of the scores and the loadings. The scores indicate correlations between the new PCs and the original axes of the injected sample. The loadings provide a measure of the perpendicularity of the original individual variable axis to the principal component axis, and is reported as the cosine of the angle between the two axes. If the loading value is zero for a variable along a PC axis, then the variable lies perpendicular to the PC axis and does not contribute to the variation described by that PC axis. Care must be taken when the PC axis does not capture much of the variance in the data; in this case, the loadings plots may have very little meaning. The loadings plots are correlated to the variables, in this case, time. Between the scores plots and the loadings plots, one can determine to a greater degree the dimensionality of the data set.

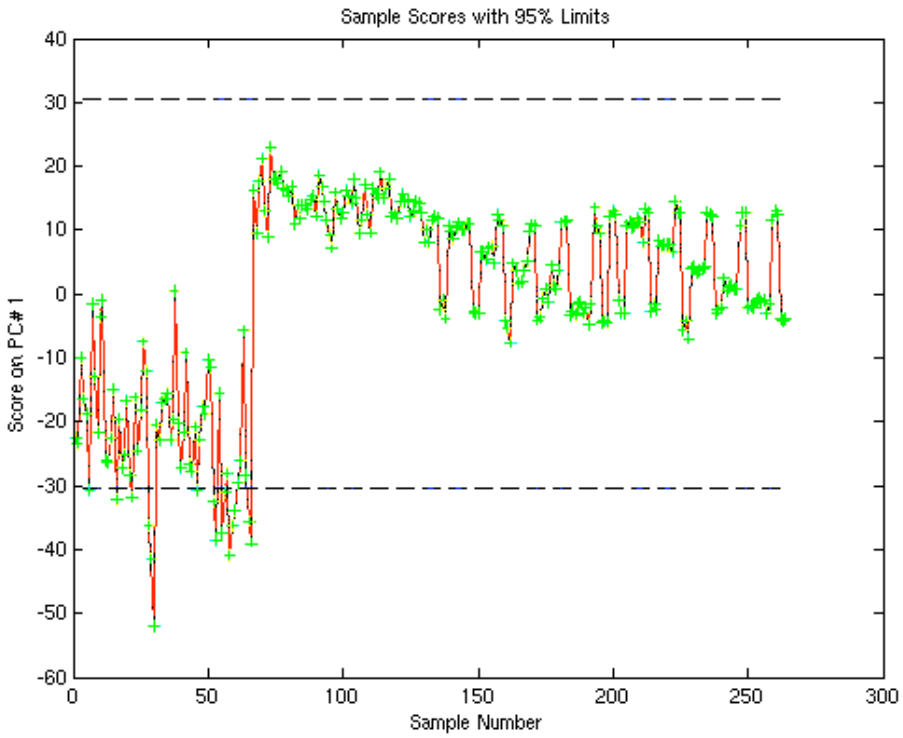


Figure 4-3: Scores plot for original data set, PC1

As seen in the scores plot, Figure 4-3 above, there appears to be some structure in the data after 66 injection samples, which roughly corresponds to the beginning of the data from the second channel. There were 66 injection samples in the data set. Each electrochemical channel data for each of the 66 injected samples is presented in sequence as one moves from sample row 1 to sample row 66 in the scores plot of Figure 4-3. The sample rows 67-132 are then composed of all the data for channel 2, sample rows 133-198 are composed of all the data from electrochemical channel 3, and, finally, the sample rows 199-264 are all the data from electrochemical channel 4. The dashed lines indicate incorporation of 95% of the values.

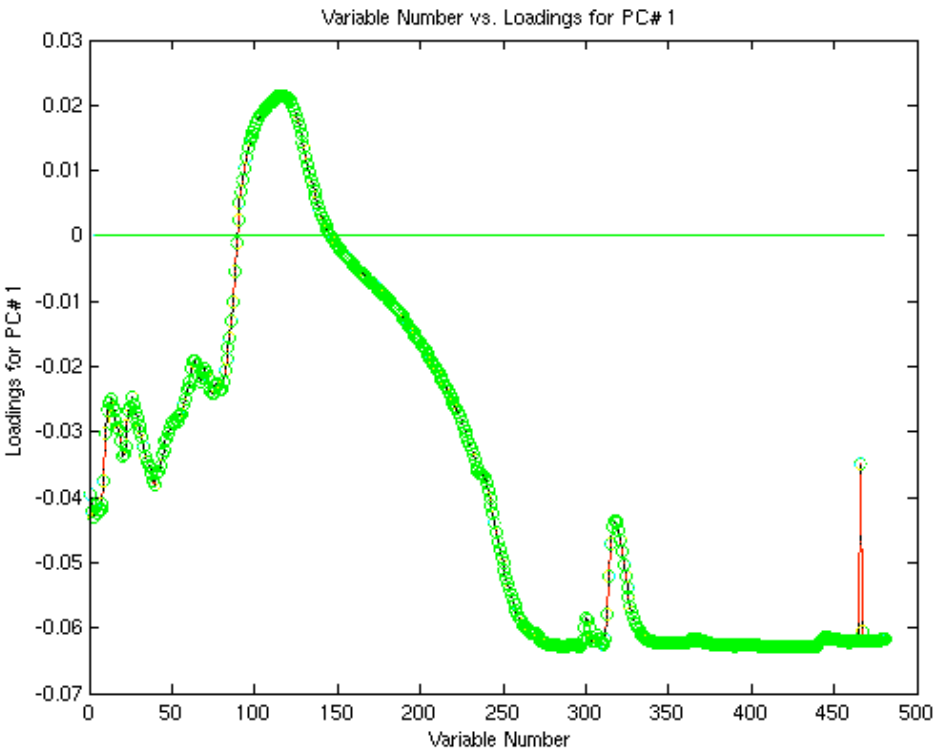


Figure 4-4: Loadings plot for original data set, PC1

The loadings plot for PC1 shows how the variables (time) contribute to PC1. The sign of the signal is irrelevant for this determination; if it is positive the axis lies in the same direction as the PC, and if it is negative it lies in the opposite direction. A loading value of exactly either 1 or -1 indicates that the variable axis lies exactly parallel to the principal axis, and any other value indicates that it lies off at an angle. Thus, the variables with the greatest magnitude from zero are those that contribute the most to the principal component. In this case, variable 100 to 125 contribute the most to PC1 in the positive direction, and variables 270 through 310 and 340 to 460 contribute in the opposite direction. Since most of this region lies where the two peaks are tailing, it is tempting to attribute this PC to the tailing of the two peaks.

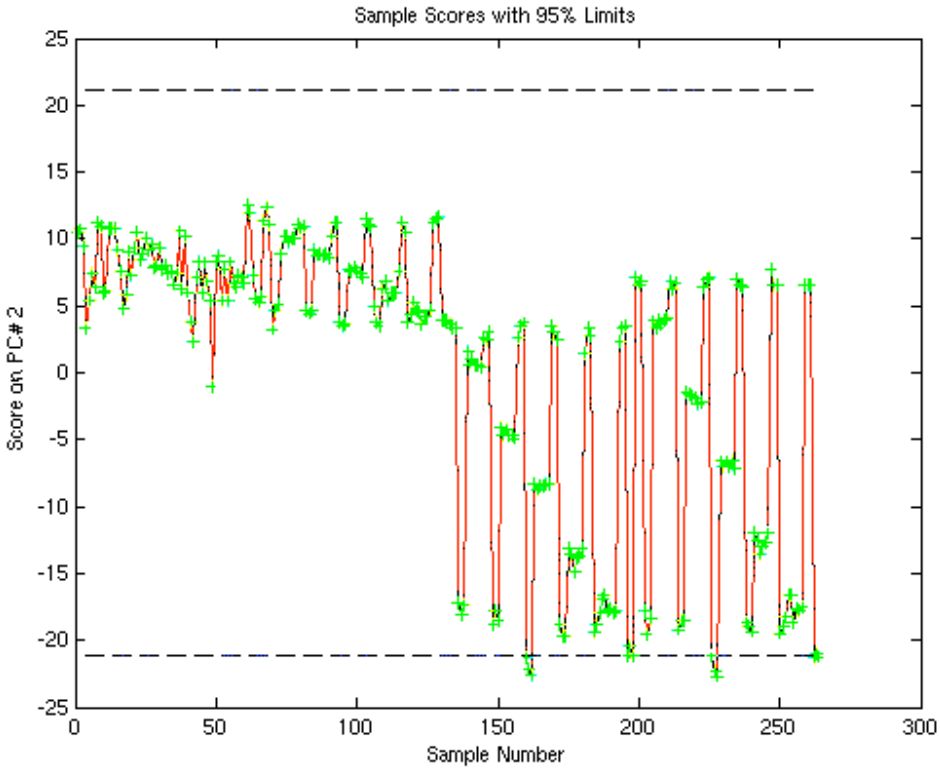


Figure 4-5: Scores plot for original data set, PC2

The scores plot for PC2 shows quite a bit of structure, with clearly four regions. As discussed previously, this must be attributed to the amount of the shift in the axes among the samples rows themselves.

For the loadings for PC2, two regions, one at variables 85 to 105 (1.541 to 1.574 minutes) and another region at 140 to 225 (1.633 to 1.774 minutes), contribute to the variance for this PC. It is tempting to assign specific chromatographic features to such ranges, but such assignments must be made with caution as erroneous assignments are easily made without careful consideration of the other PCs in the data set.

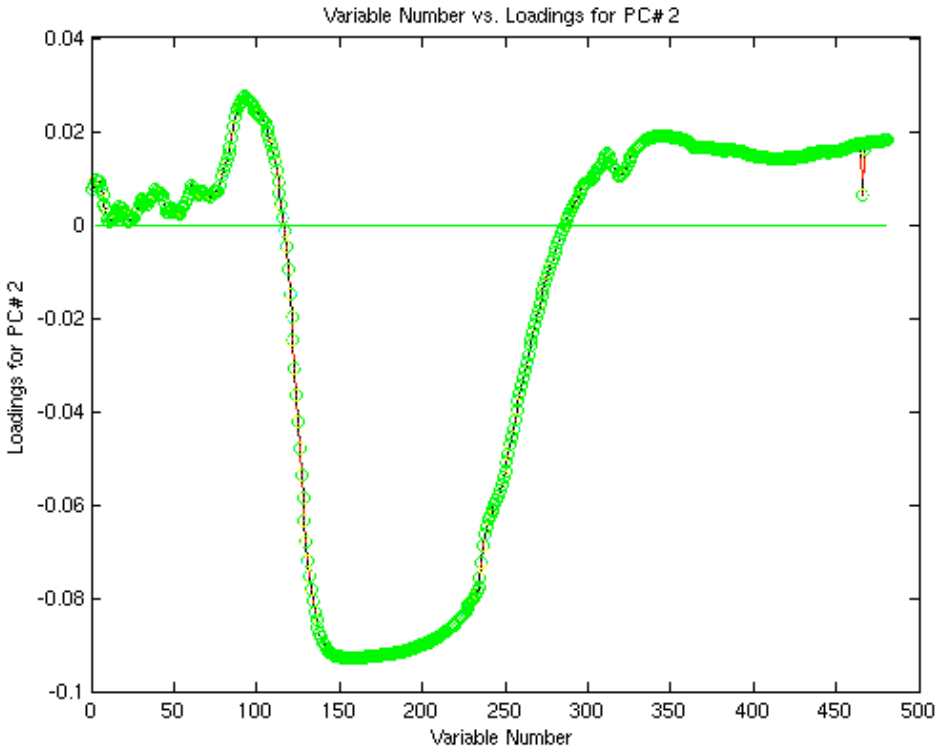


Figure 4-6: Loadings plot for original data set, PC2

The scores plot for PC3 is shown below in Figure 4-7. Structure is still present in the data, as seen by the regular "pulse" feature at about sample row 66. Below 67 sample rows, in fact, the data appears almost chaotic, and thus information concerning the injected samples using the first potential, 600 mV, have been described in previous PCs.

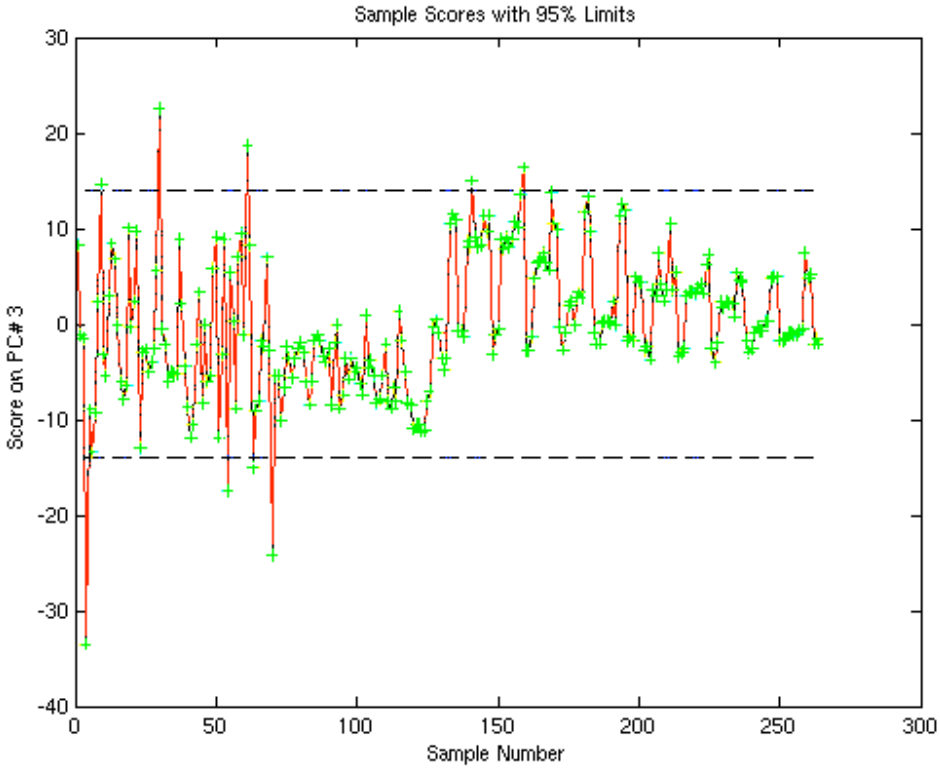


Figure 4-7: Scores plot for original data set, PC3

The loadings plot for PC3 is shown below in Figure 4-8. There is a broad feature from zero to about 125 variables (1.40 to 1.60 minutes), with the remainder of the variables (time points) contributing very little information. These additional times thus contribute noise to the PC3, and can be eliminated if the reduction of noise is desirable. The latter part of this feature appears to correspond with the arrival of Epi in the chromatogram; thus, a large part of this feature occurs before the peaks elute. However, the decision to eliminate data in this region must consider the other PCs involved in the data set.

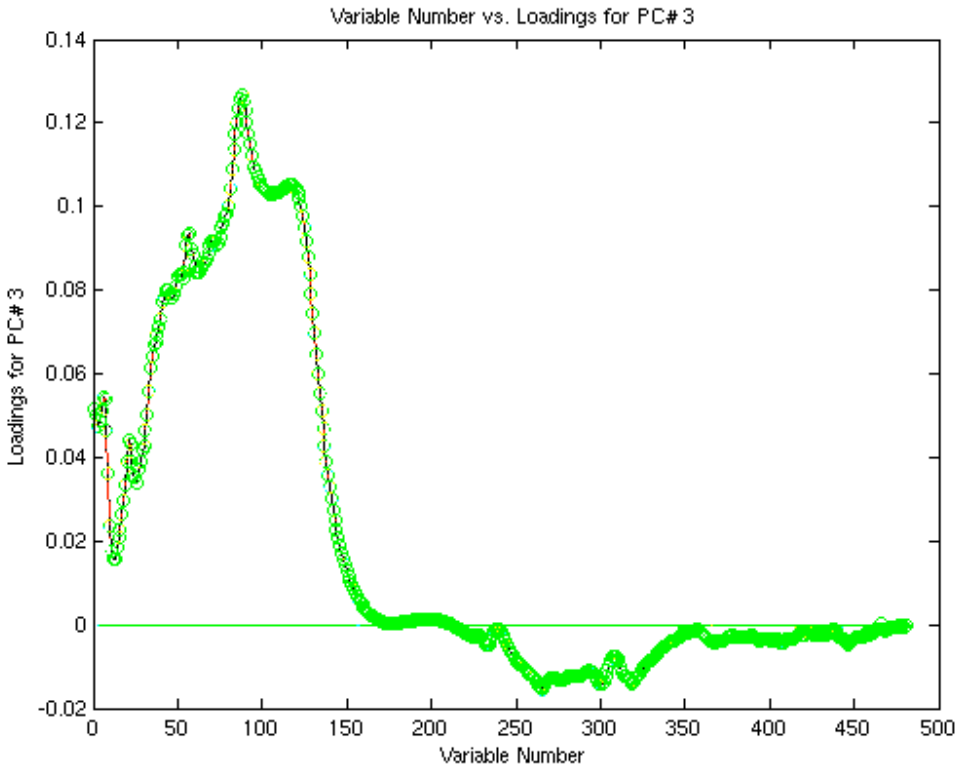


Figure 4-8: Loading plot for original data set, PC3

The scores plot for PC4 are shown below in Figure 4-9. The prominent features apparent in the earlier scores plots appears to be diminishing, and it appears much more chaotic and noisy. The informational content in the data obtained from channel 1 has been described, and channel 2 as well as channel 4 both appear to be nearly completely described. Nevertheless, there still appears to be some structure in the data.

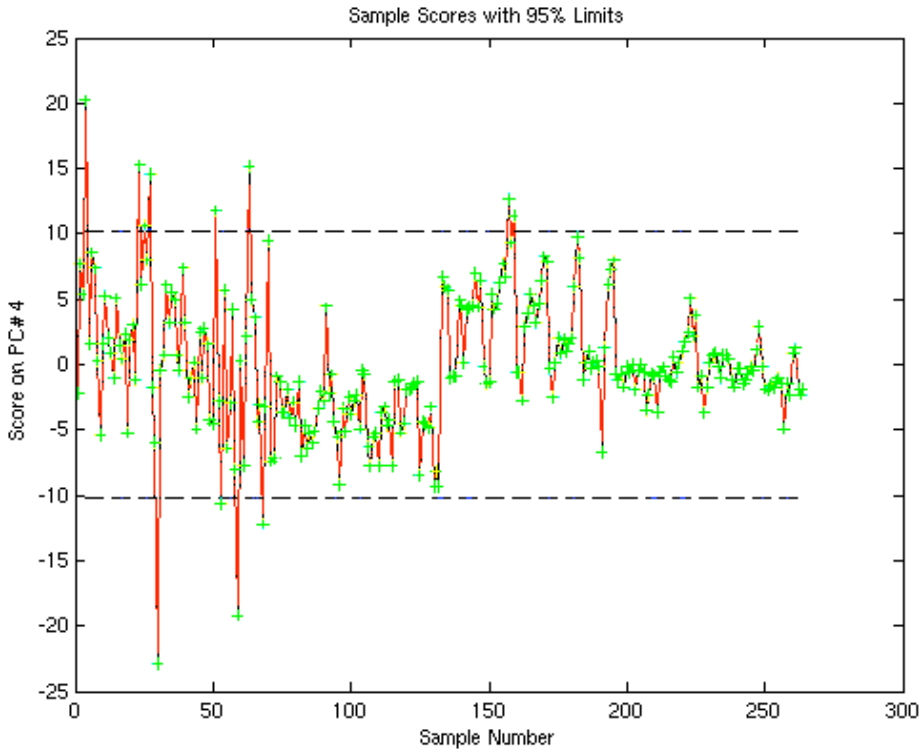


Figure 4-9: Scores plot for original data set, PC4

The loadings plot for PC4 shows both positive and negative regions that contribute most to PC4. These lie at 30-80 variable numbers (1.449 to 1.533 minutes) and 100-120 variable numbers (1.566 to 1.599 minutes). The rest of the variables appear to describe the small contaminant peak often seen on the far shoulder of Epi, but with the scale of this plot, it can be difficult to determine specifically what is causing the rise.

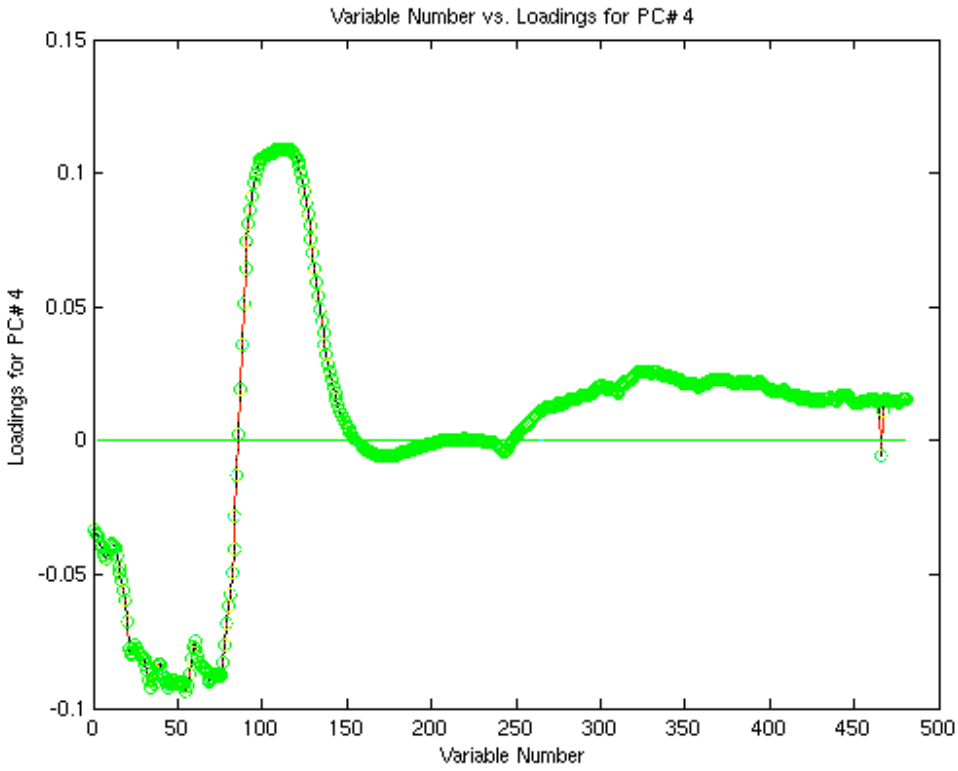


Figure 4-10: Loadings plot for original data set, PC4

Results

A total of 10 principal components were analyzed in the above manner. Subsequently, the dimensionality of the data set was determined to be approximately six, and the method subsequently was built using a total of 8 principal components. Half the data set was used to build the model (the basis set), and the remaining was used as the test set. The results of this entirely inclusive model were very poor, with little predictive ability either to the model or to the test set, as shown in the following Table 4-2. In this table, we have shown the representative data provided for one of the four electrochemical channels, in this case channel 2.

Table 4-2: Representative data for the entirely inclusive model, channel 2

Channel 2

$n = 6$, $t_{table} = 2.57$	Mix 1					Mix 2					Mix 3					Mix 4					Mix 5				
	Epi	Epi		Dopa	Dopa	Epi	Epi		Dopa	Dopa	Epi	Epi		Dopa	Dopa	Epi	Epi		Dopa	Dopa	Epi	Epi		Dopa	Dopa
Technique	Ave	Std Dev	t_{calEpi}	Ave	Std Dev	Ave	Std Dev	t_{calEpi}	Ave	Std Dev	Ave	Std Dev	t_{calEpi}	Ave	Std Dev	Ave	Std Dev	t_{calEpi}	Ave	Std Dev	Ave	Std Dev	t_{calEpi}	Ave	Std Dev
PCR Model																									
Basis – # PCs																									
5	27.6	1.3	61.95	3.17	2.91	24.6	1.0	54.63	15.0	1.9	31.83	15.5	2.0	22.54	25.3	2.0	50.67	8.04	1.84	16.70	37.3	1.0	132.50	0.40	1.30	12.14	49.8	2.2	77.42
6	27.4	1.4	60.10	2.55	3.29	24.5	1.1	53.29	14.7	1.9	33.32	15.5	2.0	22.63	25.2	2.0	50.69	7.92	1.97	15.76	37.0	1.1	131.83	0.36	1.24	12.81	49.7	2.4	71.67
7	28.6	1.4	59.14	2.24	3.53	24.5	0.9	65.02	14.7	1.8	34.14	15.2	2.1	22.53	25.2	2.0	50.56	7.99	1.51	20.47	37.0	1.1	125.72	0.32	1.27	12.65	49.7	2.4	71.46
8	30.7	1.6	47.76	-3.69	5.12	22.7	1.1	57.26	19.9	2.3	40.14	13.7	1.4	34.84	29.5	5.2	17.55	5.43	2.30	16.14	44.3	2.8	43.19	-2.54	1.37	16.76	57.9	2.7	57.51
Test Set – # PCs																									
5	27.4	1.2	73.12	5.90	0.43	24.1	1.6	36.97	17.0	0.8	72.21	15.2	0.9	50.90	26.9	0.4	224.73	8.46	1.46	20.34	35.9	5.2	26.96	2.56	0.48	22.13	47.5	1.6	110.50
6	27.4	1.1	75.58	5.76	0.54	34.80	1.7	35.22	16.8	1.0	55.53	15.2	1.0	46.97	26.8	0.7	142.14	8.42	1.52	19.5	35.8	5.6	25.53	2.54	0.40	26.76	47.5	1.7	104.06
7	27.7	1.3	64.90	5.66	0.57	33.24	1.3	44.42	16.8	1.1	50.69	14.8	0.5	91.82	26.9	0.8	119.13	7.94	1.49	20.84	35.9	5.7	24.68	2.93	0.97	9.96	47.4	1.8	97.78
8	29.8	1.7	45.26	-0.36	1.28	26.35	1.1	57.43	20.8	1.7	40.14	13.6	0.5	97.0	30.6	3.2	27.72	4.87	2.38	16.15	44.8	8.0	14.94	0.39	1.76	9.01	54.7	4.0	39.91
PLS Model																									
Basis set – # of PCs																									
5	27.5	1.3	62.46	6.86	0.93	17.15	24.7	1.1	53.09	21.2	1.9	24.75	15.7	2.0	22.41	28.5	4.5	20.73	8.20	1.78	17.03	48.7	3.0	37.23	0.67	1.23	12.28	61.8	2.6	55.42
6	27.6	1.3	62.40	1.20	3.17	9.41	24.6	1.1	54.29	27.0	1.1	29.39	15.6	2.0	22.64	40.5	5.3	12.27	8.03	1.83	16.77	59.0	6.0	14.23	0.42	1.29	12.26	77.9	6.5	16.15
7	30.4	2.3	33.81	7.36	7.47	**1.97**	25.5	1.2	47.52	28.9	1.2	22.36	16.7	1.7	25.34	43.0	4.3	13.57	9.91	1.05	24.82	63.2	7.5	9.99	2.64	1.02	10.19	82.8	7.1	12.97
8	33.9	3.1	22.22	8.26	7.71	**1.63**	24.5	1.7	33.81	28.7	1.3	21.00	16.6	1.1	40.22	43.0	4.0	14.77	9.23	1.46	19.06	63.0	7.4	10.19	2.35	1.40	7.91	82.7	7.3	12.62
Test Set – # PCs																									
5	27.3	1.1	75.99	8.53	1.72	6.92	24.2	1.7	35.00	22.4	0.5	91.23	15.4	0.9	49.44	30.1	2.1	42.01	8.58	1.29	22.82	49.7	5.9	18.32	2.69	0.45	22.61	58.8	2.5	61.10
6	27.4	1.1	73.43	2.69	4.16	6.30	24.1	1.6	36.49	29.0	4.4	6.17	15.2	0.9	51.23	41.9	3.6	17.00	8.43	1.43	20.81	58.9	14.9	5.72	2.55	0.46	22.82	67.9	1.4	94.82
7	29.0	1.2	68.75	6.21	4.37	4.02	25.2	1.5	37.41	31.4	4.3	4.98	16.3	0.4	102.72	44.2	2.4	23.47	9.58	1.23	21.87	61.4	17.4	4.54	3.87	1.23	5.96	70.8	1.6	76.98
8	32.0	1.3	55.08	6.99	4.31	3.63	25.1	1.8	30.40	31.4	4.4	4.90	16.6	1.0	45.38	44.3	2.1	27.01	8.06	2.08	14.72	61.0	17.6	4.55	3.50	3.18	2.59	70.7	2.1	57.31
SIMPLS Model																									
Basis set – # of PCs																									
5	27.6	1.4	61.28	7.18	0.88	17.31	24.8	1.1	52.99	21.0	1.9	25.15	15.8	2.1	22.08	28.3	4.4	21.49	8.36	1.76	16.97	48.4	2.9	38.46	0.83	1.24	11.94	61.4	2.5	57.00
6	27.7	1.4	61.32	1.52	3.25	8.94	24.7	1.1	54.13	27.0	1.1	29.07	15.6	2.1	22.30	40.6	5.2	12.45	8.20	1.81	16.73	59.1	6.0	14.19	0.59	1.29	11.93	77.9	6.5	16.04
7	32.3	3.1	23.53	7.98	7.18	**1.84**	25.7	1.3	42.29	28.4	1.4	21.19	17.1	1.4	29.67	42.7	4.0	15.00	10.6	0.8	31.25	62.5	7.2	10.59	3.46	1.07	7.82	82.0	7.3	12.98
8	33.9	2.7	24.97	7.21	7.32	**2.06**	24.5	1.6	35.40	28.9	1.2	22.54	16.5	1.1	39.95	43.0	4.4	13.45	9.19	1.32	21.11	63.2	7.5	10.01	1.79	2.16	5.74	82.7	7.1	13.05
Test Set – # PCs																									
5	27.4	1.1	77.46	8.79	1.65	6.82	24.3	1.7	35.26	22.3	0.4	98.19	15.4	1.0	48.33	30.0	2.0	45.60	8.77	1.27	22.8	49.3	5.8	18.89	2.86	0.46	21.17	58.5	2.3	65.40
6	27.4	1.1	74.90	2.93	4.16	6.15	24.2	1.6	36.70	29.0	4.4	6.12	15.3	0.9	49.99	41.9	3.5	17.44	8.63	1.41	20.8	58.9	15.0	5.70	2.72	0.47	21.39	68.0	1.3	97.41
7	30.2	1.2	64.77	6.86	4.29	3.72	25.7	1.5	36.22	31.2	4.4	4.98	16.8	0.5	88.75	44.1	1.9	28.69	9.89	2.05	12.78	60.7	17.3	4.68	4.51	1.95	2.94	70.5	2.2	54.85
8	31.8	1.3	58.29	6.11	4.37	4.08	25.1	1.8	31.75	31.4	4.3	4.96	16.5	1.2	37.01	44.2	2.4	23.08	8.14	2.23	13.67	61.5	17.3	4.55	3.83	**2.22**	70.8	1.5	82.34

Key: Bolded values are those that pass the Student's t-test to 95% certainty. t_{cal} values are the calculated t-values for the t-test, where n=6, $t_{table} = 2.57$

The amount of variance found in the results led to consideration of a reduction in the data set itself to exclude those times that contributed very little to the signal present in the data, and much to the resultant noise. This determination was made after careful consideration of the scores and loadings plots. A reasonable assumption was also made that more than six principal components need not be examined closely, as most of these appeared to contain more noise than signal. This data suggests that a reduced data set of 121 time points (121 dimensions) may be advantageous. Chromatographically, the time was subsequently set to range from 1.549 minutes to 1.749 minutes, a region which encompasses both the Epi and Dopa peak maxima.

B. Principal Component Analysis and Associated Chemometric Analyses for the Reduced, 121 Time Data Points

The reduced data set was then submitted for PCA analysis. The inherent dimensionality and/or number of principal components now appears to be no more than 4, as shown by the scree plot below in Figure 4-11:

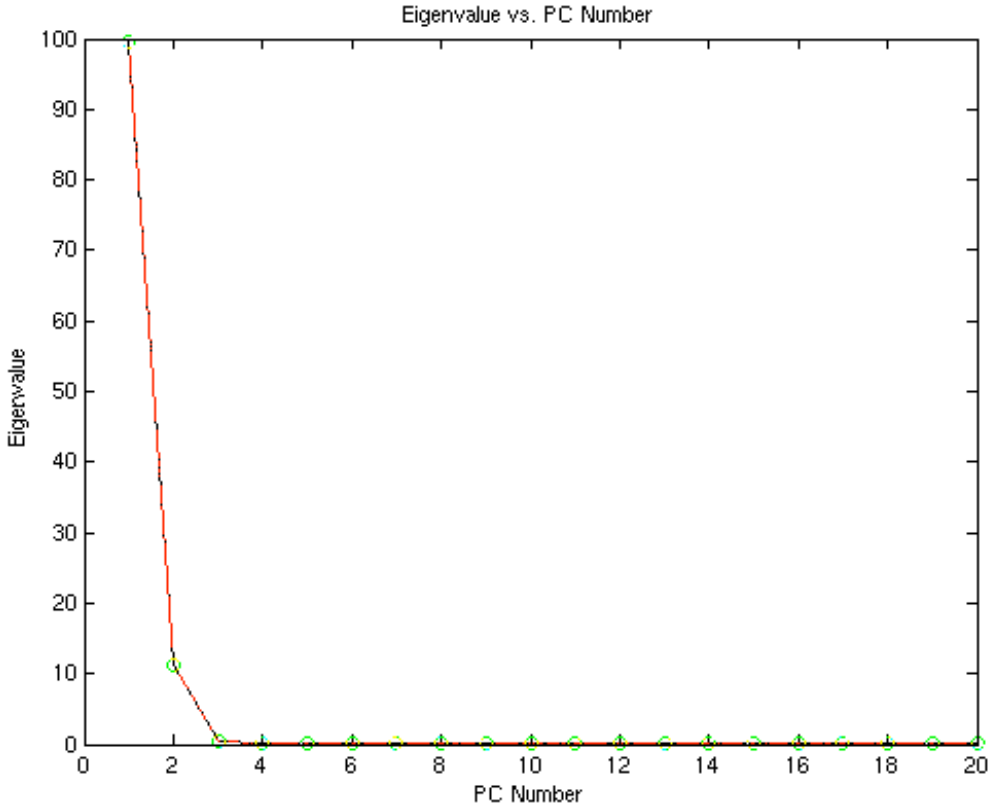


Figure 4-11: Scree plot for the reduced data set

Examination of the percent variance captured table, Table 4-3 below, confirms that with 4 PCs, well over 99.9% of the variance is captured in the model. As the chromatographic signal levels vary as much as 3%, the use of more than 4 PCs is not warranted. In fact, the percent variance captured table indicates that at the level of variance we wish to capture, two PCs would be adequate. Since there are three entities changing, concentration of Epi, concentration of Dopa, and the potential, three PCs would appear to be warranted. The principal component analysis was done using 4 PCs, as described below.

Table 4-3: Percent variance captured by the PCA Model, reduced (121 points) data set

Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC	% Variance Captured Total
1	9.96e+01	89.56	89.56
2	1.12e+01	10.11	99.67
3	3.09e-01	0.28	99.95
4	4.20e-02	0.04	99.99
5	5.81e-03	0.01	99.99
6	1.45e-03	0.00	99.99
7	7.79e-04	0.00	99.99
8	6.37e-04	0.00	100.00
9	5.04e-04	0.00	100.00

The scores plot for PC1 is shown below in Figure 4-12. The previous data set showed a scores plot for PC1 having a considerable amount of variation in the first 66 sample rows; this variation is greatly reduced, and the underlying structure of the data can be seen. The regular rise and fall of the data indicates that the two standards lie along different axes, and the relative amount of the two analytes dictate whether the mixture group will be high or low.

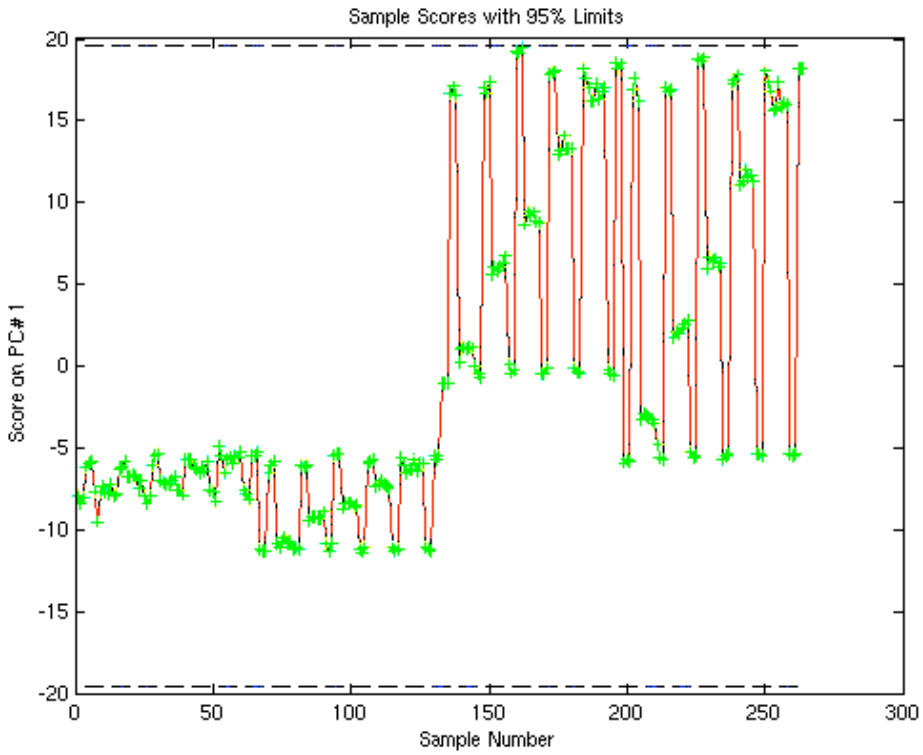


Figure 4-12: Scores plot for PC1, reduced data set

The loadings plot for PC1 shows a very ordered arrangement of the data points, and appears almost gaussian in character. Those new time numbers around 60 contribute the most to this PC.

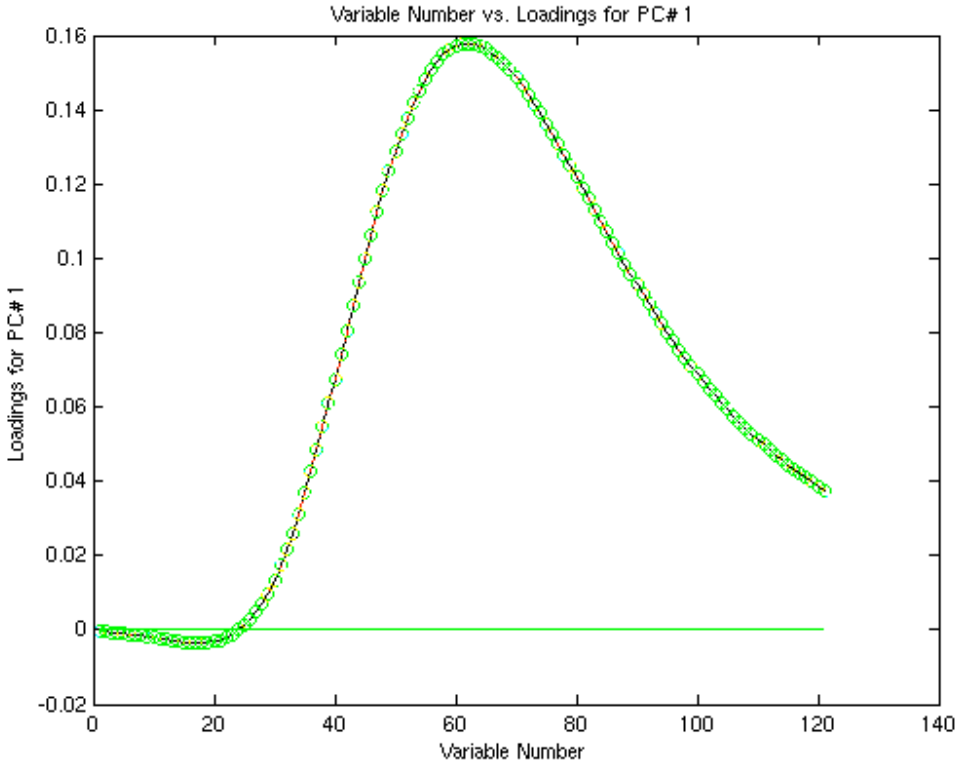


Figure 4-13: Loadings plot for PC1, reduced data set

The scores plot for PC2 also shows the regular patterns of rising and falling values as was found for the first PC, although the magnitude of these oscillations is smaller. Thus, most of the variance that can be attributed to the standards is described primarily by PC1, and secondarily by PC2. The loadings plot for this PC has the appearance of a first derivative of the gaussian function, with the most variance described by around variable numbers 30-35.

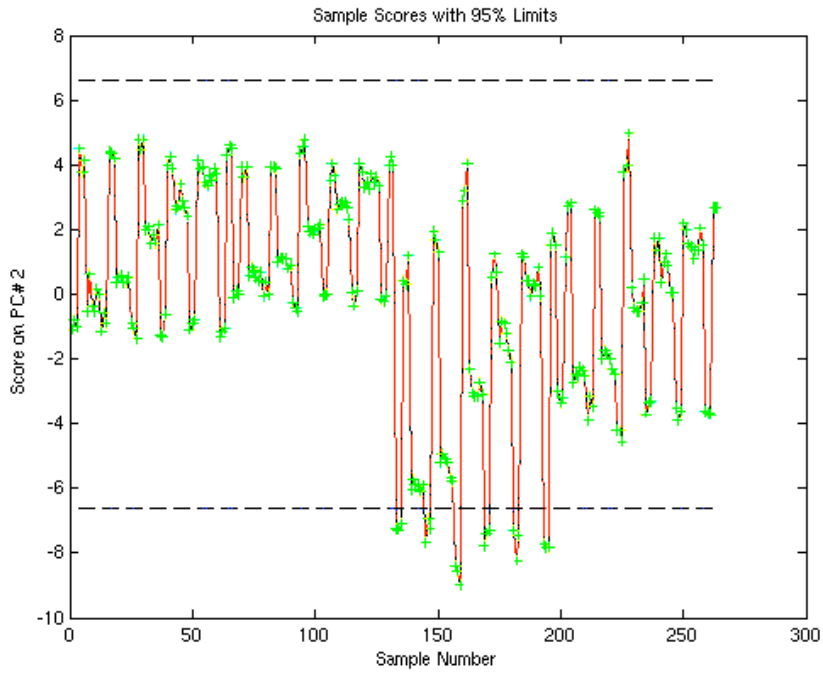


Figure 4-14: Scores plot for PC2, reduced data set

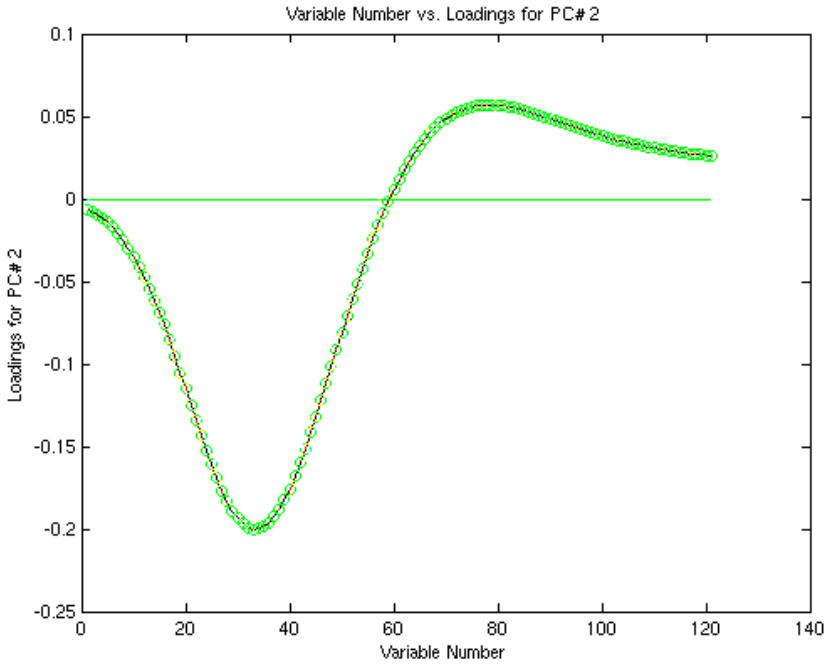


Figure 4-15: Loadings plot for PC2, reduced data set

The scores plot for PC3 shows a loss of structure, indicating more noise is being captured by this PC. However, there still appears to be some structure present; as the model uses three changing variables to determine, three principal components are required. The loadings plot for PC3 justifies this as well, as there still appears to be some structure present in the data set from sample rows 60 through 120 (channel 2 data), with less from 120 through 190 (channel 3 data).

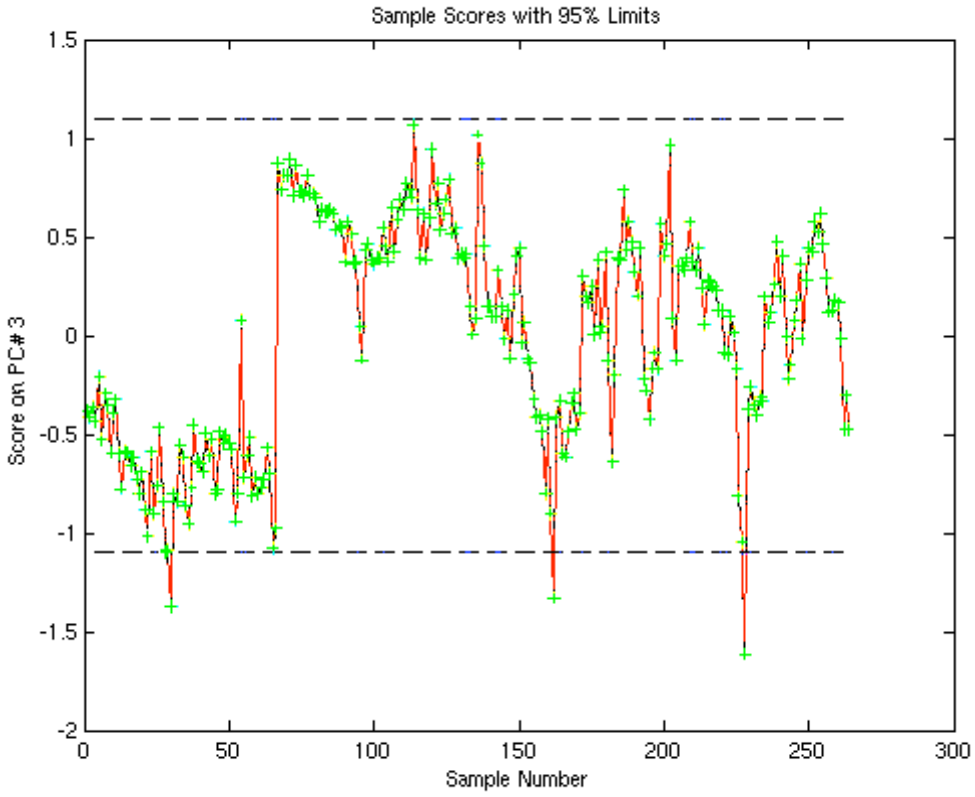


Figure 4-16: Scores plot for PC3, reduced data set

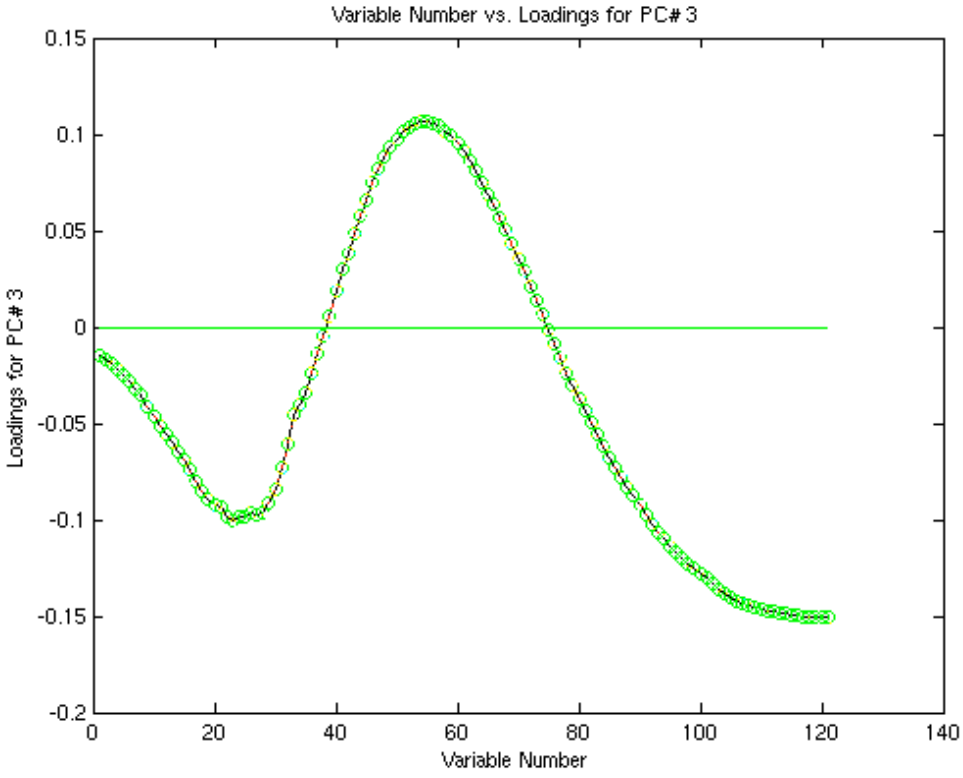


Figure 4-17: Loadings plot for PC3, reduced data set

The scores plot for PC4 show very little structure is left in the data set, as confirmed by the loadings plot for this PC. Now there are four regions that contribute the most to the PC, and it crosses the zero axis three times. This plot appears to be predominantly chaotic in nature and thus has little practical value.

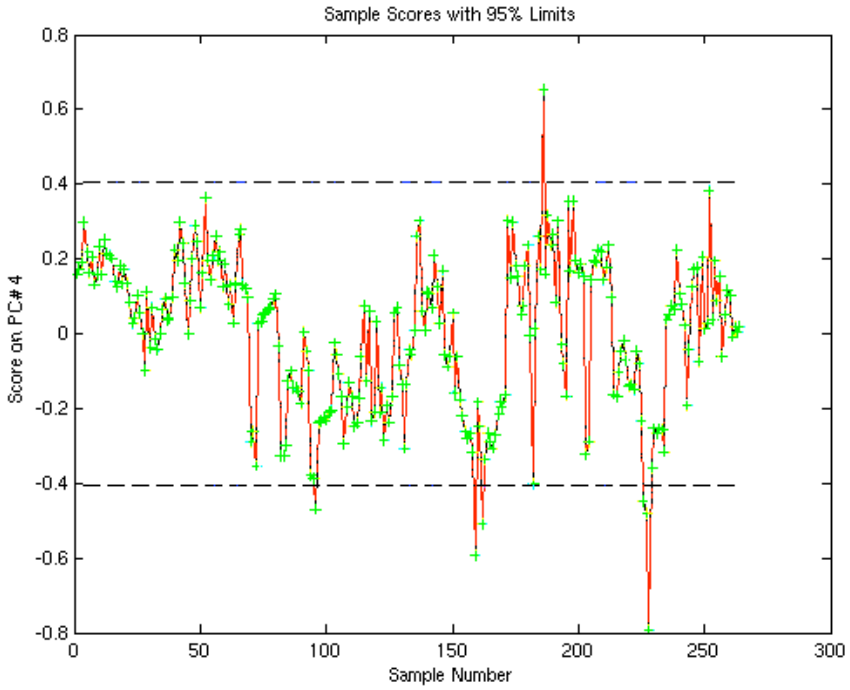


Figure 4-18: Scores plot for PC4, reduced data set

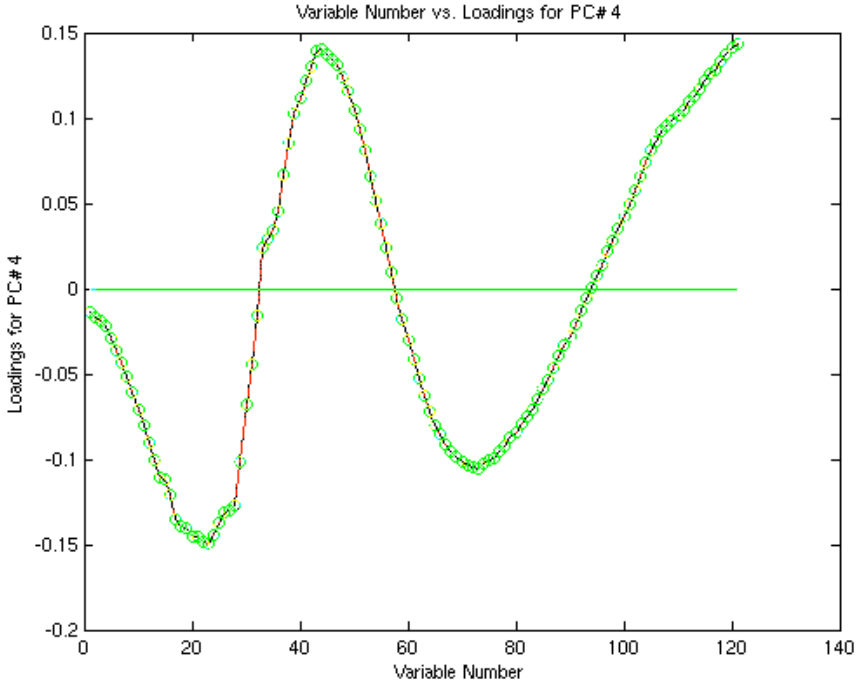


Figure 4-19: Loadings plot for PC4, reduced data set

Results

The PCA analysis for the reduced data set included 10 PCs. The potential PCs beyond four showed increasing amounts of noise and contained little, if any, additional information. The chemometric models for the reduced data set were, thus, built using 4 principal components. The predictive results of the derived chemometric approaches were again disappointing. First, the models did a very poor job predicting themselves, indicating they could not be relied upon to predict the test set. The results from the prediction of the test set confirmed that the models had poor predictive ability. The individual predicted outcomes for each of the four electrochemical channels is shown below in Tables 4-4 through 4-7.

Upon examination of the scores and loadings plots for the reduced data set, a further reduction in the time values employed appeared to be in order to enhance the signal to noise content. This further reduced data set consisted of some 21 time points, and thus was 21-dimensional. Chromatographically, the corresponding time ranges from 1.6044 minutes to 1.6394 minutes, which barely encompasses both peak maxima.

Table 4-4: Outcome for reduced data set, channel 1

Channel 1

$n = 6$, $t_{table} = 2.57$	Mix 1						Mix 2						Mix 3						Mix 4						Mix 5					
Technique	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$
PCR Model																														
Basis -- # PCs																														
2	35.8	3.1	20.6	15.4	3.0	1.62	29.3	0.7	63.96	25.3	0.9	41.65	18.9	2.1	17.91	33.8	2.4	33.98	10.0	2.0	12.93	46.3	1.1	101.30	1.29	1.80	7.57	56.1	1.9	84.53
3	42.4	3.5	**2.57**	33.0	4.8	8.91	40.2	2.0	**1.10**	54.6	3.0	10.02	29.3	2.0	6.16	61.7	5.9	**2.14**	19.7	1.3	**1.71**	72.4	4.8	10.91	11.8	2.1	5.69	84.2	3.8	23.24
4	46.7	3.7	21.4	34.3	4.5	10.33	41.4	1.3	**0.43**	54.9	2.5	12.54	29.2	2.3	5.43	61.7	6.3	**2.01**	24.1	4.1	**2.10**	73.8	4.2	11.55	16.0	2.8	8.06	85.6	3.4	24.79
Test Set -- # PCs																														
2	33.8	4.3	16.0	14.8	2.0	1.74	29.7	1.0	44.55	24.3	2.7	14.18	16.0	1.5	30.63	36.6	2.4	30.46	8.20	3.59	8.45	46.8	2.9	39.70	2.41	1.57	6.93	56.2	2.3	69.18
3	41.0	4.1	2.57	34.3	4.6	10.13	39.6	0.8	5.13	51.0	6.4	3.38	27.2	0.3	64.31	66.6	6.8	**0.13**	18.2	2.2	**2.67**	73.7	7.1	6.91	13.6	1.0	16.79	86.3	4.8	17.57
4	46.2	5.5	14.1	35.9	4.8	10.46	41.0	1.9	**0.14**	51.4	6.8	3.37	27.2	1.6	10.88	66.6	7.0	**0.12**	22.0	5.5	**0.62**	74.8	6.2	7.50	18.7	1.9	15.50	87.8	4.4	18.10
PLS Model																														
Basis set -- # of PCs																														
2	35.7	3.1	20.8	15.5	3.0	1.72	29.1	0.7	67.29	25.5	0.8	42.37	18.7	2.1	18.06	34.1	2.4	33.34	9.76	2.04	13.01	46.5	1.2	97.52	1.08	1.80	7.86	56.4	1.9	83.59
3	42.6	3.6	**2.57**	33.3	4.7	9.36	40.5	2.0	**0.81**	54.8	2.9	10.78	29.5	2.0	5.90	61.9	6.1	**2.04**	20.0	1.3	**0.98**	72.9	4.8	10.68	12.1	2.1	6.06	84.6	3.7	23.44
4	47.0	4.3	18.7	35.1	3.8	12.67	41.0	1.4	**0.30**	55.0	1.7	17.95	29.1	2.2	5.86	61.7	6.7	**1.91**	24.7	4.2	**2.38**	74.8	4.2	10.86	16.2	2.6	8.82	86.3	3.1	26.70
Test Set -- # PCs																														
2	33.6	4.3	16.1	14.9	2.0	1.90	29.4	1.0	44.07	24.5	2.8	13.81	15.8	1.5	30.35	36.9	2.5	29.68	7.99	3.62	8.52	47.0	2.9	38.85	2.19	1.59	7.20	56.5	2.3	68.15
3	41.3	4.2	2.6	34.7	4.6	10.29	39.8	0.8	4.03	51.2	6.5	3.41	27.5	0.3	50.93	66.8	6.8	**0.05**	18.6	2.2	**2.25**	74.1	7.0	6.84	14.0	1.0	17.60	86.7	4.7	17.53
4	46.9	6.3	12.7	37.0	4.9	10.82	40.8	2.0	**0.40**	51.6	7.0	3.32	27.6	1.8	9.03	66.8	7.1	**0.02**	23.1	5.0	**1.22**	76.0	5.9	7.32	18.7	1.9	15.06	88.6	4.2	18.51
SIMPLS Model																														
Basis set -- # of PCs																														
2	35.7	3.1	20.8	15.5	3.0	1.71	29.1	0.7	67.25	25.5	0.8	42.36	18.7	2.1	18.06	34.1	2.4	33.35	9.76	2.04	13.01	46.5	1.2	97.56	1.08	1.80	7.86	56.4	1.9	83.60
3	42.6	3.6	**2.57**	33.3	4.7	9.36	40.5	2.0	**0.81**	54.8	2.9	10.78	29.5	2.0	5.90	61.9	6.1	**2.04**	20.0	1.3	**0.98**	72.9	4.8	10.68	12.1	2.1	6.06	84.6	3.7	23.44
4	47.0	4.3	18.7	35.1	3.8	12.65	41.0	1.4	**0.29**	55.0	1.8	17.88	29.1	2.2	5.86	61.7	6.7	**1.91**	24.7	4.2	**2.37**	74.8	4.2	10.87	16.2	2.6	8.82	86.3	3.1	26.68
Test Set -- # PCs																														
2	33.6	4.3	16.1	14.9	2.0	1.90	29.4	1.0	44.07	24.5	2.8	13.81	15.8	1.5	30.36	36.9	2.5	29.69	8.00	3.62	8.51	47.0	2.9	38.86	2.19	1.59	7.20	56.5	2.3	68.16
3	41.3	4.2	2.6	34.7	4.6	10.29	39.8	0.8	4.03	51.2	6.5	3.41	27.5	0.3	50.92	66.8	6.8	**0.05**	18.6	2.2	**2.25**	74.1	7.0	6.84	14.0	1.0	17.60	86.7	4.7	17.53
4	46.9	6.3	12.7	37.0	4.9	10.82	40.8	2.0	**0.40**	51.6	7.0	3.32	27.6	1.8	9.03	66.8	7.1	**0.02**	23.1	5.0	**1.22**	76.0	5.9	7.32	18.7	1.9	15.06	88.6	4.2	18.51

Key: Bolded values are those that pass the Student's t-test to 95% certainty. t_{Cal} values are the calculated t-values for the t-test, where n=6, t_{Table} = 2.57

Table 4-5: Outcome for reduced data set, channel 2

Channel 2

$n = 6$, $t_{Table} = 2.57$	Mix 1						Mix 2						Mix 3						Mix 4						Mix 5					
Technique	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$
PCR Model																														
Basis – # PCs																														
2	29.9	1.4	55.00	6.75	1.56	10.41	25.1	1.2	47.48	18.8	0.4	137.50	16.7	1.2	36.65	28.9	0.4	232.39	10.2	2.3	11.08	40.6	2.6	49.65	1.42	1.32	10.08	54.6	2.8	58.39
3	26.6	1.1	**2.57**	-2.09	3.04	14.07	24.1	2.2	18.98	16.2	2.5	25.46	17.0	1.3	33.30	29.7	0.5	189.95	7.02	0.80	41.60	31.9	8.1	18.68	0.0421	0.9966	16.76	50.9	4.4	39.04
4	30.9	1.5	27.32	-0.76	3.00	13.16	22.7	0.9	50.39	15.7	2.1	31.11	13.4	1.3	38.55	28.6	0.6	161.52	6.22	2.94	11.97	31.7	7.4	20.55	-3.47	1.90	13.34	49.8	4.0	43.13
Test Set – # PCs																														
2	29.0	1.2	64.87	7.83	0.23	60.28	25.9	1.3	41.82	17.1	1.4	41.50	16.2	0.9	50.35	30.2	0.4	252.87	9.52	1.31	20.62	41.8	2.1	60.48	3.60	0.84	9.51	51.7	1.3	130.61
3	26.5	1.2	**2.57**	1.13	0.14	244.53	24.4	1.4	29.59	13.0	2.2	32.98	16.5	0.7	58.89	31.0	1.6	54.33	7.13	1.71	19.31	35.4	1.0	137.45	1.30	0.63	21.46	45.5	2.6	70.10
4	30.7	1.5	27.28	2.44	0.29	107.79	23.1	1.1	41.08	12.6	2.1	34.33	13.0	0.3	192.56	29.9	1.7	53.94	5.41	3.07	12.11	34.8	0.6	227.19	-0.13	1.06	16.16	45.1	2.4	76.85
PLS Model																														
Basis set – # of PCs																														
2	30.0	1.4	54.56	6.63	1.57	10.50	25.1	1.2	48.42	18.8	0.4	144.12	16.7	1.2	36.75	28.9	0.4	232.16	10.3	2.3	10.80	40.5	2.7	48.70	1.46	1.33	9.94	54.6	2.8	58.05
3	26.7	1.1	**2.57**	-1.81	3.02	13.94	24.0	2.2	19.03	16.1	2.5	25.95	17.0	1.3	33.66	29.6	0.5	178.29	6.95	0.80	41.80	31.9	8.0	18.81	-0.0424	1.0118	16.71	50.8	4.3	39.43
4	32.0	1.5	29.25	0.40	2.86	12.83	23.3	1.1	39.82	15.8	2.0	32.44	13.7	1.1	45.96	28.2	0.7	134.27	6.64	2.47	13.80	31.7	7.2	21.00	-2.96	1.88	12.78	49.6	3.9	44.49
Test Set – # PCs																														
2	29.1	1.2	64.75	7.73	0.22	62.05	25.9	1.3	41.80	17.0	1.4	41.45	16.2	0.9	50.11	30.2	0.4	245.64	9.57	1.31	20.64	41.7	2.1	60.71	3.65	0.86	9.19	51.7	1.3	130.87
3	26.6	1.2	2.57	1.40	0.17	203.74	24.3	1.4	29.69	13.0	2.2	33.09	16.5	0.7	60.54	30.9	1.6	55.75	7.07	1.74	19.00	35.3	1.0	147.22	1.23	0.67	20.68	45.5	2.6	70.41
4	31.6	1.5	28.69	3.49	0.37	77.73	23.4	1.1	37.81	12.6	2.1	34.69	13.5	0.2	313.75	29.7	1.8	51.84	6.03	3.00	11.88	34.9	0.5	295.27	0.13	1.35	12.25	45.0	2.4	77.73
SIMPLS Model																														
Basis set – # of PCs																														
2	30.0	1.4	54.56	6.63	1.57	10.50	25.1	1.2	48.41	18.8	0.4	144.06	16.7	1.2	36.75	28.9	0.4	232.18	10.3	2.3	10.81	40.5	2.7	48.71	1.46	1.33	9.94	54.6	2.8	58.05
3	26.7	1.1	**2.57**	-1.81	3.02	13.94	24.0	2.2	19.03	16.1	2.5	25.95	17.0	1.3	33.66	29.6	0.5	178.28	6.95	0.80	41.80	31.9	8.0	18.81	-0.0425	1.0118	16.71	50.8	4.3	39.43
4	32.0	1.5	29.24	0.39	2.86	12.83	23.3	1.1	39.85	15.8	2.0	32.42	13.7	1.1	45.88	28.2	0.7	134.51	6.63	2.48	13.77	31.7	7.2	20.99	-2.97	1.88	12.78	49.6	3.9	44.47
Test Set – # PCs																														
2	29.1	1.2	64.75	7.73	0.22	62.03	25.9	1.3	41.80	17.0	1.4	41.45	16.2	0.9	50.11	30.2	0.4	245.71	9.57	1.31	20.64	41.7	2.1	60.70	3.65	0.86	9.20	51.7	1.3	130.87
3	26.6	1.2	**2.57**	1.40	0.17	203.74	24.3	1.4	29.69	13.0	2.2	33.09	16.5	0.7	60.54	30.9	1.6	53.75	7.07	1.74	19.00	35.3	1.0	147.23	1.23	0.67	20.68	45.5	2.6	70.41
4	31.6	1.5	28.68	3.48	0.37	77.98	23.4	1.1	37.83	12.6	2.1	34.69	13.5	0.2	313.30	29.7	1.8	51.86	6.02	3.00	11.88	34.9	0.5	294.29	0.13	1.34	12.28	45.0	2.4	77.72

Key: Bolded values are those that pass the Student's t-test to 95% certainty. t_{Cal} values are the calculated t-values for the t-test, where n=6, $t_{Table} = 2.57$

Table 4-6: Outcome for reduced data set, channel 3

Channel 3

n = 6, t_{Table} = 2.57	Mix 1						Mix 2						Mix 3						Mix 4						Mix 5					
Technique	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$
PCR Model																														
Basis -- # PCs																														
2	79.9	1.0	42.58	12.9	3.0	**0.41**	71.2	3.5	16.14	45.2	1.6	7.65	47.9	3.4	9.83	78.4	1.8	15.57	33.7	5.3	6.05	111	5	8.58	15.31	2.41	8.59	143	4	15.97
3	76.2	1.0	2.571	2.94	4.24	7.19	69.7	5.3	13.23	41.3	3.2	**0.64**	50.2	3.9	9.93	84.5	0.5	80.16	30.4	3.0	8.03	102	11	**1.75**	10.24	1.06	7.81	130	7	3.55
4	78.8	2.1	73.55	3.75	4.18	6.81	63.3	2.1	25.48	39.3	2.2	3.15	40.4	4.1	3.67	81.5	0.4	94.12	34.3	4.6	7.25	103	11	**2.12**	14.32	2.93	6.24	131	6	4.34
Test Set -- # PCs																														
2	78.6	0.9	44.10	15.0	0.8	5.17	71.6	2.6	22.58	43.1	1.3	5.64	50.9	0.7	57.67	77.0	2.2	11.18	33.6	4.0	8.03	112	6	7.06	19.55	1.38	22.59	138	2	22.13
3	75.5	1.0	2.571	6.70	0.46	46.56	70.2	3.5	20.15	39.3	2.5	2.78	52.3	1.9	23.80	80.8	4.5	7.51	29.5	5.3	4.11	101	4	4.54	13.72	0.47	35.71	123	5	**1.23**
4	77.8	0.9	169.65	7.41	0.57	34.23	63.4	1.7	32.10	37.2	2.1	5.75	43.5	1.4	15.79	78.1	4.0	6.76	33.0	6.9	4.44	102	3	5.92	21.33	1.09	32.51	125	5	2.59
PLS Model																														
Basis set -- # of PCs																														
2	80.0	1.1	42.09	12.7	3.0	**0.56**	71.2	3.5	16.33	45.1	1.6	7.70	47.8	3.4	9.82	78.4	1.8	15.68	33.8	5.4	6.03	111	5	8.43	15.44	2.44	8.61	143	4	15.80
3	76.2	0.9	2.571	3.06	4.21	7.17	69.5	5.2	13.31	40.8	3.0	**1.06**	50.0	3.9	9.84	83.9	0.6	75.58	30.4	3.0	8.18	102	11	**1.80**	10.18	1.05	7.74	130	6	3.60
4	79.2	3.0	53.49	4.28	4.25	6.40	62.3	1.7	30.61	37.8	1.4	7.42	39.1	3.4	3.45	79.4	0.7	45.92	34.9	4.2	8.28	104	11	**2.31**	13.85	2.59	6.61	131	6	4.56
Test Set -- # PCs																														
2	78.6	0.9	44.30	14.8	0.8	4.65	71.6	2.5	22.78	43.0	1.3	5.49	50.9	0.7	58.99	77.0	2.2	11.15	33.7	4.0	8.15	112	6	7.04	19.70	1.40	22.45	138	2	21.78
3	75.5	1.0	2.571	6.81	0.48	44.04	70.0	3.5	20.15	38.8	2.5	3.34	52.1	1.9	23.58	80.2	4.4	7.35	29.5	5.4	4.06	101	4	4.77	13.73	0.47	35.66	123	5	**1.37**
4	78.2	1.0	151.08	7.91	0.78	23.38	62.2	1.4	37.26	35.5	2.0	8.10	42.0	1.3	13.89	76.0	3.7	6.06	33.6	7.0	4.58	103	3	6.94	20.90	0.98	35.04	126	4	3.08
SIMPLS Model																														
Basis set -- # of PCs																														
2	80.0	1.1	42.10	12.7	3.0	**0.56**	71.2	3.5	16.33	45.1	1.6	7.70	47.8	3.4	9.82	78.4	1.8	15.68	33.8	5.4	6.03	111	5	8.43	15.44	2.44	8.61	143	4	15.81
3	76.2	0.9	2.571	3.06	4.21	7.17	69.5	5.2	13.31	40.8	3.0	**1.06**	50.0	3.9	9.83	83.9	0.6	75.58	30.4	3.0	8.18	102	11	**1.80**	10.18	1.05	7.74	130	6	3.60
4	79.2	3.0	53.55	4.27	4.25	6.40	62.3	1.7	30.59	37.9	1.4	7.37	39.1	3.4	3.44	79.4	0.7	46.22	34.9	4.2	8.27	104	11	**2.31**	13.86	2.59	6.61	131	6	4.56
Test Set -- # PCs																														
2	78.6	0.9	44.30	14.8	0.8	4.65	71.6	2.5	22.78	43.0	1.3	5.50	50.9	0.7	58.97	77.0	2.2	11.15	33.7	4.0	8.15	112	6	7.04	19.70	1.40	22.45	138	2	21.78
3	75.5	1.0	2.571	6.81	0.48	44.03	70.0	3.5	20.15	38.8	2.5	3.34	52.1	1.9	23.57	80.2	4.4	7.35	29.5	5.4	4.06	101	4	4.77	13.73	0.47	35.67	123	5	**1.37**
4	78.2	1.0	151.21	7.91	0.78	23.47	62.2	1.4	37.24	35.6	2.0	8.07	42.0	1.3	13.91	76.0	3.7	6.07	33.6	7.0	4.58	103	3	6.93	20.91	0.98	35.03	126	4	3.07

Key: Bolded values are those that pass the Student's t-test to 95% certainty. t_{Cal} values are the calculated t-values for the t-test, where n=6, t_{Table} = 2.57

Table 4-7: Outcome for reduced data set, channel 4

Channel 4

$n = 6, t_{table} = 2.57$	Mix 1						Mix 2						Mix 3						Mix 4						Mix 5					
Technique	Epi Ave	Epi Std Dev	$t_{cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{cal\ Dopa}$
PCR Model																														
Basis -- # PCs																														
2	52.5	1.5	15.0	20.1	1.5	10.72	45.6	3.2	**1.86**	51.4	1.4	20.36	27.7	3.1	5.27	84.7	2.1	21.24	17.4	4.3	**1.83**	118	4	14.55	6.19	2.75	**0.60**	150	4	18.06
3	49.4	2.4	2.571	12.0	2.9	2.88	44.4	4.4	**1.82**	48.4	1.9	8.26	30.6	2.8	3.32	92.3	3.0	20.58	16.7	2.9	3.22	116	8	6.60	1.73	0.60	20.78	138	8	5.21
4	55.6	1.2	84.8	13.9	2.7	**1.32**	41.9	2.5	**0.69**	47.6	1.3	10.64	22.2	3.9	7.60	89.7	2.6	21.30	18.0	5.6	**1.12**	116	7	7.58	2.01	2.69	4.41	138	8	5.55
Test Set -- # PCs																														
2	51.6	1.3	19.9	20.9	2.0	9.40	45.5	2.0	3.21	50.3	1.2	20.99	27.3	4.5	3.80	86.6	3.2	15.09	18.0	5.5	**1.17**	118	7	8.76	8.98	1.58	3.29	145	2	29.84
3	49.1	1.3	2.571	14.2	2.0	**1.49**	44.3	3.1	**2.53**	47.2	4.4	2.84	29.9	4.3	**2.47**	93.8	4.2	15.51	16.4	3.9	2.63	114	14	3.53	4.72	0.72	7.24	134	8	3.98
4	55.3	0.5	186.7	16.1	2.3	**0.75**	42.3	1.6	**1.78**	46.6	3.9	2.78	22.1	4.8	6.14	91.3	4.0	14.80	16.7	8.0	**1.20**	114	13	3.92	7.91	1.52	**1.69**	135	8	4.62
PLS Model																														
Basis set -- # of PCs																														
2	52.5	1.5	15.1	20.0	1.5	10.57	45.6	3.2	**1.87**	51.4	1.3	20.65	27.7	3.1	5.31	84.8	2.1	21.19	17.4	4.3	**1.80**	118	4	14.40	6.31	2.79	**0.48**	150	4	17.92
3	49.6	2.4	2.57	12.4	2.9	**2.56**	44.4	4.4	**1.80**	48.3	1.7	8.88	30.4	2.8	3.43	91.9	3.0	20.39	16.8	3.0	3.17	116	8	6.72	1.57	0.58	22.40	138	8	5.15
4	56.9	0.9	112.1	15.4	2.7	**0.05**	42.0	1.9	**1.07**	47.3	0.8	16.24	22.0	3.5	8.61	88.4	2.6	19.97	18.5	5.6	**0.90**	117	7	8.14	0.67	2.22	6.83	138	8	5.38
Test Set -- # PCs																														
2	51.6	1.2	19.9	20.8	2.0	9.24	45.5	1.9	3.23	50.2	1.2	20.52	27.2	4.5	3.83	86.7	3.2	15.08	18.0	5.6	**1.14**	118	7	8.67	9.09	1.63	3.36	145	2	28.91
3	49.2	1.3	2.571	14.6	2.1	**0.93**	44.3	3.1	**2.50**	47.2	4.3	2.83	29.8	4.3	**2.55**	93.4	4.3	15.12	16.4	3.9	2.60	114	14	3.57	4.63	0.76	7.17	134	8	3.96
4	56.7	0.6	171.4	17.7	2.6	**2.23**	42.6	1.4	2.65	46.5	3.7	2.90	22.2	4.0	7.36	90.2	4.3	13.29	17.2	7.7	**1.07**	114	12	4.10	6.66	0.98	**0.50**	134	8	4.55
SIMPLS Model																														
Basis set -- # of PCs																														
2	52.5	1.5	15.1	20.0	1.5	10.57	45.6	3.2	**1.87**	51.4	1.3	20.65	27.7	3.1	5.31	84.8	2.1	21.19	17.4	4.3	**1.80**	118	4	14.40	6.31	2.79	**0.48**	150	4	17.92
3	49.6	2.4	2.571	12.4	2.9	**2.56**	44.4	4.4	**1.80**	48.3	1.7	8.88	30.4	2.8	3.43	91.9	3.0	20.39	16.8	3.0	3.17	116	8	6.72	1.57	0.58	22.40	138	8	5.15
4	56.9	0.9	111.9	15.4	2.7	**0.04**	42.0	1.9	**1.06**	47.3	0.8	16.18	22.0	3.5	8.61	88.4	2.6	19.99	18.5	5.6	**0.91**	117	7	8.13	0.68	2.22	6.80	138	8	5.38
Test Set -- # PCs																														
2	51.6	1.2	19.9	20.8	2.0	9.24	45.5	1.9	3.23	50.2	1.2	20.52	27.2	4.5	3.82	86.7	3.2	15.08	18.0	5.6	**1.14**	118	7	8.67	9.09	1.63	3.36	145	2	28.92
3	49.2	1.3	2.571	14.6	2.1	**0.93**	44.3	3.1	**2.50**	47.2	4.3	2.83	29.8	4.3	**2.55**	93.4	4.3	15.12	16.4	3.9	2.60	114	14	3.57	4.63	0.76	7.17	134	8	3.96
4	56.7	0.6	171.5	17.7	2.6	**2.22**	42.6	1.4	2.64	46.5	3.7	2.90	22.2	4.0	7.35	90.2	4.3	13.31	17.2	7.7	**1.08**	114	12	4.10	6.67	0.99	**0.47**	134	8	4.55

Key: Bolded values are those that pass the Student's t-test to 95% certainty. t_{cal} values are the calculated t-values for the t-test, where n=6, $t_{Table} = 2.57$

C. Principal Component Analysis and Associated Chemometric Analyses for Very Reduced, 21 Time Data Points

The very reduced data set was then submitted to PCA analysis. The inherent dimensionality appeared to be no more than 3, as shown by the scree plot:

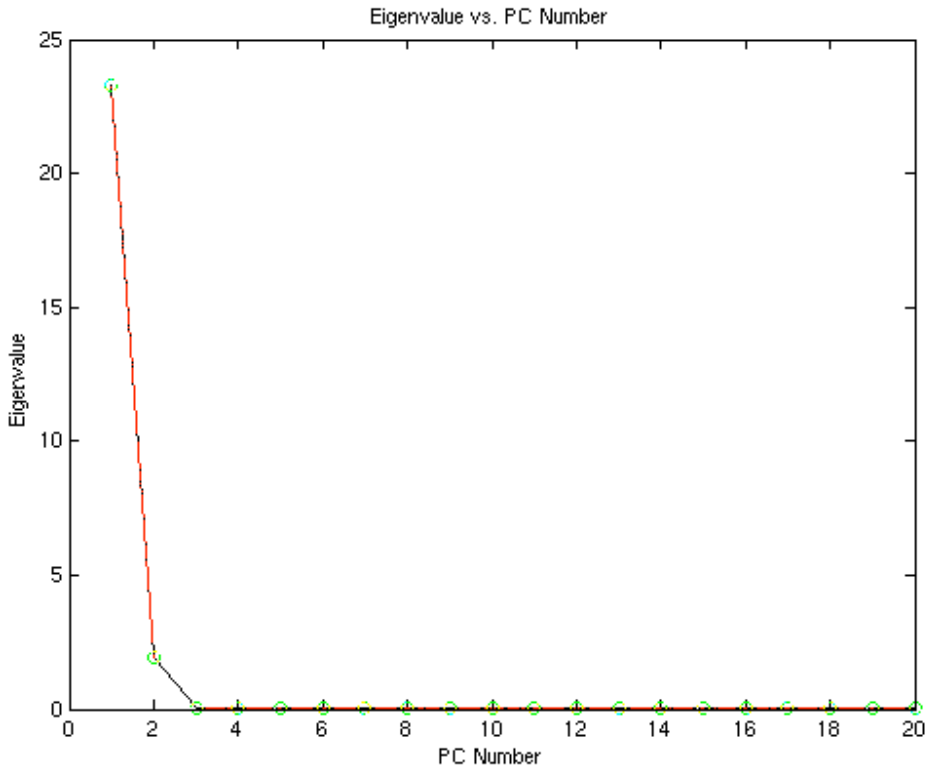


Figure 4-20: Scree plot for the very reduced data set

This need to use no more than three PCs for the very reduced data set was confirmed by examination of the percent variance captured, shown in Table 4-8 below. This table indicates that three PCs encompassed virtually all the variance found in the data set, with diminishingly small amounts of variance found using more than this

number. Obviously, the model incorporates some variance associated with noise that could be eliminated by the use of fewer PCs; however, as there are three parameters that are changing in this model, concentration of Epi, concentration of Dopa, and the potential of the electrodes, the use of less than three is not warranted.

Table 4-8: Percent variance captured by the PCA Model, very reduced (21 points) data set

Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC	% Variance Captured Total
1	2.32e+01	92.33	92.33
2	1.92e+00	7.66	99.99
3	1.70e-03	0.01	100.00
4	3.45e-04	0.00	100.00

The scores plot for PC1 is shown in Figure 4-21. The regular rise and fall of the data seen in the previous models is largely gone. The loadings plot, Figure 4-22, shows a very organized arrangement for the data points, with the most contribution for PC1 coming from the later time points. These later time points have more Dopa present than Epi, and thus this PC might reasonably be coupled to the presence of Dopa in the injected sample at that time.

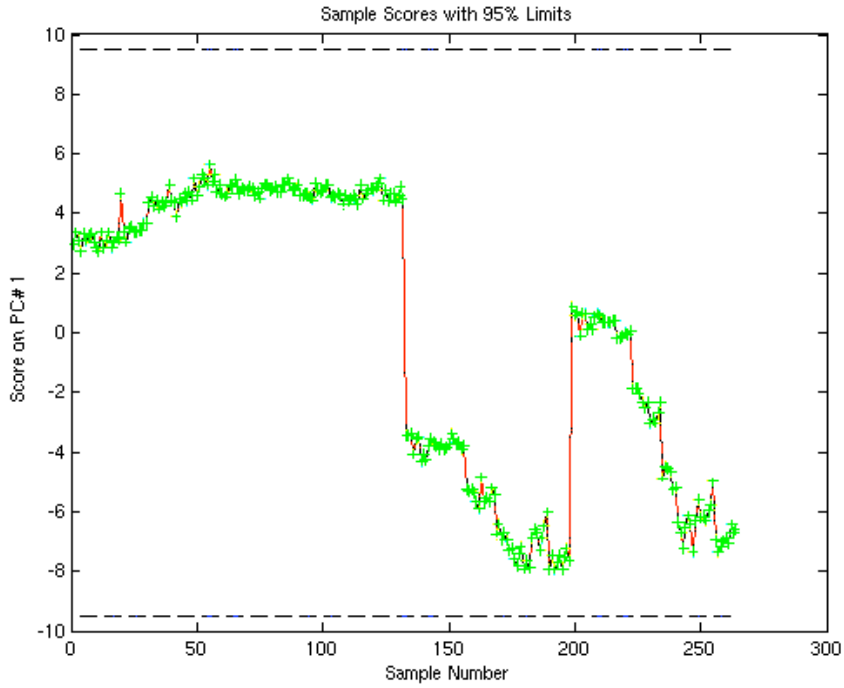


Figure 4-21: Scores plot for PC1, reduced data set

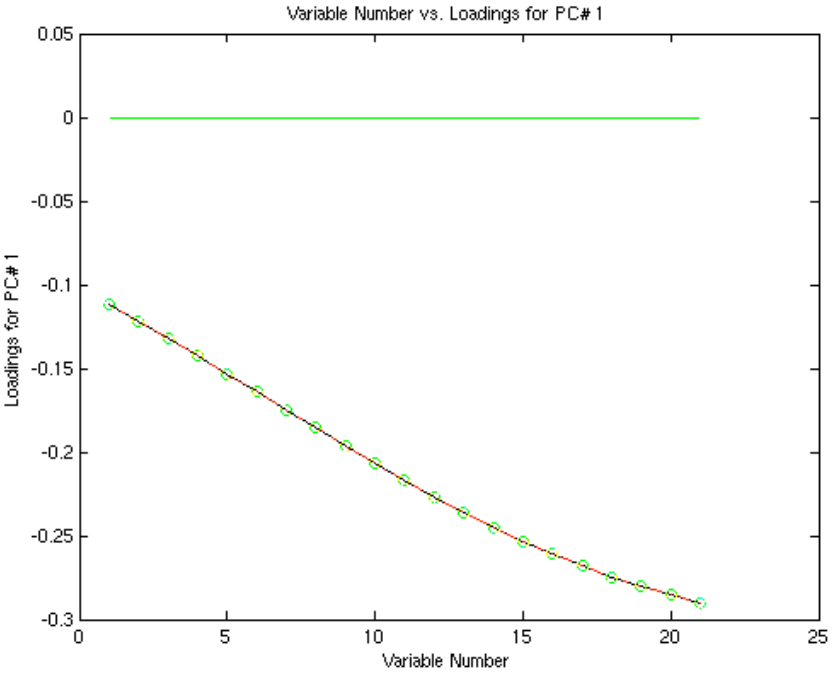


Figure 4-22: Loadings plot for PC1, reduced data set

The scores plot for PC2 also shows the regular patterns of rising and falling values as was previously seen. The upper range of each "rise" can be attributed to those injected samples with larger amounts of Dopa in them for each potential; the ordered arrangement of four such patterns indicates that this perhaps is related to the potential of the samples.

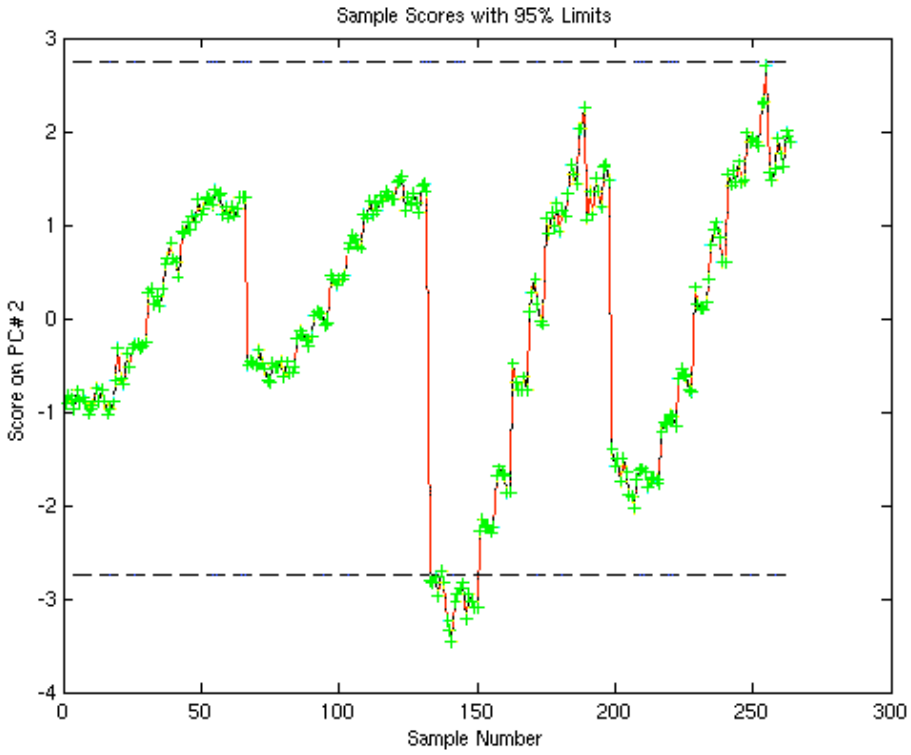


Figure 4-23: Scores plot for PC2, reduced data set

The loadings plot for this data set shows a slightly sigmoidal shape with an inflection point around variable number 11, which corresponds to the retention time of Epi, 1.621 minutes. Thus, the portion of the curve below this point is related to an

increasing signal for Epi, while the portion of the curve above this point is related to the declining signal for Epi.

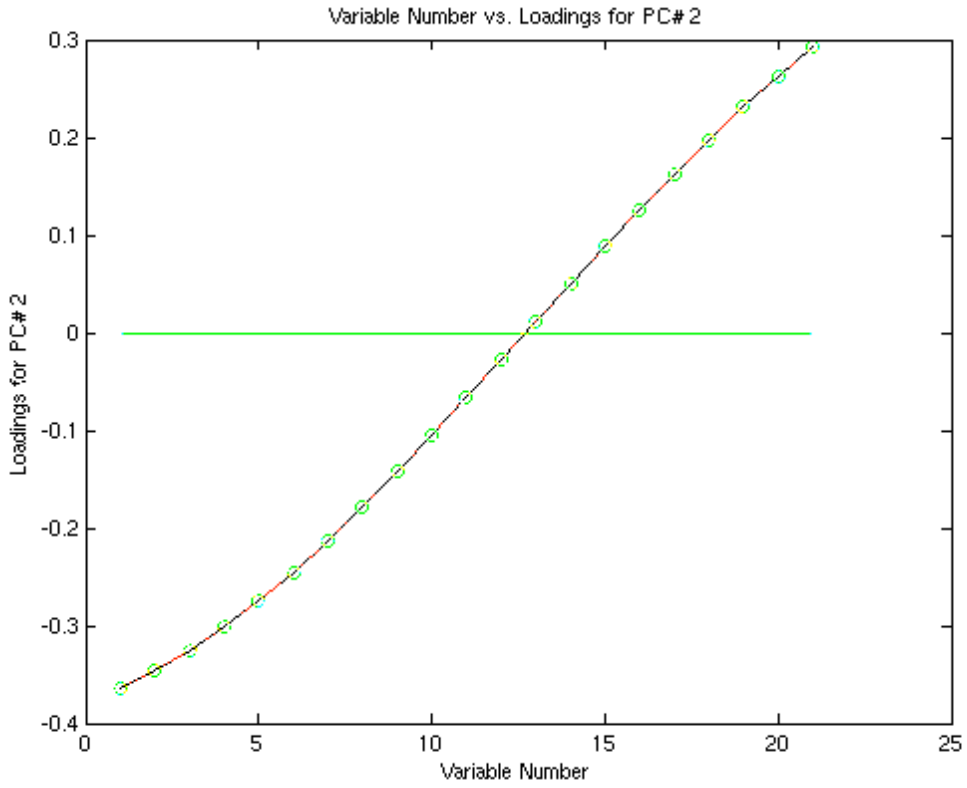


Figure 4-24: Loadings plot for PC2, reduced data set

The scores plot for PC3, Figure 4-25, shows a loss of structure, indicating that fundamentally no further information to noise enhancement is to be gained by trying to reduce the data set. Although little structure appears to be present in the data, there remain three principal and fundamental components associated with this data. The loadings plot for PC3 justifies this as well, as there still appears to be some structure present in the data set, but not much.

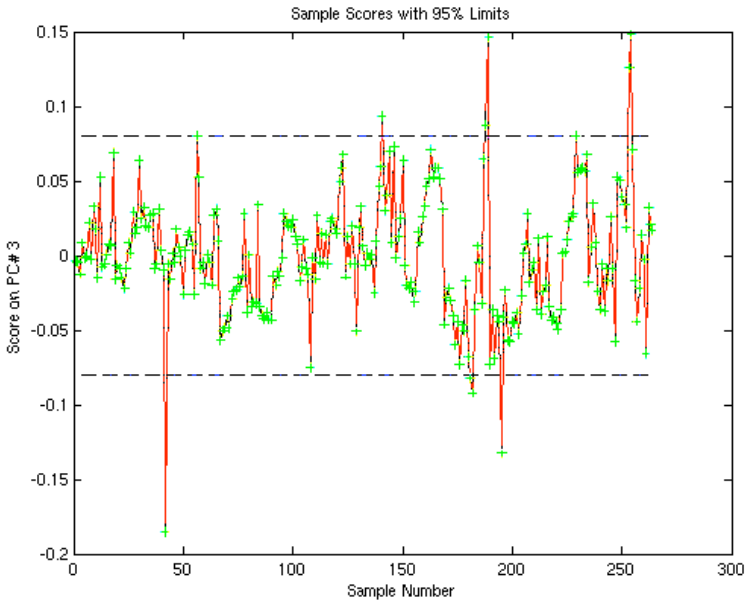


Figure 4-25: Scores plot for PC3, reduced data set

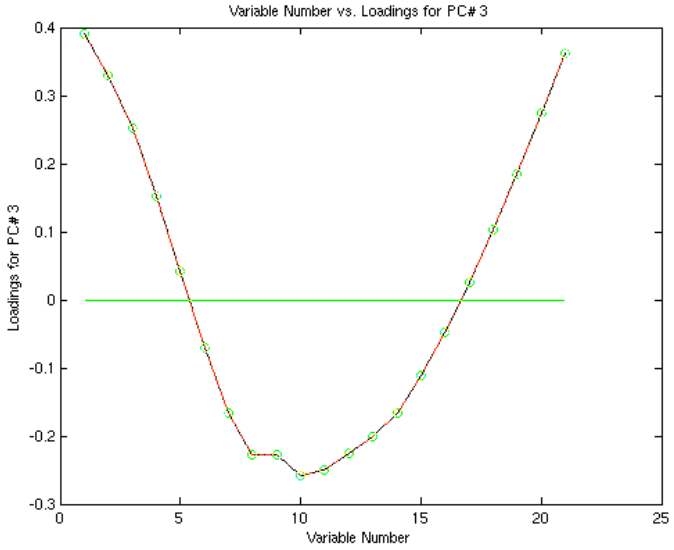


Figure 4-26: Loadings plot for PC3, reduced data set

Results

As the captured variance table indicated that three PCs would capture all the variance in the data, the PCA analysis included only 3 PCs. The results of this work were again disappointing, as the model did a very poor job predicting itself and thus could not be depended on to predict the test set. The results of the test set confirmed that the model had poor predictive ability. This is shown below in the Tables 4-9 through 4-12, corresponding to the four electrochemical channels reconstructed by the model.

Table 4-9: Results for the PCA Model, very reduced (21 points) data set, channel 1

Channel 1

n = 6, t_{table} = 2.57	Mix 1						Mix 2						Mix 3						Mix 4						Mix 5					
	Epi	Epi		Dopa	Dopa		Epi	Epi		Dopa	Dopa		Epi	Epi		Dopa	Dopa		Epi	Epi		Dopa	Dopa		Epi	Epi		Dopa	Dopa	
Technique	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$
PCR Model																														
Basis -- # PCs																														
2	38.3	3.4	17.0	11.7	3.4	**1.19**	32.4	0.7	52.49	18.8	0.9	56.60	21.8	1.7	17.60	26.8	1.5	65.56	14.0	2.83	5.73	38.2	1.2	109.77	4.35	1.98	3.10	48.3	1.7	104.77
3	38.3	3.4	2.571	10.6	2.5	4.76	32.4	0.7	29.81	26.0	6.4	6.18	21.8	1.7	17.58	33.9	1.9	42.22	13.9	2.77	5.87	27.3	23.7	6.85	4.36	1.98	3.10	51.6	4.0	42.18
4	38.4	3.4	17.2	10.2	2.9	4.40	32.6	0.8	26.94	24.8	8.0	5.32	22.1	1.9	15.95	32.1	2.5	33.98	14.5	3.72	4.00	23.8	29.7	5.76	4.55	1.81	3.14	50.4	3.1	54.86
Test Set -- # PCs																														
2	36.8	4.7	13.1	10.2	1.4	5.57	32.6	0.5	78.95	18.3	1.5	36.68	19.7	0.8	42.24	28.1	1.2	82.01	11.94	2.85	7.43	38.5	0.7	206.38	5.69	1.34	**2.13**	48.2	1.6	113.46
3	36.8	4.7	2.571	10.3	1.3	9.89	32.6	0.5	44.17	24.1	2.7	16.24	19.7	0.8	42.35	35.5	2.2	35.33	11.9	2.84	7.45	41.6	5.1	24.87	5.70	1.34	**2.13**	49.7	3.7	46.94
4	36.8	4.6	12.0	10.0	1.0	13.14	32.7	0.5	45.18	23.7	2.2	20.26	20.0	0.8	44.38	33.4	1.8	45.93	12.2	2.66	7.73	40.0	4.2	31.03	5.66	1.31	**2.26**	50.0	3.5	49.24
PLS Model																														
Basis set -- # of PCs																														
2	38.3	3.4	17.0	11.7	3.4	**1.19**	32.4	0.7	52.38	18.8	0.9	56.77	21.8	1.7	17.60	26.8	1.5	65.54	13.97	2.83	5.71	38.2	1.3	108.71	4.35	1.98	3.10	48.3	1.7	104.70
3	38.3	3.4	2.571	10.4	2.4	5.09	32.4	0.7	29.72	27.0	7.8	4.77	21.8	1.7	17.58	34.7	2.0	38.74	13.9	2.79	5.83	25.9	25.9	6.40	4.36	1.98	3.10	51.9	3.8	43.64
4	38.3	3.4	17.4	10.5	4.1	2.96	32.4	0.7	29.83	25.2	9.9	4.20	21.8	1.7	17.61	31.2	4.6	19.00	14.0	2.81	5.78	31.6	7.3	20.96	4.36	1.99	3.09	49.6	3.0	58.23
Test Set -- # PCs																														
2	36.8	4.7	13.1	10.2	1.4	5.57	32.6	0.5	78.81	18.3	1.5	36.68	19.7	0.8	42.22	28.1	1.2	81.97	11.94	2.85	7.43	38.5	0.7	205.69	5.69	1.34	**2.13**	48.2	1.6	113.37
3	36.8	4.7	2.571	10.3	1.1	11.12	32.6	0.5	44.12	25.2	3.3	12.63	19.7	0.8	42.31	36.5	2.4	31.54	11.9	2.84	7.45	42.0	5.5	23.07	5.70	1.34	**2.13**	50.2	3.9	43.86
4	36.8	4.7	11.7	11.4	1.2	8.25	32.6	0.5	44.59	26.7	10.2	3.73	19.7	0.8	42.30	34.9	3.0	26.43	11.9	2.85	7.43	41.4	4.8	26.92	5.70	1.34	**2.12**	52.2	2.7	63.06
SIMPLS Model																														
Basis set -- # of PCs																														
2	38.3	3.4	17.0	11.7	3.4	**1.19**	32.4	0.7	52.38	18.8	0.9	56.77	21.8	1.7	17.60	26.8	1.5	65.54	13.97	2.83	5.71	38.2	1.3	108.72	4.35	1.98	3.10	48.3	1.7	104.70
3	38.3	3.4	2.571	10.4	2.4	5.09	32.4	0.7	29.72	27.0	7.8	4.77	21.8	1.7	17.58	34.7	2.0	38.74	13.9	2.79	5.83	25.9	25.9	6.40	4.36	1.98	3.10	51.9	3.8	43.64
4	38.3	3.4	17.4	10.5	4.1	2.96	32.4	0.7	29.83	25.2	9.9	4.20	21.8	1.7	17.61	31.2	4.6	19.00	14.0	2.81	5.78	31.6	7.3	20.96	4.36	1.99	3.09	49.6	3.0	58.24
Test Set -- # PCs																														
2	36.8	4.7	13.1	10.2	1.4	5.57	32.6	0.5	78.81	18.3	1.5	36.68	19.7	0.8	42.22	28.1	1.2	81.97	11.94	2.85	7.43	38.5	0.7	205.69	5.69	1.34	**2.13**	48.2	1.6	113.37
3	36.8	4.7	2.571	10.3	1.1	11.12	32.6	0.5	44.12	25.2	3.3	12.63	19.7	0.8	42.31	36.5	2.4	31.54	11.9	2.84	7.45	42.0	5.5	23.07	5.70	1.34	**2.13**	50.2	3.9	43.86
4	36.8	4.7	11.7	11.4	1.2	8.25	32.6	0.5	44.59	26.7	10.2	3.74	19.7	0.8	42.30	34.9	3.0	26.43	11.9	2.85	7.43	41.4	4.8	26.92	5.70	1.34	**2.12**	52.2	2.7	63.06

Key: Bolded values are those that pass the Student's t-test to 95% certainty. t_{cal} values are the calculated t-values for the t-test, where n=6, t_{Table} = 2.57

Table 4-10: Results for the PCA Model, very reduced (21 points) data set, channel 2

Channel 2

$n = 6$, $t_{Table} = 2.57$	Mix 1					Mix 2					Mix 3					Mix 4					Mix 5				
Technique	Epi Ave	Epi Std Dev	t_{Cal} Epi	Dopa Ave	Dopa Std Dev	t_{Cal} Dopa	Epi Ave	Epi Std Dev	t_{Cal} Epi	Dopa Ave	Dopa Std Dev	t_{Cal} Dopa	Epi Ave	Epi Std Dev	t_{Cal} Epi	Dopa Ave	Dopa Std Dev	t_{Cal} Dopa	Epi Ave	Epi Std Dev	t_{Cal} Epi	Dopa Ave	Dopa Std Dev	t_{Cal} Dopa
PCR Model																								
Basis – # PCs																								
2	29.8	1.4	57.09	9.38	1.25	7.82	25.0	1.4	39.59	19.4	1.1	45.29	16.7	1.1	40.10	28.1	0.2	448.71	9.1	1.2	23.70	42.4	0.4	335.56	0.81	1.51	9.82	54.1	2.6	61.84
3	29.8	1.4	2.571	4.10	1.82	15.15	25.0	1.4	27.50	21.3	4.0	12.90	16.7	1.1	40.09	34.3	1.2	67.67	9.14	1.16	24.06	39.3	6.8	19.44	0.821	1.503	9.84	59.4	5.5	27.11
4	29.6	1.3	28.24	5.12	1.69	14.87	25.2	1.7	22.90	20.3	2.5	21.65	16.9	1.0	40.84	33.6	1.0	85.26	8.74	0.50	58.09	41.9	2.0	62.67	0.85	1.58	9.28	59.1	6.1	24.63
Test Set – # PCs																								
2	29.1	1.3	62.23	9.91	0.23	36.90	25.5	1.2	45.25	18.4	1.2	45.97	16.4	0.6	71.19	29.1	0.1	1211.64	9.05	1.50	18.84	42.2	2.1	59.16	2.59	0.84	12.41	52.5	2.0	81.61
3	29.1	1.3	2.571	4.38	0.50	53.55	25.5	1.2	31.37	19.3	0.6	99.81	16.4	0.6	70.72	35.1	1.7	46.82	9.05	1.49	18.93	43.4	4.9	25.03	2.60	0.84	12.45	53.4	2.8	59.57
4	29.0	1.3	26.95	5.14	0.06	424.48	25.4	1.2	32.77	19.5	0.8	71.63	16.6	0.7	62.09	33.9	1.8	44.97	9.07	1.34	21.06	43.2	3.9	31.65	2.66	0.85	12.13	53.0	3.0	54.47
PLS Model																								
Basis set -- # of PCs																								
2	29.8	1.4	57.09	9.38	1.25	7.82	25.0	1.4	39.64	19.4	1.1	45.38	16.7	1.1	40.10	28.1	0.2	447.97	9.1	1.2	23.65	42.4	0.4	337.15	0.81	1.51	9.81	54.1	2.6	61.78
3	29.8	1.4	2.571	3.57	1.89	15.32	25.0	1.4	27.60	21.3	3.8	13.36	16.7	1.1	40.08	35.1	1.2	62.67	9.14	1.17	23.92	39.6	6.9	19.10	0.819	1.503	9.85	61.2	7.8	18.67
4	29.8	1.4	27.52	6.74	1.82	11.61	25.0	1.4	27.83	18.4	6.0	9.64	16.7	1.1	39.97	32.4	2.6	32.87	9.15	1.18	23.69	46.4	2.7	43.27	0.83	1.49	9.94	71.4	29.2	4.11
Test Set – # PCs																								
2	29.1	1.3	62.22	9.91	0.23	36.95	25.5	1.2	45.27	18.4	1.2	45.99	16.4	0.6	71.25	29.2	0.1	1200.57	9.05	1.50	18.83	42.2	2.1	59.11	2.59	0.84	12.40	52.5	2.0	81.61
3	29.1	1.3	2.571	3.64	0.35	81.07	25.5	1.2	31.40	19.6	0.5	110.82	16.4	0.6	70.87	35.8	1.9	40.60	9.05	1.49	18.90	43.6	5.1	24.03	2.59	0.84	12.44	53.7	3.0	54.80
4	29.1	1.3	28.05	5.08	1.80	13.99	25.5	1.2	31.47	21.6	2.4	20.56	16.4	0.6	71.07	33.5	2.3	35.01	9.05	1.50	18.86	43.7	2.6	47.60	2.60	0.84	12.47	56.4	9.5	16.44
SIMPLS Model																								
Basis set -- # of PCs																								
2	29.8	1.4	57.08	9.38	1.25	7.82	25.0	1.4	39.64	19.4	1.1	45.38	16.7	1.1	40.10	28.1	0.2	447.97	9.1	1.2	23.65	42.4	0.4	337.15	0.81	1.51	9.81	54.1	2.6	61.78
3	29.8	1.4	2.571	3.57	1.89	15.32	25.0	1.4	27.60	21.3	3.8	13.36	16.7	1.1	40.08	35.1	1.2	62.67	9.14	1.17	23.92	39.6	6.9	19.10	0.819	1.503	9.85	61.2	7.8	18.67
4	29.8	1.4	27.52	6.74	1.82	11.61	25.0	1.4	27.83	18.4	6.0	9.64	16.7	1.1	39.97	32.4	2.6	32.87	9.15	1.18	23.69	46.4	2.7	43.26	0.83	1.49	9.94	71.4	29.2	4.11
Test Set – # PCs																								
2	29.1	1.3	62.22	9.91	0.23	36.95	25.5	1.2	45.27	18.4	1.2	45.99	16.4	0.6	71.25	29.2	0.1	1200.57	9.05	1.50	18.83	42.2	2.1	59.11	2.59	0.84	12.40	52.5	2.0	81.61
3	29.1	1.3	2.571	3.64	0.35	81.07	25.5	1.2	31.40	19.6	0.5	110.82	16.4	0.6	70.87	35.8	1.9	40.60	9.05	1.49	18.90	43.6	5.1	24.03	2.59	0.84	12.44	53.7	3.0	54.80
4	29.1	1.3	28.05	5.08	1.80	13.98	25.5	1.2	31.47	21.6	2.4	20.56	16.4	0.6	71.06	33.5	2.3	35.02	9.05	1.50	18.86	43.7	2.6	47.60	2.60	0.84	12.47	56.4	9.5	16.44

Key: Bolded values are those that pass the Student's t-test to 95% certainty. t_{Cal} values are the calculated t-values for the t-test, where n=6, $t_{Table} = 2.57$

Table 4-11: Results for the PCA Model, very reduced (21 points) data set, channel 3

Channel 3

$n = 6$, $t_{table} = 2.57$	Mix 1						Mix 2						Mix 3						Mix 4						Mix 5					
Technique	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$
PCR Model																														
Basis -- # PCs																														
2	79.7	1.0	46.36	17.3	2.5	3.80	70.0	3.2	16.81	48.0	2.0	9.72	47.4	3.2	10.04	77.2	1.4	18.26	33.5	3.9	8.22	112	2	28.57	12.94	1.79	8.32	147	3	18.75
3	79.7	0.9	2.571	12.94	1.38	4.35	70.0	3.2	22.00	53.2	2.6	10.64	47.4	3.2	10.08	88.5	3.2	16.30	33.5	3.9	8.22	103	5	4.50	12.92	1.78	8.33	138	5	8.47
4	79.6	0.9	178.00	13.58	1.09	4.05	70.1	3.1	23.15	52.7	3.5	7.48	47.6	3.2	10.26	86.9	3.4	14.63	33.3	3.7	8.44	104	4	6.62	12.69	1.76	8.13	140	5	9.14
Test Set -- # PCs																														
2	78.5	0.9	45.84	18.9	0.7	18.22	70.5	2.6	20.94	46.0	1.6	9.12	50.1	1.0	39.50	76.8	2.3	10.76	32.9	4.9	6.24	113	7	6.67	17.27	1.04	24.58	143	2	33.94
3	78.5	0.9	2.571	14.44	0.69	3.32	70.5	2.6	27.30	51.0	1.9	11.22	50.1	1.0	39.63	85.4	4.7	9.60	32.9	4.8	6.24	105	10	2.81	17.23	1.03	24.63	128	3	7.74
4	78.4	0.9	182.91	15.36	0.89	0.05	70.6	2.7	27.12	50.5	1.6	12.94	50.3	0.9	42.25	84.1	4.2	10.01	32.8	4.9	6.09	106	10	2.87	16.83	0.85	28.88	131	2	16.88
PLS Model																														
Basis set -- # of PCs																														
2	79.7	1.0	46.35	17.3	2.5	3.80	70.0	3.2	16.82	48.0	2.0	9.74	47.4	3.2	10.04	77.2	1.4	18.26	33.6	3.9	8.22	112	2	28.52	12.94	1.79	8.32	147	3	18.74
3	79.7	0.9	2.571	12.23	1.47	5.25	70.0	3.2	22.02	54.0	3.7	7.85	47.4	3.2	10.07	89.8	3.4	16.25	33.5	3.9	8.22	101	6	3.37	12.92	1.78	8.32	137	6	7.26
4	79.7	0.9	168.53	12.03	3.63	2.26	70.0	3.2	21.98	52.8	8.9	2.92	47.4	3.2	10.06	84.4	2.4	18.08	33.5	3.9	8.23	101	7	2.90	12.92	1.79	8.31	138	7	5.98
Test Set -- # PCs																														
2	78.5	0.9	45.85	18.9	0.7	18.21	70.5	2.6	20.95	46.0	1.6	9.13	50.0	1.0	39.49	76.8	2.3	10.76	33.0	4.9	6.24	113	7	6.66	17.27	1.04	24.57	142	2	33.93
3	78.5	0.9	2.571	13.89	0.72	5.07	70.5	2.6	27.31	51.7	2.3	10.08	50.1	1.0	39.59	86.4	5.1	9.41	32.9	4.8	6.24	104	10	2.38	17.24	1.03	24.61	127	3	5.31
4	78.5	0.9	174.67	15.38	1.84	0.00	70.5	2.6	27.32	50.9	2.0	10.92	50.1	1.0	39.49	82.5	4.6	8.35	32.9	4.8	6.24	106	12	**2.53**	17.24	1.04	24.56	129	4	5.54
SIMPLS Model																														
Basis set -- # of PCs																														
2	79.7	1.0	46.35	17.3	2.5	3.80	70.0	3.2	16.82	48.0	2.0	9.74	47.4	3.2	10.04	77.2	1.4	18.26	33.6	3.9	8.22	112	2	28.53	12.94	1.79	8.32	147	3	18.74
3	79.7	0.9	2.571	12.23	1.47	5.25	70.0	3.2	22.02	54.0	3.7	7.85	47.4	3.2	10.07	89.8	3.4	16.25	33.5	3.9	8.22	101	6	3.37	12.92	1.78	8.32	137	6	7.26
4	79.7	0.9	168.53	12.03	3.63	2.26	70.0	3.2	21.98	52.8	8.9	2.92	47.4	3.2	10.06	84.4	2.4	18.08	33.5	3.9	8.23	101	7	2.90	12.92	1.79	8.31	138	7	5.98
Test Set -- # PCs																														
2	78.5	0.9	45.85	18.9	0.7	18.21	70.5	2.6	20.95	46.0	1.6	9.13	50.0	1.0	39.49	76.8	2.3	10.76	33.0	4.9	6.24	113	7	6.66	17.27	1.04	24.57	142	2	33.93
3	78.5	0.9	2.571	13.89	0.72	5.07	70.5	2.6	27.31	51.7	2.3	10.08	50.1	1.0	39.59	86.4	5.1	9.41	32.9	4.8	6.24	104	10	**2.38**	17.24	1.03	24.61	127	3	5.31
4	78.5	0.9	174.67	15.38	1.84	0.00	70.5	2.6	27.32	50.9	2.0	10.92	50.1	1.0	39.49	82.5	4.6	8.35	32.9	4.8	6.24	106	12	**2.53**	17.24	1.04	24.56	129	4	5.54

Key: Bolded values are those that pass the Student's t-test to 95% certainty. t_{Cal} values are the calculated t-values for the t-test, where n=6, $t_{Table} = 2.57$

Table 4-12: Results for the PCA Model, very reduced (21 points) data set, channel 4

Channel 4

$n = 6$, $t_{table} = 2.57$	Mix 1						Mix 2						Mix 3						Mix 4						Mix 5					
	Epi	Epi		Dopa	Dopa		Epi	Epi		Dopa	Dopa		Epi	Epi		Dopa	Dopa		Epi	Epi		Dopa	Dopa		Epi	Epi		Dopa	Dopa	
Technique	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$	Ave	Std Dev	$t_{cal\ Epi}$	Ave	Std Dev	$t_{cal\ Dopa}$
PCR Model																														
Basis -- # PCs																														
2	53.1	1.6	13.3	22.4	1.8	12.18	45.2	2.9	**2.35**	52.3	1.3	23.04	27.9	2.7	5.79	81.5	1.0	34.89	17.5	4.0	**1.88**	116	2	22.58	2.58	1.75	5.98	154	4	21.03
3	53.1	1.6	2.571	15.0	0.4	2.68	45.3	2.9	3.48	56.0	1.6	21.45	27.9	2.7	5.78	94.4	3.7	18.42	17.5	3.9	**1.89**	114	7	6.60	2.58	1.74	6.01	153	4	18.56
4	53.0	1.6	58.5	15.8	0.5	**2.28**	45.3	2.8	3.67	55.7	2.1	16.07	28.2	2.6	5.69	92.3	2.9	21.13	17.7	3.9	**1.85**	113	7	6.80	2.51	1.55	6.89	154	4	21.08
Test Set -- # PCs																														
2	52.5	1.1	19.9	22.7	1.8	12.74	45.4	2.1	3.12	50.9	0.8	33.72	27.7	3.9	4.13	83.2	1.9	20.78	17.6	4.7	**1.53**	117	4	15.60	5.91	0.53	4.42	149	1	116.49
3	52.4	1.1	2.571	15.3	0.8	**0.30**	45.4	2.1	5.07	53.6	2.2	12.68	27.7	3.9	4.12	95.0	3.2	21.74	17.6	4.7	**1.53**	116	10	5.41	5.90	0.52	4.54	144	2	25.00
4	52.3	1.2	79.0	16.1	1.0	**1.79**	45.5	2.1	5.06	52.8	1.9	13.68	28.2	3.9	3.84	91.7	2.8	21.40	17.8	4.5	**1.55**	115	8	6.25	5.62	0.34	8.99	145	1	54.25
PLS Model																														
Basis set -- # of PCs																														
2	53.1	1.6	13.3	22.4	1.8	12.18	45.2	2.9	**2.35**	52.3	1.3	23.07	27.9	2.7	5.79	81.5	1.0	34.86	17.5	4.0	**1.88**	116	2	22.55	2.58	1.75	5.97	154	4	21.04
3	53.1	1.6	2.57	14.0	0.3	10.22	45.3	2.9	3.48	56.3	2.0	17.66	27.9	2.7	5.78	95.8	4.0	17.89	17.5	3.9	**1.89**	113	8	5.86	2.58	1.75	6.00	153	5	17.27
4	53.1	1.6	60.0	16.4	1.2	**2.07**	45.2	2.9	3.47	52.7	1.7	14.94	27.9	2.7	5.79	89.0	3.5	15.34	17.5	3.9	**1.89**	111	7	6.19	2.58	1.75	5.99	149	6	12.47
Test Set -- # PCs																														
2	52.5	1.1	19.9	22.7	1.8	12.74	45.4	2.1	3.12	50.9	0.8	33.75	27.7	3.9	4.13	83.2	1.9	20.78	17.6	4.7	**1.53**	117	4	15.59	5.91	0.53	4.40	149	1	116.79
3	52.4	1.1	2.571	14.2	0.8	3.71	45.4	2.1	5.07	54.0	2.6	11.00	27.7	3.9	4.12	96.2	3.3	21.57	17.6	4.7	**1.53**	115	10	5.11	5.90	0.52	4.51	143	3	20.95
4	52.4	1.1	81.6	15.8	2.7	**0.37**	45.4	2.1	5.07	53.1	3.2	8.49	27.7	3.9	4.12	88.9	3.2	17.08	17.6	4.7	**1.53**	113	3	13.34	5.90	0.52	4.50	143	2	28.27
SIMPLS Model																														
Basis set -- # of PCs																														
2	53.1	1.6	13.3	22.4	1.8	12.18	45.2	2.9	**2.35**	52.3	1.3	23.07	27.9	2.7	5.79	81.5	1.0	34.86	17.5	4.0	**1.88**	116	2	22.55	2.58	1.75	5.97	154	4	21.04
3	53.1	1.6	2.571	14.0	0.3	10.22	45.3	2.9	3.48	56.3	2.0	17.66	27.9	2.7	5.78	95.8	4.0	17.89	17.5	3.9	**1.89**	113	8	5.86	2.58	1.75	6.00	153	5	17.27
4	53.1	1.6	60.0	16.4	1.2	**2.07**	45.2	2.9	3.47	52.7	1.7	14.94	27.9	2.7	5.79	89.0	3.5	15.34	17.5	3.9	**1.89**	111	7	6.19	2.58	1.75	5.99	149	6	12.47
Test Set -- # PCs																														
2	52.5	1.1	19.9	22.7	1.8	12.74	45.4	2.1	3.12	50.9	0.8	33.75	27.7	3.9	4.13	83.2	1.9	20.78	17.6	4.7	**1.53**	117	4	15.59	5.91	0.53	4.40	149	1	116.79
3	52.4	1.1	2.571	14.2	0.8	3.71	45.4	2.1	5.07	54.0	2.6	11.00	27.7	3.9	4.12	96.2	3.3	21.57	17.6	4.7	**1.53**	115	10	5.11	5.90	0.52	4.51	143	3	20.95
4	52.4	1.1	81.6	15.8	2.7	**0.37**	45.4	2.1	5.07	53.1	3.2	8.49	27.7	3.9	4.12	88.9	3.2	17.08	17.6	4.7	**1.53**	113	3	13.34	5.90	0.52	4.50	143	2	28.27

Key: Bolded values are those that pass the Student's t-test to 95% certainty. t_{cal} values are the calculated t-values for the t-test, where n=6, $t_{table} = 2.57$

Cramer's Rule Approach

The Cramer's Rule approach may be considered to be the basis from which the chemometric methods were developed. Indeed, such chemometric approaches typically employ Cramer's Rule multiple times to achieve a more comprehensive data treatment. The chemometric methods employed in the current investigation may have failed due to factors that are specific and/or overly inclusive with these methods. Alternatively, chemical interactions between the analytes and/or simple noise content in the data set could have been the underlying source of the inadequacy of the chemometric approaches to provide the desired deconvolution capabilities.

However, while the chemometric approaches had not provided the concentration predictive capabilities desired, they did demonstrate that a substantial amount of related informational content was contained in the data set. Further, these approaches seemed to indicate that the maximal amount of informational content with a relatively minimal amount of noise content was contained in the very reduced data set incorporating 21 time points. Thus, we decided to examine if the even simpler application of Cramer's rule to the very reduced data set using only the two-dimensional signal vs. potential plane at specific time points might be more favorable. This is simply labeled by us as the "Cramer's Rule" or CR method wherein data from all four channels was used at a single specified time point. In fact, it was repeated 21 times, once for each of the time points contained in the very reduced data set.

The chemometric approach indicated that a maximal information content appeared to be associated with the signal vs. potential plane; this was indicated usually in the second scores plot, which showed a regular pattern of four jagged plateaus. The

scores and loadings plots seem to indicate that no further reduction in the original time parameter could be undertaken to provide more signal to noise information. Stated another way, none of the 21 time points of the very reduced data set indicated strikingly better signal to noise informational content than any of the other time points in this very reduced data set. Therefore, all 21 points in time for the very reduced data set were examined using the CR approach.

The CR method was done as follows. All the data for the Epi standards were fitted to Equation 1.79, previously described, giving current as a function of the potential of the electrode:

$$i = \frac{-K_R C_R^*}{1 + 10^{-\left[\frac{(E - E_c)s}{0.05916}\right]}}$$

Equation 1.79

where E_c is the experimentally determined half-wave potential, s is overall experimentally determined number of electrons involved in the reaction, i is the current, E is the potential of the electrode, and C_R is the concentration of the reduced species. All of these parameters are constant for a given component (Epi or Dopa) except for the current and the potential. The potential vs. signal data was fitted to this equation using a residual-sum-of-squares method in Excel, using Excel's Solver function, which uses the Levenberg-Marquardt method of iteration for convergence. This was then repeated for the Dopa standards. This procedure thus yielded constants at given potentials related to the currents (Equation 1.80, reproduced below):

$$i = -K_R^{'} C_R^*$$

Equation 1.80

for both compounds, which now could be used in the final simultaneous linear equation solutions as derived in Chapter 1 and again reproduced here:

$$\begin{aligned}
i_{tot,Ch1} &= \left(K_O^{'} C_O^*\right)_{Ch1,Cmpd1} + \left(K_O^{'} C_O^*\right)_{Ch1,Cmpd2} + \cdots + \left(K_O^{'} C_O^*\right)_{Ch1,Cmpd(n)} \\
i_{tot,Ch2} &= \left(K_O^{'} C_O^*\right)_{Ch2,Cmpd1} + \left(K_O^{'} C_O^*\right)_{Ch2,Cmpd2} + \cdots + \left(K_O^{'} C_O^*\right)_{Ch2,Cmpd(n)} \\
i_{tot,Ch3} &= \left(K_O^{'} C_O^*\right)_{Ch3,Cmpd1} + \left(K_O^{'} C_O^*\right)_{Ch3,Cmpd2} + \cdots + \left(K_O^{'} C_O^*\right)_{Ch3,Cmpd(n)} \\
i_{tot,Ch4} &= \left(K_O^{'} C_O^*\right)_{Ch4,Cmpd1} + \left(K_O^{'} C_O^*\right)_{Ch4,Cmpd2} + \cdots + \left(K_O^{'} C_O^*\right)_{Ch4,Cmpd(n)}
\end{aligned}$$

Equation 1.87

Results

The results for the Cramer's Rule approach are given in Table 4-13 below. As this approach uses the data from all four channels, it is presented by each mixture set. In general, the method is generally accurate and reasonably precise, given the actual amounts of the analytes in the injected samples themselves. Although the relative percent RSD values appear high for the smaller values for both Epi and Dopa in the extreme samples, they are within the amount of signal variance expected for the data.

The results that provided proper predictions for the data according to the criteria selected by the Student's t-test (95% confidence levels) are bolded in Table 4-1. As can

be seen, only three times produced successful results. These were 1.6110, 1.6127, and 1.6210 minutes, where the latter corresponds to the retention time for Epi.

The retention times of Epi and Dopa are fixed as much as possible due to the time corrections. The time corrections fixed the solvent front time and the time of the internal standard, Dopamine. The variance in the retention times of both Epi and Dopa is reduced considerably, and is limited to the experimental sampling period. The data on each channel was also normalized in its response, as all the internal standard peaks for a given channel were made to be the same height. Thus, any changes in the current as seen by the internal standard could be eliminated. Since the mixtures are of known composition, there are no effects expected from the injected sample matrix.

In summary, the PCR chemometric approach clearly guided us to the region of the data collected which yielded the maximal amount of signal to noise information. However, this approach may be simply too inclusive in its application to the prediction of concentrations of the individual components. The much more simplistic Cramer's Rule approach did work in three discrete time domains near the retention time for Epi.

The fact that the Cramer's Rule approach seemed to consistently yield better predictability than the more powerful PCR chemometric approach leads us to believe that there is, indeed, an underlying more optimal approach to the deconvolution of these severely chromatographically and electrochemically overlapped analytes. In fact, one might note that, given the substantial overlap of these two analytes in both the chromatography and electrochemistry, it is remarkable that modestly reliable prediction of concentrations of the individual components within the mixtures was achieved.

Table 4-13: Summary of the data for the Cramer's Rule Approach

$n = 6$, $t_{table} = 2.57$	Mix 1						Mix 2						Mix 3						Mix 4						Mix 5					
Technique	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$	Epi Ave	Epi Std Dev	$t_{Cal\ Epi}$	Dopa Ave	Dopa Std Dev	$t_{Cal\ Dopa}$
Cramer's Rule Approach																														
At Difference Minimum (1.597 min)	56.7	9.8	**1.27**	16.2	45.4	**0.15**	48.6	2.7	**0.54**	54.6	23.3	**1.52**	30.3	3.8	2.58	92.1	22.9	2.69	14.5	3.3	4.45	131	19	4.80	3.31	3.10	2.81	149	16	4.35
At Epi Retention Time (1.621 min)	58.1	7.6	**1.15**	15.8	11.6	**0.52**	48.6	2.5	**0.60**	43.1	6.0	**1.23**	32.4	2.6	**1.75**	66.9	6.2	**0.00**	20.4	4.4	**0.08**	98.1	6.1	**1.77**	6.25	3.40	**0.44**	127	7	**2.41**
At Crossing Point (1.622 min)	61.5	8.1	**0.06**	16.2	11.3	**0.60**	51.5	2.7	3.21	43.5	5.8	**1.42**	34.4	2.7	**0.08**	67.7	6.0	**0.31**	21.5	4.4	**0.51**	99.4	5.7	**2.43**	6.47	3.76	**0.25**	129	7	2.97
At Ave. Retention Time (1.629 min)	58.3	7.6	**1.12**	16.1	7.4	**0.88**	48.6	2.7	**0.54**	43.1	3.7	**1.92**	32.0	1.9	3.00	68.4	3.3	**1.10**	20.2	2.5	**0.38**	97.5	2.8	3.29	6.25	4.15	**0.36**	126	6	**2.27**
At Dopa Retention Time (1.637 min)	58.2	7.9	**1.11**	16.0	6.2	**1.03**	48.6	3.3	**0.45**	43.1	3.4	**2.16**	31.6	1.2	5.42	69.5	1.9	3.32	20.4	2.2	**0.23**	96.8	2.6	3.01	6.77	4.51	**0.05**	124	5	**1.65**
At times (minutes)																														
1.6044	57.0	8.0	**1.43**	16.6	23.7	**0.33**	48.7	1.6	**1.23**	46.3	10.7	**1.43**	32.5	3.0	**1.44**	70.9	12.8	**0.78**	17.6	2.7	2.67	108	10	3.65	5.78	2.74	**0.96**	131	9	2.86
1.6060	57.1	7.9	**1.41**	16.2	21.8	**0.32**	48.8	1.6	**1.35**	44.9	9.7	**1.22**	32.7	2.8	**1.32**	68.9	11.5	**0.45**	18.3	2.7	**1.99**	105	9	3.13	6.51	2.65	**0.31**	128	8	**2.41**
1.6077	57.2	7.9	**1.39**	15.9	20.3	**0.31**	48.8	1.8	**1.14**	44.3	9.2	**1.11**	32.7	2.7	**1.41**	68.2	10.3	**0.32**	18.7	2.7	**1.67**	104	9	2.91	6.84	2.72	**0.01**	127	8	**2.15**
1.6094	57.1	7.7	**1.43**	15.9	18.4	**0.34**	48.6	1.9	**0.87**	44.0	8.6	**1.10**	32.6	2.7	**1.50**	67.8	9.6	**0.24**	19.0	3.0	**1.26**	103	8	2.66	6.76	2.81	**0.07**	128	8	**2.19**
1.6110	57.1	7.6	**1.47**	16.0	17.0	**0.38**	48.4	2.2	**0.51**	43.9	8.4	**1.11**	32.5	2.7	**1.60**	67.5	9.0	**0.17**	19.5	3.4	**0.79**	101	8	**2.49**	6.43	3.04	**0.34**	128	8	**2.46**
1.6127	57.0	7.6	**1.48**	16.0	16.0	**0.40**	48.1	2.4	**0.15**	44.1	8.2	**1.19**	32.4	2.7	**1.68**	67.3	8.5	**0.14**	19.6	3.9	**0.57**	101	8	**2.29**	6.39	3.37	**0.33**	128	8	**2.38**
1.6144	57.1	7.7	**1.44**	15.8	15.1	**0.40**	48.0	2.5	**0.06**	43.9	7.8	**1.18**	32.3	2.7	**1.80**	67.3	8.0	**0.15**	19.8	4.3	**0.41**	100	8	**2.10**	6.46	2.84	**0.33**	128	7	2.75
1.6160	57.0	7.5	**1.52**	16.1	13.8	**0.49**	47.9	2.5	**0.08**	43.9	7.3	**1.28**	32.3	2.6	**1.86**	67.2	7.4	**0.11**	20.3	4.5	**0.15**	99.2	7.2	**1.88**	6.65	2.93	**0.16**	128	7	2.68
1.6177	57.1	7.6	**1.46**	16.0	13.1	**0.49**	47.7	2.6	**0.21**	44.0	7.0	**1.36**	32.0	2.5	**2.23**	67.6	6.7	**0.27**	20.1	4.6	**0.26**	99.4	7.0	**2.03**	6.10	3.17	**0.58**	129	7	2.98
1.6194	57.2	7.5	**1.44**	16.0	12.3	**0.52**	47.9	2.5	**0.07**	43.7	6.4	**1.36**	32.0	2.6	**2.16**	67.5	6.4	**0.25**	20.2	4.5	**0.21**	99.2	6.5	**2.11**	6.19	3.20	**0.51**	128	7	2.97
1.6210	58.1	7.6	**1.15**	15.8	11.6	**0.52**	48.6	2.5	**0.60**	43.1	6.0	**1.23**	32.4	2.6	**1.75**	66.9	6.2	**0.00**	20.4	4.4	**0.08**	98.1	6.1	**1.77**	6.25	3.40	**0.44**	127	7	**2.41**
1.6227	56.6	7.5	**1.66**	16.2	11.1	**0.62**	47.3	2.5	**0.60**	43.5	5.6	**1.49**	31.5	2.4	2.72	67.7	5.7	**0.38**	19.7	3.9	**0.53**	99.3	5.4	2.61	5.93	3.46	**0.65**	128	7	3.00
1.6244	56.8	7.5	**1.60**	16.1	10.5	**0.64**	47.2	2.5	**0.69**	43.7	5.3	**1.67**	31.5	2.3	2.85	67.9	5.3	**0.49**	19.7	3.5	**0.57**	99.2	4.7	2.96	5.90	3.47	**0.67**	128	6	3.09
1.6260	56.9	7.6	**1.55**	16.2	10.0	**0.69**	47.3	2.4	**0.64**	43.7	5.0	**1.79**	31.5	2.3	2.92	68.1	5.0	**0.60**	19.7	3.4	**0.61**	99.2	4.4	3.14	5.98	3.61	**0.59**	128	6	3.03
1.6277	57.0	7.5	**1.50**	16.2	9.5	**0.72**	47.4	2.5	**0.53**	43.7	4.7	**1.86**	31.5	2.2	3.07	68.3	4.6	**0.77**	19.7	3.1	**0.67**	99.2	4.0	3.42	5.95	3.76	**0.58**	128	6	3.00
1.6294	57.1	7.6	**1.47**	16.3	9.2	**0.77**	47.5	2.6	**0.46**	43.7	4.7	**1.92**	31.4	2.0	3.40	68.6	4.2	**0.99**	19.8	2.9	**0.62**	98.9	3.7	3.54	5.94	3.94	**0.56**	128	6	2.97
1.6310	57.3	7.6	**1.41**	16.2	8.8	**0.80**	47.7	2.6	**0.24**	43.6	4.4	**1.94**	31.5	2.1	3.13	68.6	4.1	**1.03**	19.8	2.9	**0.62**	98.9	3.6	3.62	5.97	3.96	**0.54**	128	6	2.98
1.6327	57.6	7.5	**1.32**	16.1	8.3	**0.82**	47.9	2.7	**0.02**	43.5	4.3	**1.96**	31.6	1.9	3.32	68.7	3.8	**1.19**	19.8	2.8	**0.63**	98.8	3.3	3.85	6.05	3.98	**0.50**	127	6	2.94
1.6344	57.7	7.7	**1.25**	16.2	8.1	**0.87**	48.2	2.8	**0.24**	43.4	4.1	**1.97**	31.8	2.0	2.98	68.8	3.7	**1.29**	20.0	2.5	**0.51**	98.6	3.0	4.06	6.06	4.15	**0.47**	127	6	2.89
1.6360	58.0	7.6	**1.17**	16.2	7.7	**0.89**	48.4	2.7	**0.40**	43.5	3.9	**2.13**	31.9	2.0	2.91	69.0	3.5	**1.47**	20.1	2.4	**0.41**	98.4	2.9	4.16	6.20	4.17	**0.38**	127	6	2.85
1.6377	58.2	7.6	**1.10**	16.2	7.4	**0.94**	48.6	2.8	**0.61**	43.4	3.7	**2.18**	31.9	1.8	3.12	69.1	3.2	**1.75**	20.1	2.5	**0.43**	98.5	2.8	4.27	6.24	4.20	**0.35**	127	6	2.81
1.6394	58.5	7.8	**0.98**	16.1	7.3	**0.93**	48.9	3.1	**0.78**	43.4	3.9	**2.05**	32.1	1.7	3.25	69.2	2.9	**2.01**	20.3	2.4	**0.25**	98.3	2.7	4.29	6.36	4.33	**0.28**	127	6	2.71

Bolded values are those quantities passing the Student't t-test at 95% certainty.

Chapter 5 — Conclusions

Discriminating between and determining individual contributions from severely overlapping chromatographic peaks has long been a vexing problem in analytical chemistry. The application of three-dimensional instrumental techniques to such difficult problems has allowed for some discrimination; the added third dimension typically targets a specific property of the analyte, such as its volatility or absorption of light. The utilization of electrochemical information as the third dimension, discussed in Chapter 1, has not commonly been found in conjunction with chromatography.

The instrument used in this work is HPLC with electrochemical detection. The detector contains four glassy carbon serial electrodes, each set to a progressively higher potential. Thus, the data included time, signal, and the potential of each electrode. Two analytes were chosen: Epi and Dopa. These two are overlapped chromatographically so severely ($R_s = 0.13$) that the two peaks, if constructed to be of equal height individually, appear to be a single peak with no discernible valley or other indication to the observer that they are indeed two peaks.

Five mixture sets were made to encompass a range of molar response ratios for Epi:Dopa using the signals values on the first electrochemical channel. The mixtures ratios for Epi:Dopa signal were: 90:10, 70:30, 50:50, 30:70, and 10:90. All Epi:Dopa mixtures and all Epi and Dopa standards also contained an internal standard, Dopamine. Six replicate injections were made of each set of five mixtures, and three replicates of Epi and Dopa standards were injected both before and after each mixture set.

The subsequent data set contained several known and/or assumed sources of variance. These included: (1) shifts in the solvent front peak times due to manually injecting the samples onto the HPLC-ECD system; (2) long-term changes in the retention time of the internal standard presumed due to long-term flow rate oscillations in the reciprocating pump; and (3) changes in the signal height due to differences in the injection volume due at least in part to the use of manual, rather than automated, injections. Fortunately, these known and substantial sources of "noise" in our data set were easily corrected, for the most part, through standardization of the individual chromatorgrams in both time and signal strength using the retention time of the solvent peak, the retention time of the internal standard, and the signal strength of the internal standard.

Two major methods to deconvolute such highly overlapping chromatographic peaks were investigated. The first method was curve fitting, where twelve common curve-fitting models having associated mathematical equations were evaluated. The mathematical models included the bi-Gaussian, Exponentially Modified Gaussian (EMG), Fraser-Suzuki, log-normal, Haarhoff-van der Linde, Cauchy-Gaussian, Chesler-Cram, Giddings, GMG, GEMG4, GEMG5, EMG+GMG. Each of these twelve models was examined for its ability to fit the Epi and Dopa standard injection data sets. The equations that produced the best results were the EMG, GEMG4, GEMG5, EMG+GMG and Cauchy-Gaussian. These five models were subsequently used to deconvolute the two highly overlapped peaks using information from the time vs. signal domain for each of the four electrochemical detectors, where both area under the curve and peak heights were examined.

A second approach involved using common chemometric approaches with an initial principal component analysis being followed by submission of the data to PCR, PLS, and SIMPLS for deconvolution and prediction of Epi and Dopa concentrations in "unknown" samples. An auxiliary to the second approach examined the potential vs. signal domain with Cramer's rule at single data time points. An acceptable accuracy in the predictability of results was set so that the predictions would fall with the 95% confidence level (Student's t-test) for both analytes. Any results that fell short of this requirement were deemed to be unacceptable. Additionally, predictions which met these criteria were further examined to guarantee that they possessed a relatively low percent relative standard deviation, indicating a reasonable amount of precision in the predictions as well. Ideally, the outcomes should be both accurate and precise.

The use of curve fitting methods failed to adequately deconvolute the data as presented for all four data channels for the range of molar response ratios studied. When the curve fitting methods were able to deconvolute the data for an individual mixture set, it rarely was able to do so with both good accuracy and good precision. For a given mixture examined on a single electrochemical channel, several curve fit equations were found to be useful in properly deconvoluting the data. However, the same individual equations were not consistently capable of successful and acceptable deconvolution from one mixture set to another mixture set and/or from one electrochemical channel to another.

An argument could be made that the mixtures on the far extremes, i.e., Mixes 1 and 5, both of which contained 10 % of the minor species, are more difficult to properly deconvolute and thus could be expected to fail due to the large difference in molar

response ratios. The argument further could indicate that a method was considered successful if it predicted the composition of the "middle" mixture sets, i.e., mixtures 2 through 4, but this was not observed.

These methods may have failed because of the nature of the surface to be fitted. An iterative process was used to fit the observed data to the various mathematical models. The "combined" function used to deconvolute the data is in actuality two identical functions added together, and then the parameters are allowed to change until a minimum in the sum-of-squares of the residual (SSR) is found. The minimum in the SSR is shallow, and a large number of possible combinations of the parameters for the equations will satisfy the SSR requirements, even when constricting the values to those that are considered to be reasonably or physically meaningful. Some of these combination of parameters, when put back into the equations, were found to be unable to accurately quantify the solutions. The found parameters yielded equations that when added together, resulted in a very low SSR value, however, these were not able to accurately and precisely determine the known amounts of the analytes in the mixtures.

The use of curve fitting methods has previously been shown to work reasonably well when the peaks are separated by a distance sufficient to generate an easily discernible valley between them. But once the peaks overlap to the extent that valley disappears there is nothing for the equation to fit but the peak as a whole. As the two peaks merge into one, the remaining single peak takes on characteristics of a peak containing only a single analyte. The curve fitting methods cannot, unfortunately, easily distinguish between the merged double peak and a single peak.

The chemometric methods required the data to be presented to the algorithms in a structured way, depending on how one wishes to investigate the system. This structure then determines the nature of the information one can and does obtain. The procedure for these methods included (1) the determination of the number of dimensions inherent in the n-dimensional data space using the scree plots, percent variance tables, and the loadings and scores plots using all the data available; (2) building the model using a subset of the data determined to be the basis set; and (3) subsequent prediction of the values for the dependent variables based on the remainder of the data, designated the test set. Building the model depends on the correlations constructed with the known values for the dependent variables; the subsequent prediction of the "unknowns" depends on the strengths of these correlations.

The chemometric methods (PCR, PLS, and SIMPLS) all had difficulties. In all three data sets—the entire data set comprised of 481 time points, the reduced data set comprised of 121 data points, and the very reduced data set comprised of 21 data points—the predictive abilities of the models were overall poor. However, the PCA approaches clearly worked better than the curve fitting methods, as there was some informational content which we were able to glean from the data. The degree of accuracy and precision, however, in using the PCA approaches for predictions ranged from virtually unusable to modest at best. The most information concerning prediction of the concentrations of Epi and Dopa in the mixtures was obtrained using the simpler Cramer's Rule approach. Several explanations may be pertinent to this outcome, and some of these are discussed below.

Dopa, Epi, and Dopamine are all biogenic amines which almost certainly chemisorb to some degree onto the surface of the glassy carbon electrodes used in this study. Previous work involving DOPAC and 5-HIAA in this laboratory[141] found that both compounds absorbed appreciably in a concentration dependent manner onto powdered glassy carbon at room temperature with no applied potential employed concentrations similar to those used in the current investigation. The adsorptions of these molecules were described reasonably well by Langmuir isotherms. These results, as well as other work concerning biogenic amines and glassy carbon are consistent with the proposal that these compounds form a monolayer of neutral species on the electrode surface.[142] One could also reasonably surmise that the oxidation products of these molecules would likewise undergo chemisorption.

The hydrodynamic voltammograms for the two analytes are shown below in Figure 5-1. At the low potential of 600mV, Epi is expected to have a negligible response. At 900mV, Epi is expected to contribute appreciably to the current measured at the electrode surface. In the experiment, however, at 600mV there was a considerable response from Epi. Clearly, something is decreasing the effective oxidation potential for Epi; we are suggesting that this could due to a chemisorption-coupled or otherwise surface alteration enhancement of the electron transfer process.

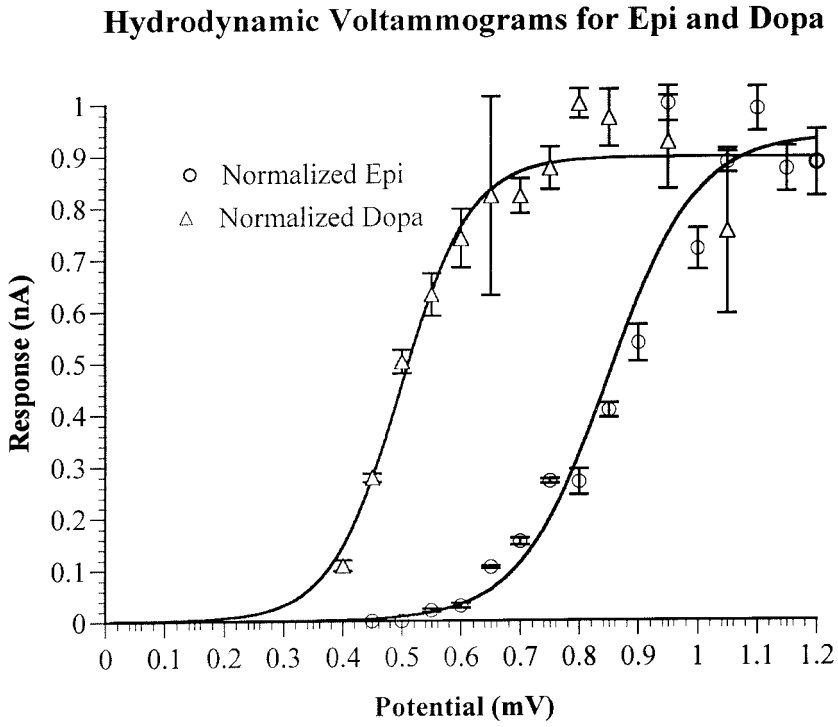


Figure 5-1: Hydrodynamic voltammograms of Epi and Dopa

The transfer of electrons in electrochemical detection occurs at the surface of the electrode both as inner-sphere and outer-sphere reactions.[143] Transmission of electrons via other molecules or ions away from the surface of the electrode does not contribute to the electrode response. Inner-sphere reactions involve specific absorption of species involved in the reaction with the electrode. Any chemisorbed species on the electrode surface that accepts electrons from compounds in solution can be expected to either facilitate or dampen the electron transfer between the analytes and the electrode. One-component solutions, not mixtures, were used to determine the hydrodynamic voltammograms of the analytes, thus, the oxidation potentials determined may not accurately reflect the oxidation potentials of the mixtures due to this chemisorption phenomenon.

When a biogenic amine passes over the surface of the electrode, the electrode provides the energy for a one-electron oxidation to occur. The biogenic amine briefly becomes a radical cation that is expected to immediately deprotonate to form a neutral radical that is known to rapidly oxidize and deprotonate to form the quinone; this second electron transfer typically occurs too fast to be seen as a second wave in the HDV, but can occasionally be seen in cyclic voltammetry Each of the species involved have the capability of chemisorbing onto the electrode surface.

Chemisorbed species can cause a change in the electrochemical behavior of the electrode surface, which results in a subsequent shift in experimentally determined $E_{1/2}$ of the hydrodynamic voltammogram. The reversibility of the reaction at the surface of electrode also changes, indicating a shift in the slope of the HDV to a steeper angle. The changes to the HDV are two-fold: both a shift to lower potentials and a steeper slope of the curve, as evidenced by the response for Epi discussed above.

Epi is used to illustrate the complexity of the reactions generated by the electrode, following the initial two-electron oxidation of Epi to its quinoid form, the unprotonated form of the epinephrine quinoid undergoes a further reaction to yield leucoadrenochrome, a cyclized hydroquinone. Leucoadrenochrome is easier to oxidize than epinephrine itself and it reacts with another molecule of ephinephrine quinone, giving adrenochrome and regenerating epinephrine. This electron transfer-chemical reaction-chemical (ECC) reaction mechanism results in an apparent four electron oxidation of ephinephrine to adrenochrome.[144] Should the leucoadrenochrome not leave the surface of the electrode before it reacts with the ephinephrine quinone, it will result in increased current for Epi.

The standards and samples used herein contained three biogenic amines: Epi, Dopa, and Dopamine. Dopamine is known to interact with the electrode surface, and can adsorb to it.[142] All three compounds are very similar chemically to one another, so it would not be surprising to find that any of the three, either singly or in combination, may be involved in binding to the electrode surface. Such films that develop on the surface of the carbon electrode have spurred research into the field of modified glassy carbon electrodes, where a desirable film is bonded to the surface of the electrode and serves usually to eliminate an interference.[145, 146] In this case, the film is temporary, and can be removed to varying levels of success by several methods, including reversing the potential on the electrode and allowing the film to dissipate. With enough time, the film will slowly disappear. A quicker and more thorough method to remove the film is to polish the electrode.

Studies concerning the nature of the glassy carbon electrode surface have found that a freshly polished surface has poor electron-transfer kinetics, with the active sites for electron transfer present on the graphite edges inherent in the structure of glassy carbon[147] and some species can attack basal plane sites as well.[148] Activation of the electrode surface has been attributed to the formation of graphitic oxide,[149] oxygen containing functional groups, and edge plane density of the electrode surface itself.[150] Further investigation has shown that an absorption mechanism based on specific chemical sites such as oxides or surface radicals cannot account for the absorption seen at the surface of the electrode, but rather the absorption of quinones on glassy carbon depends on effects such as electrostatic attraction between the adsorbate and partial surface charges.[151] Thus, the activated areas are larger than the edge plane regions on the

electrode surface. However, if the surface is contaminated with oxygen bound to the surface of the electrode, deactivation of the glassy carbon electrode can occur.[152] This also accounts for the disappearance of the surface layer when the potential is turned off as the electrostatic conditions change.

The model can apparently partially model this chemisoption phenomenon. But since the effect is also dependent on the potential of the electrode as well as the concentration of the analyte in a complex manner, it fails with the rise in potentials and the change in the analyte concentrations. Additionally, the amount of data one submits to the model has a large part in the ability of the model to accurately return the expected outcomes. The entire data set, which had 481 time measurements, clearly had some measurements that contained no useful information. When these are eliminated from the data set, the predictive ability of the subsequent model increases. However, when even more time points suspected to contain more noise than information are taken out of the data, the predictive ability of the new model decreases. The conclusions one must draw are two-fold.

The first conclusion is there is some optimum number of variable (time) points to be used for the model that eliminates the additional error associated with noise, but retains enough of the informational content to enable the models to predict accurately the values of the unknowns. This optimum number of variables (time) is not easily determined using these methods, and may be a matter of experience of the investigator.

The second conclusion is that the predictive abilities of the chemometric methods are unable to be successfully employed due to the noise and inherent uncertainty in the time vs. signal plane. The separation power found using the Cramer's Rule approach on

individual data points in the potential vs. signal plane is lost in the inability of the models to separate the two analytes in the time vs. signal plane. Some predictive ability is clearly being expressed by the chemometric methods, but may be lost in the inherent uncertainties found on this time vs. signal plane. Either the number of PCs used to form the models are incorrect, or the model itself has some inherent difficulties in using the information present in the third dimension, potential.

Increasing the number of PCs from two to higher numbers was investigated; an examination of the scree plots for both the reduced model and the very reduced model indicates that about 98% of the variance is encompassed using only two to three PCs. Inclusion of more than the minimum number of PCs than necessary will add more noise into the model and the results, as the PCs capturing the most variance are those that describe the information contained in the signal; the PCs capturing the least amount of variance contain noise and thus are not used in the model. There is still noise embedded in the signal that cannot be separated from the signal, however. The use of a smaller number of PCs than required causes more information to be lost and can adversely affect the results. The models assume three changing entities: the concentration of Epi, the concentration of Dopa, and the potential. Thus, a minimum of three PCs are required to build each model and subsequent predictions. This source of noise found in the third PC contributes to the difficulties in the deconvolution, especially in the very reduced data set.

Another possible explanation for the failure of the chemometric methods is that these methods assume a linear inner dimension. This arises from the requirement that the responses found both in the time vs. signal plane and the potential vs. signal plane form homogeneous systems of linear equations. This means that the signal from one electrode

cannot be a multiple of another electrode. This is not normally expected in the data. However, there is a high degree of correlation between the responses of the analytes at nearby times, and thus the problem is rather ill-conditioned. The noise and experimental error found in nearby data points are expected to be similar, and thus contributing to the problem. However, the noise and experimental error for a given analyte is expected to be at their lowest at the top of the analyte peak, exactly where the highest degree of correlation are expected to occur. In addition, the vector of "inner-relationship" coefficients which relate the scores from the independent variables and the scores from the dependent variables is expected to be linear. This will not be the case when the system itself is highly nonlinear. If the same value found in the independent variables yield multiple values in the dependent variables, these methods can fail. Thus, it appears that the signal values originally input into the model still vary too much, and thus the method fails.

The Cramer's Rule method also can yield correct answers to 95% confidence, but only at carefully selected points on the chromatogram. This method is a precursor to the chemometric methods also in this work. The chemometric methods make use of Cramer's rule in determining the models and hence the results, but the chemometric methods also have inherent noise smoothing characteristics that the Cramer's Rule method presented herein does not. More importantly, the use of Cramers Rule in this work assumes there is one constant that can be obtained regarding the response of the pure analyte at a given electrode, per unit time. The PCA/PCR/PLS/SIMPLS models above assume this constant does not change per unit time. This is not the case using

Cramer's Rule. Thus, Cramer's Rule has a greater chance of success where the chemometric methods failed.

Can the chemometric methods be used to deconvolute chromatographic data using electrochemical detection? The answer is yes, but with caution. More work needs to be done to determine why specifically the system behaves as it does. This includes investigation as to whether small changes in the dependent variable matrix has a discernable effect on the outcomes. While this investigation may be useful as a diagnostic, it is not expected to allow complete deconvolution to occur, as this implies a high degree of instability in the algorithms.

A second avenue of investigation involves the nature of the adsorption occurring at the electrode surface. Pretreatment of all four electrodes with multiple injections of Dopa prior to an HDV analysis of Epi would give a more realistic picture of the true chemical system involved in this study. The HDV used in this work was done using the first electrode, as the others were assumed to act similarly. Comparison of the pretreated HDVs to untreated HDVs then will give more complete information concerning the nature of the shift in the electrochemical properties of the electrode surface. Any bias introduced because the electrode surface was unstable in regard to building of the surface monolayer would then be eliminated. As it is unknown which one of the three compounds is responsible for the effects seen herein, the experiment needs to be repeated. A comparison of the amounts of analyte used in the pretreatment would also be useful; the models built in this work indicated that the concentrations used were also a factor in the failure of the models. The methods can be applied to other binary or ternary

mixtures only with additional investigations regarding the nature of the interactions of the specific compounds on the electrode surface.

References

1. Ettre, L. S., *M.S. Tswett and the invention of chromatography*, LCGC North America **2003**, *21*, 458-467.
2. Papoff, P.; Ceccarini, A.; Lanza, F.; Fanelli, N., *Enhancing the quality of information obtained by a comparison between experimental and deconvolved peak parameters in ion chromatography*, Journal of Chromatography A **1997**, *789*, 51-65.
3. Brooks, S. H.; Dorsey, J. G., *Moment analysis for evaluation of flow-injection analysis*, Analytica Chimica Acta **1990**, *229*, 35-46.
4. Schumauch, L. J., *Response time and flow sensitivity of detectors for gas chromatography*, Analytical Chemistry **1959**, *31*, 225-230.
5. Grall, A.; Leonard, C.; Sacks, R., *Peak capacity, peak-capacity production rate, and boiling point resolution for temperature-programmed GC with very high programming rates*, Analytical Chemistry **2000**, *72*, 591-598.
6. Roston, D. A.; Kissinger, P. T., *Series dual-electrode electrode for liquid chromatography/electrochemistry*, Analytical Chemistry **1982**, *54*, 429-434.
7. Bard, A. J.; Faulkner, L. R. *Electrochemical Methods: Fundamentals and Applications*, 2 ed.; John Wiley & Sons, Inc: New York, 2001.
8. Geiger, W. E. In *Laboratory Techniques in Electroanalytical Chemistry,* 2 ed.; Kissinger, P. T., Heineman, W. R., Eds.; Marcel Dekker, Inc.: New York, 1996, pp 683-717.

9. Newman, J. In *Electroanalytical Chemistry, A Series of Advances, Vol. 6,* 1 ed.; Bard, A. J., Ed.; Marcel Dekker, Inc.: New York, 1973, pp 187-352.

10. Elderfield, A. J.; Truscott, R. J. W.; Gan, I. E. T.; Schier, G. M., *Separation of tryptophan metabolites by reversed-phase high-performance liquid chromatography with amperometric and fluorescence detection*, Journal of Chromatography A **1989**, *495*, 71-80.

11. Kissinger, P. T.; Preddy, C. R.; Shoup, R. E.; Heineman, W. R. In *Laboratory Techniques in Electroanalytical Chemistry,* 2 ed.; Kissinger, P. T., Heineman, W. R., Eds.; Marcel Dekker, Inc.: New York, 1996, pp 11-50.

12. Roosendaal, E. M. M.; Poppe, H., *Diffusion-limited mass transfer in thin-layer (flow-through) electrochemical detectors with single and dual electrodes*, Analytica Chimica Acta **1984**, *158*, 323-333.

13. Lunte, C. E.; Ridgway, T. H.; Heineman, W. R., *Voltammetric-amperometric dual-electrode detection for flow injection analysis and liquid chromatography*, Analytical Chemistry **1987**, *59*, 761-766.

14. LaCourse, W. R. In *Techniques in Analytical Chemistry: Pulsed Electrochemical Detection in High-Performance Liquid Chromatography,* 1 ed.; John Wiley & Sons, Inc.: New York, 1997, pp 60-85.

15. Gunasingham, H.; Fleet, B. In *Electroanalytical Chemistry: A Series of Advances, Vol. 16,* 1 ed.; Bard, A. J., Ed.; Marcel Dekker, Inc.: New York, 1989, pp 89-180.

16. Stulík, K.; Pacáková, V. In *Electroanalytical Measurements in Flowing Liquids,* 1 ed.; Ellis Horewood Ltd.: Chichester, 1987, pp 82-161.

17. Lunte, S. M.; Lunte, C. E.; Kissinger, P. T. In *Laboratory Techniques in Electroanalytical Chemistry,* 2 ed.; Kissinger, P. T., Heineman, W. R., Eds.; Marcel Dekker, Inc.: New York, 1996, pp 813-854.

18. Skoog, D. A.; Leary, J. J. In *Principles of Instrumental Analysis,* 4th ed.; Saunders College Publishing: Fort Worth, TX, 1992, pp 600.

19. Chesler, S. N.; Cram, S. P., *Iterative curve fitting of chromatographic peaks*, Analytical Chemistry **1973**, *45*, 1354-1359.

20. Gong, F.; Liang, Y.-Z.; Xu, Q.-S.; Chau, F.-T., *Gas chromatography-mass spectrometry and chemometric resolution applied to the determination of essential oils in Cortex Cinnamomi*, Journal of Chromatography A **2001**, *905*, 193-205.

21. Service, K. M.; Brereton, R. G.; Harris, S., *Analysis of badger urine volatiles using gas chromatography-mass spectrometry and pattern recognition techniques*, Analyst **2001**, *126*, 615-623.

22. Demir, C.; Hindemarch, P.; Brereton, R. G., *Deconvolution of a three-component co-eluting peak cluster in gas chromatography-mass spectrometry*, Analyst **2000**, *125*, 287-292.

23. Ruiz-Calero, V.; Saurina, J.; Galceran, M. T.; Hernández-Cassou, S.; Puignou, L., *Potentiality of proton nuclear magnetic resonance and multivariate calibration methods for the determination of dermatan sulfate contamination in heparin samples*, Analyst **2000**, *125*, 933-938.

24. Wiberg, K.; Hagman, A.; Burén, P.; Jacobsson, S. P., *Determination of the content and identity of lidocaine solutions with UV-visible spectroscopy and multivariate calibration*, Analyst **2001**, *126*, 1142-1148.

25. Walmsley, A. D.; Loades, V. C., *Determination of acetonitrile and ethanol in water by guided microwave spectroscopy with multivariate calibration*, Analyst **2001**, *126*, 417-420.

26. Sánchez, F. C.; Vandeginste, B. G. M.; Hancewicz, T. M.; Massart, D. L., *Resolution of complex liquid chromatography-fourier transform infrared spectroscopy data*, Analytical Chemistry **1997**, *69*, 1477-1484.

27. Sentellas, S.; Saurina, J.; Hernández-Cassou, S.; Galceran, M. T.; Puignou, L., *Multivariate calibration methods for quantification in strongly overlapping capillary electrophoretic peaks*, Journal of Chromatography A **2001**, *909*, 259-269.

28. Ortiz-Estarelles, O.; Martin-Biosca, Y.; Medina-Hernánez, M. J.; Sagrado, S.; Bonet-Domingo, E., *Multivariate data analysis of quality parameters in drinking water*, Analyst **2001**, *126*, 91-96.

29. Domingo, C.; García-Carmona, J.; Fanovich, M. A.; Saurina, J., *Application of chemometric techniques to the characterization of impregnated materials obtained following supercritical fluid technology*, Analyst **2001**, *126*, 1792-1796.

30. Mattioni, B. E.; Jurs, P. C., *Development of quantitative structure-activity relationship and classification models for a set of carbonic anhydrase inhibitors*, Journal of Chemical Information and Computational Science **2002**, *42*, 94-102.

31. Peña, R. M.; García, S.; Iglesias, R.; Barro, S.; Herrero, C., *Authentication of Galician (N.W. Spain) quality brand potatoes using metal analysis. Classical pattern recognition techniques versus a new vector quantization-based classification procedure*, Analyst **2001**, *126*, 2186-2193.

32. Padin, P. M.; Peña, R. M.; García, S.; Iglesias, R.; Herrero, C., *Characterization of Galician (N.W. Spain) quality brand potatoes: a comparison study of several pattern recognition techniques*, Analyst **2001**, *126*, 97-103.

33. Latorre, M. J.; Peña, R.; García, S.; Herrero, C., *Authentication of Galician (N.W. Spain) honeys by multivariate techniques based on metal content data*, Analyst **2000**, *125*, 307-312.

34. Mello, C.; Poppi, R. J.; de Andrade, J. C.; Cantarella, H., *Pruning neural network architecture optimization applied to near-infrared spectroscopic measurements. Determination of the nitrogen content in wheat leaves*, Analyst **1999**, *124*, 1669-1674.

35. Palma, M.; Barroso, C. G.; Pérez-Bustamante, J. A., *Evaluation of similarities between natural and accelerated browning of fino sherry wines by chemometric techniques*, Analyst **2000**, *125*, 1151-1154.

36. Myors, R. B.; Crisp, P. T.; Skopec, S. V.; Wells, R. J., *Investigation of heroin profiling using trace organic impurities*, Analyst **2001**, *126*, 679-689.

37. Welsh, W. J.; Lin, W.; Tersigni, S. H.; Collantes, E.; Duta, R.; Carey, M. S., *Pharmaceutical fingerprinting: evaluation of neural networks and chemometric techniques for distinguishing among same-product manufacturers*, Analytical Chemistry **1996**, *68*, 3473-3482.

38. Hernández-Arteseros, J. A.; Compaño, R.; Ferrer, R.; Prat, M. D., *Application of principal component regression to luminescence data for the screening of ciprofloxacin and enrofloxacin in animal tissues*, Analyst **2000**, *125*, 1155-1158.

39. Dreassi, E.; Corti, P.; Bezzini, F.; Furlanetto, S., *High-performance liquid chromatographic assay of erythromycin from biological matrix using electrochemical or ultraviolet detection*, Analyst **2000**, *125*, 1077-1081.

40. Lunte, C. E.; Kissinger, P. T.; Shoup, R. E., *Difference mode detection with thin-layer dual-electrode liquid chromatography/electrochemistry*, Analytical Chemistry **1985**, *57*, 1541-1546.

41. Koivu, T. J.; Piironen, V. I.; Henttonen, S. K.; Mattila, P. H., *Determination of phylloquinone in vegetables, fruits, and berries by high-performance liquid chromatography with electrochemical detection*, Journal of Agriculture and Food Chemistry **1997**, *45*, 4644-4649.

42. Selavka, C. M.; Krull, I. S., *Liquid chromatography-photolysis-electrochemical detection for organoiodides. 1. Optimization and application*, Analytical Chemistry **1987**, *59*, 2699-2703.

43. Zhong, M.; Zhou, J.; Lunte, S. M.; Zhao, G.; Giolando, D. M.; Kirchhof, J. R., *Dual-electrode detection for capillary electrophoresis/electrochemistry*, Analytical Chemistry **1996**, *68*, 203-207.

44. Cabanillas, A. G.; Díaz, T. G.; Díaz, N. M. M.; Saiinas, F.; Burguilios, J. M. O.; Viré, J.-C., *Resolution by polarographic techniques of atrizine-simazine and terbutryn-prometryn binary mixtures by using PLS calibration and artificial neural networks*, Analyst **2000**, *125*, 909-914.

45. Ubibe, C.; Lizarreta, J.; Grau, J.; López-Cuerto, G., *Multivariate analysis of reagent profiles for slow reactions using PLS analysis*, Analyst **2000**, *125*, 1709-1712.

46. Díaz-Cruz, J. M.; Agulló, J.; Díaz-Cruz, M. S.; Ariño, C.; Esteban, M.; Tauler, R., *Implementation of a chemical equilibrium constraint in the multivariate curve resolution of voltammograms from systems with excessive metal complexes*, Analyst **2001**, *126*, 371-377.

47. Gladney, H. M.; Dowden, B. F.; Swalen, J. D., *Computer-assisted gas-liquid chromatography*, Analytical Chemistry **1969**, *41*, 883-888.

48. Pápai, Z.; Pap, T. L., *Determination of chromatographic peak parameters by non-linear curve fitting using statistical moments*, Analyst **2002**, *127*, 494-498.

49. De Weljer, A. P.; Lucasius, C. B.; Buydens, L.; Kateman, G., *Curve fitting using natural computation*, Analytical Chemistry **1994**, *66*, 23-31.

50. Lundeen, J. T.; Juvet, J., R.S., *Quantitative resolution of severely overlapping chromatographic peaks*, Analytical Chemistry **1981**, *53*, 1369-1372.

51. Anderson, A. H.; Gibb, T. C.; Littlewood, A. B., *Computer analysis of unresolved non-Gaussian gas chromatograms by curve-fitting*, Analytical Chemistry **1970**, *42*, 434-440.

52. Vaidya, R. A.; Hester, R. D., *Deconvolution of overlapping chromatographic peaks using constrained non-linear optimization*, Journal of Chromatography **1984**, *287*, 231-244.

53. Marquardt, D. W., *An algorithm for least-squares estimation of nonlinear parameters*, Journal of the Society for Industrial and Applied Mathematics **1963**, *11*, 431-441.

54. Goodman, K. J.; Brenna, J. T., *Curve fitting for restoration of accuracy for overlapping peaks in gas chromatography/combustion isotope ratio mass spectrometry*, Analytical Chemistry **1994**, *66*, 1294-1301.

55. Reh, E., *Peak-shape analysis of unresolved peaks in chromatography: comparison of algorithms*, Trends in Analytical Chemistry **1995**, *14*, 1-5.

56. Arena, J. V.; Mazzarella, C. R.; Gluodenis, R. J., *Software-based deconvolution of gas chromatograms: An experiment for the instrumental analysis laboratory*, Journal of Chemical Education **1994**, *71*, 483-488.

57. Felinger, A.; Pap, T.; Inczédy, J., *Curve fitting to asymmetrical chromatograms by the extended Kalman filter in frequency domain*, Talanta **1994**, *41*, 1119-1126.

58. Phillips, M. L.; White, R. L., *Dependence of chromatogram peak areas obtained by curve-fitting on the choice of peak shape function*, Journal of Chromatographic Science **1997**, *35*, 75-81.

59. Pitha, J.; Jones, R. N., *An evaluation of mathematical functions to fit infrared band envelopes*, Canadian Journal of Chemistry **1967**, *45*, 2347-2352.

60. Youn, D. Y.; Yun, S. J.; Jung, K.-H., *Improved algorithm for resolution of overlapped asymmetric chromatographic peaks*, Journal of Chromatography **1992**, *591*, 19-29.

61. Grushka, E., *Characterization of exponentially modified gaussian peaks in chromatography*, Analytical Chemistry **1972**, *44*, 1733-1738.

62. Barber, W. E.; Carr, P. W., *Graphical method for obtaining retention time and number of theoretical plates from tailed chromatographic peaks*, Analytical Chemistry **1981**, *53*, 1939-1942.

63. Roberts, S. M.; Wilkinson, D. H.; Walker, L. R., *Practical least squares approximation of chromatograms*, Analytical Chemistry **1970**, *42*, 886-893.

64. Arena, J. V.; Leu, T. M., *Deconvolution of gas chromatograms with Excel*, Journal of Chemical Education **1999**, *76*, 867.

65. Davis, J. M.; Giddings, J. C., *Origin and characterization of departures from the statistical model of component-peak overlap in chromatography*, Journal of Chromatography **1984**, *289*, 277-298.

66. Jeansonne, M. S.; Foley, J. P., *Improved equations for the calculation of chromatographic figures of merit for ideal and skewed chromatographic peaks*, Journal of Chromatography **1992**, *594*, 1-8.

67. Grimalt, J.; Iturriaga, H.; Tomas, X., *The resolution of chromatograms with overlapping peaks by means of different statistical functions*, Analytica Chimica Acta **1982**, *139*, 155-166.

68. Pirogov, A. V.; Obrezkov, O. N.; Shpigun, O. A., *"Chromatogram generator" chromatogram modelling software*, Journal of Chromatography A **1995**, *706*, 31-36.

69. Yau, W. W., *Characterizing skewed chromatographic band broadening*, Analytical Chemistry **1977**, *49*, 395-398.

70. Pauls, R. E.; Rogers, L. B., *Band broadening studies using parameters for an exponentially modified gaussian*, Analytical Chemistry **1977**, *49*, 625-628.

71. Glajch, J. L.; Warren, D. C.; Kaiser, M. A.; Rogers, L. B., *Effects of operating variables on peak shape in gel permeation chromatography*, Analytical Chemistry **1978**, *50*, 1962-1967.

72. Lochmüller, C. H.; Sumner, M., *Estimation of extra-column dead volume effects using a mixing cell model*, Journal of Chromatographic Science **1980**, *18*, 159-165.

73. Felinger, A.; Pasti, L.; Dondi, F., *Fourier analysis of multicomponent chromatograms. Theory and models*, Analytical Chemistry **1990**, *62*, 1846-1853.

74. Felinger, A.; Pasti, L.; Reschiglian, P.; Dondi, F., *Fourier analysis of multicomponent chromatograms. Numerical evaluation of statistical parameters*, Analytical Chemistry **1990**, *62*, 1854-1860.

75. Berthod, A., *Mathematical series for signal modeling using exponentially modified functions*, Analytical Chemistry **1991**, *63*, 1879-1884.

76. Jeansonne, M. S.; Foley, J. P., *Review of the exponentially modified gaussian (EMG) function since 1983*, Journal of Chromatographic Science **1991**, *29*, 258-266.

77. Yau, W. W.; Kirkland, J. J., *Improved computer algorithm for characterizing skewed chromatographic band broadening. I. Method*, Journal of Chromatography **1991**, *556*, 111-118.

78. Yau, W. W.; Rementer, S. W.; Boyajian, J. M.; DeStefano, J. J.; Graff, J. F.; Lim, K. B.; Kirkland, J. J., *Improved computer algorithm for characterizing skewed chromatographic band broadening. II. Results and comparisons*, Journal of Chromatography **1993**, *630*, 69-77.

79. Felinger, A., *Deconvolution of overlapping skewed peaks*, Analytical Chemistry **1994**, *66*, 3066-3072.

80. Morton, D. W.; Young, C. L., *Analysis of peak profiles using statistical moments*, Journal of Chromatographic Science **1995**, *33*, 514-524.

81. Olivé, J.; Grimalt, J. O., *Relationships between different chromatographic peak description functions and numerical solutions of the mass balance equation*, Journal of Chromatographic Science **1995**, *33*, 194-203.

82. Le Vent, S., *Simulation of chromatographic peaks by simple functions*, Analytica Chimica Acta **1995**, *312*, 263-270.

83. Di Marco, V. B.; Bombi, G. G., *Mathematical functions for the representation of chromatographic peaks*, Journal of Chromatography A **2001**, *931*, 1-30.

84. Li, J., *A simplified exponentially modified gaussian function for modeling chromatographic peaks*, Journal of Chromatographic Science **1995**, *33*, 568-572.

85. Fraser, R. D. B.; Suzuki, E., *Resolution of overlapping bands: functions for simulating band shapes*, Analytical Chemistry **1969**, *41*, 37-39.

86. Pitha, J.; Jones, R. N., *A comparison of optimization methods for fitting curves to infrared band envelopes*, Canadian Journal of Chemistry **1966**, *44*, 3031-3050.

87. Olivé, J.; Grimalt, J. O., *Log-normal derived equations for the characterization of on-line acquired chromatographic peaks*, Journal of Chromatographic Science **1991**, *29*, 70-77.

88. Mott, S. D.; Grushka, E., *Chromatographic peak shape II. The use of shape models for the generation of moments*, Journal of Chromatography **1978**, *148*, 305-320.

89. Foley, J. P.; Dorsey, J. G., *Equations for calculation of chromatographic figures of merit for ideal and skewed peaks*, Analytical Chemistry **1983**, *55*, 730-737.

90. Delly, R., *Modifying the gaussian peak shape with more than one time constant*, Analytical Chemistry **1986**, *58*, 2344-2346.

91. Kolev, S. D., *Mathematical modelling of flow-injection systems*, Analytica Chimica Acta **1995**, *308*, 36-66.

92. Lan, K.; Jorgenson, J. W., *A hybrid of exponential and gaussian functions as a simple model of asymmetric chromatographic peaks*, Journal of Chromatography A **2001**, *915*, 1-13.

93. McDowell, L. M.; Barber, W. E.; Carr, P. W., *Effect of detector nonlinearity on the height, area, width and moments of peaks in liquid chromatography with absorbance detectors*, Analytical Chemistry **1981**, *53*, 1373-1376.

94. Haarhoff, P. C.; van der Linde, H. J., *Concentration dependence of elution curves in non-ideal gas chromatography*, Analytical Chemistry **1966**, *38*, 573-582.

95. Golshan-Shirazi, S.; Guiochon, G., *Effect of the difference of the column saturation capacities for the two components of a mixture on the relative intensities of the displacement and tag-along effects in nonlinear chromatography*, Analytical Chemistry **1990**, *62*, 217-220.

96. Grushka, E.; Myers, M. N.; Giddings, J. C., *Moments analysis for the discernment of overlapping chromatographic peaks*, Analytical Chemistry **1970**, *42*, 21-26.

97. Reh, E., *An algorithm for peak-shape analysis for differentiating unresolved peaks in chromatography*, Trends in Analytical Chemistry **1993**, *12*, 192-194.

98. Grimalt, J.; Iturriaga, H.; Olive, J., *An experimental study of the efficiency of different statistical functions for the resolution of chromatograms with overlapping peaks*, Analytica Chimica Acta **1987**, *201*, 193-205.

99. Olivé, J.; Grimalt, J. O.; Iturriaga, H., *Resolution of overlapping peaks in gas and liquid chromatography. Evaluation of different statistical and empirical functions*, Analytica Chimica Acta **1989**, *219*, 257-272.

100. Grimalt, J. O.; Olivé, J., *Log-normal derived equations for the determination of chromatographic peak parameters from graphical measurements*, Analytica Chimica Acta **1991**, *248*, 59-70.

101. Olivé, J.; Manuel, A.; Rodriguez, P.; Grimalt, J. O., *A low cost, open and flexible on-line data acquisition system for instrumental chromatography*, Analytica Chimica Acta **1995**, *306*, 285-295.

102. Dondi, F.; Betti, A.; Blo, G.; Bighi, C., *Statistical analysis of gas chromatographic peaks by the Gram-Charlier series of type A and the Edgeworth-Cramér series*, Analytical chemistry **1981**, *53*, 496-504.

103. Dondi, F., *Approximation properties of the Edgeworth-Cramer series and determination of peak parameters of chromatographic peaks*, Analytical Chemistry **1982**, *54*, 473-477.

104. Buys, T. S.; de Clerk, K., *Bi-gaussian fitting of skewed peaks*, Analytical Chemistry **1972**, *44*, 1273-1275.

105. Li, J., *Development and evaluation of flexible empirical peak functions for processing chromatographic peaks*, Analytical Chemistry **1997**, *69*, 4452-4462.

106. Grushka, E.; Mott, S. D., *Chromatographic peak shape III. Influence of kinetics parameters on solute profiles*, Journal of Chromatography **1979**, *186*, 117-127.

107. Jin, D.; Pardue, H. L., *Algorithms for time-dependent chromatographic peak areas: 1. Algorithms evaluated for fully resolved peaks*, Analytica Chimica Acta **2000**, *422*, 1-10.

108. Jin, D.; Pardue, H. L., *Algorithms for time-dependent chromatographic peak areas II. Resolution of overlapped chromatograms for two- and three-component samples*, Analytica Chimica Acta **2000**, *422*, 11-20.

109. Zuloaga, O.; Etxebarria, N.; Zubiaur, J.; Fernández, L. A.; Madariaga, J. M., *MultiSimplex optimization of purge-and-trap extraction of volatile organic compounds in soil samples*, Analyst **2000**, *125*, 477-480.

110. Vaid, T. P.; Burl, M. C.; Lewis, N. S., *Comparison of the performance of different discriminant algorithms in analyte discrimination tasks using an array of carbon black-polymer composite vapor detectors*, Analytical Chemistry **2001**, *73*, 321-331.

111. Wise, B. M.; Gallagher, N. B. *PLS_Toolbox 2.1*; Eigenvector Research, Inc.: Manson, WA, 2000.

112. Hernández, O.; Martín, E.; Jiménez, F.; Jiménez, A. I.; Arias, J. J., *Use of partial least-squares regression for multicomponent determinations based on kinetic spectrofluorimetric data. Simultaneous determination of canrenone and spironolactone in urine*, Analyst **2000**, *125*, 1159-1165.

113. del Olmo, M.; Zafra, A.; Vilchez, J. L., *Trace determination of phenol, bisphenol A and bisphenol A diglycidyl ether in mixtures by excitation fluorescence*

following micro liquid-liquid extraction using partial least squares regression, Analyst **1999**, *124*, 385-390.

114. Coscione, A. R.; de Andrade, J. C.; Poppi, R. J., *PLS regression using real sample calibration for aluminum and iron determination in plant extracts*, Analyst **2002**, *127*, 135-139.

115. Sales, F.; Callao, M. P.; Rius, F. X., *Multivariate standardization techniques on ion-selective sensor arrays*, Analyst **1999**, *124*, 1045-1051.

116. Saurina, J.; Hernández-Cassou, S.; Fâbregas, E.; Alegret, S., *Determination of tryptophan in feed samples by cyclic voltammetry and multivariate calibration methods*, Analyst **1999**, *124*, 733-737.

117. Domingo, C.; Arcis, R. W.; Osorio, E.; Toledano, M.; Saurina, J., *Principal component analysis and cluster analysis for the characterization of dental composites*, Analyst **2000**, *125*, 2044-2048.

118. Macho, S.; Boqué, R.; Larrechi, M. S.; Ruis, F. X., *Multivariate determination of several compositional parameters related to the content of hydrocarbon in naphtha by MIR spectroscopy*, Analyst **1999**, *124*, 1827-1831.

119. Cao, G.; Hoshino, T., *Analysis of overlapped chromatographic peaks by multi-channel electrochemical detection*, Journal of Chromatography A **1996**, *722*, 151-156.

120. Cao, G.-M.; Hoshino, T., *Simultaneous determination of 3,4-dihydroxymandelic acid, 4-hydroxy-3-methoxymandelic acid, 3,4-dihydroxyphenylglycol, 4-hydroxy-3-methoxyphenylglycol, and their precursors, in human urine by HPLC with electrochemical detection*, Chromatographia **1998**, *47*, 396-400.

121. Turk, D. J. *Development and Utilization of a Multi-Column, Multi-Electrode Detector for the Simultaneous Determination of Multiple Neurochemical Species*, 1 ed.; University of Oklahoma: Norman, 1991.

122. Ikarashi, Y.; Itoh, K.; Satoh, H.; Blank, C. L.; Way, E. L.; Maruyama, Y. In *Basic, Clinical, and Therapeutic Aspects of Alzheimer's and Parkinson's Diseases, Vol. 1*; Nagatsu, T., Fisher, A., Yoshida, M., Eds.; Plenum Press: New York, 1990, pp 525-528.

123. Satoh, H.; Yumakura, K.; Miller, R. B.; Turk, D. J.; Howard, S.; Maruyama, Y.; Blank, C. L. In *Basic, Clinical, and Therapeutic Aspects of Alzheimer's and Parkinson's Diseases, Vol. 1*; Nagatsu, T., Fisher, A., Yoshida, M., Eds.; Plenum Press: New York, 1990, pp 521-524.

124. Turk, D. J.; Blank, C. L. In *Basic, Clinical, and Therapeutic Aspects of Alzheimer's and Parkinson's Diseases, Vol. 1*; Nagatsu, T., Fisher, A., Yoshida, M., Eds.; Plenum Press: New York, 1990, pp 517-520.

125. Ikarashi, Y.; Blank, C. L.; Itoh, K.; Satoh, H.; Inoue, H. K.; Maruyama, Y., *Development of a liquid chromatograph/multiple electrochemical detector (LCMC) and its application in neuroscience*, Folia Pharmacologica Japonica **1991**, *97*, 51-64.

126. Ikarashi, Y.; Blank, C. L.; Inoue, H. K.; Maruyama, Y., *Measurement of neurochemical levels in ventricular cerebrospinal fluid of Parkinson's patients using multicolumn, multielectrode liquid chromatography with electrochemical detection*, Biogenic Amines **1992**, *8*, 175-190.

127. Achilli, G.; Cellerino, G. P.; Gamache, P. H., *Identification and determination of phenolic constituents in natural beverages and plant extracts by means of a coulometric electrode array system*, Journal of Chromatography A **1993**, *632*, 111-117.

128. Owens, A. L.; Denison, T. J.; Versnel, H.; Rebbert, M.; Peckerar, M.; Shamma, S. A., *Multi-electrode array for measuring evoked potentials from surface of ferret primary auditory cortex*, Journal of Neuroscience Methods **1995**, *58*, 209-220.

129. Hoogvliet, J. C.; Reijn, J. M.; van Bennekom, W. P., *Multichannel amperometric detection system for liquid chromatography and flow injection analysis*, Analytical Chemistry **1991**, *63*, 2418-2423.

130. Wang, C.-L.; Creasy, K. E.; Shaw, B. R., *Ring-modified carbon fiber microelectrodes and multi-microelectrode devices*, Journal of Electroanalytical Chemistry **1991**, *s300*, 365-375.

131. Takeba, K.; Matsumoto, M.; Nakazawa, H., *Determination of nitroxynil in cow milk by reversed-phase high-performance liquid chromatography with dual-electrode coulometric detection*, Journal of Chromatography A **1992**, *596*, 67-71.

132. Liu, Z.; Li, T.; Li, J.; Wang, E., *Detection of menadione sodium bisulfite (vitamin K_3) by reversed-phase high performance liquid chromatography with series dual-electrode amperometric detector*, Analytica Chimica Acta **1997**, *338*, 57-62.

133. Fung, Y.-S.; Mo, S.-Y., *Determination of amino acids and proteins by dual-electrode detection in a flow system*, Analytical Chemistry **1995**, *67*, 1121-1124.

134. Achilli, G.; Cellerino, G. P.; d'Eril, G. M.; Bird, S., *Simultaneous determination of 27 phenols and herbicides in water by high-performance liquid chromatography*

with multi-electrode electrochemical detection, Journal of Chromatography A **1995**, *697*, 357-362.

135. Gerhardt, G. A.; Drebing, C. J.; Stephen, C.; Freedman, R., *Direct determination of unconjugated HVA in human plasma filtrates by HPLC coupled with dual-electrode coulometric electrochemical detection*, Biomedical Chromatography **1989**, *3*, 105-109.

136. Hou, W.; Wang, E., *Liquid chromatography with series dual-electrode electrochemical detection for riboflavin*, Analyst **1990**, *115*, 139-141.

137. Dempsey, C. A.; Lavicky, J.; Dunn, A. J., *Short Communication: Apparent inter-channel interference in dual-electrode electrochemical detection*, Journal of Chromatography A **1992**, *596*, 110-113.

138. Fraga, C. G.; Prazen, B. J.; Synovec, R. E., *Objective data alignment and chemometric analysis of comprehensive two-dimensional separations with run-to-run peak shifting on both dimensions*, Analytical Chemistry **2001**, *73*, 5833-5840.

139. Fraga, C. G.; Prazen, B. J.; Synovec, R. E., *Comprehensive two-dimensional gas chromatography and chemometrics for the high-speed quantitative analysis of aromatic isomers in a jet fuel using the standard addition method and an objectie retention time alignment algorithm*, Analytical Chemistry **2000**, *72*, 4154-4162.

140. Walsh, S.; Diamond, D., *Non-linear curve fitting using Microsoft Excel Solver*, Talanta **1995**, *42*, 561-572.

141. DeWitt, C. Senior Thesis, University of Oklahoma, Norman, 1997.

142. McCreery, R. L. In *Neuromethods, Vol. 27: Voltammetric Methods in Brain Systems,* 1 ed.; Boulton, A., Baker, G., Adams, R. N., Eds.; Humana Press: New York, 1995, pp 1-26.

143. Bard, A. J.; Faulkner, L. R. *Electrochemical Methods: Fundamentals and Applications*, Second ed.; John Wiley & Sons, Inc.: New York, 2001.

144. Underwood, A. L.; Burnett, R. W. In *Electroanalytical Chemistry: A Series of Advances. Volume 6*; Bard, A. J., Ed.; Marcel Dekker, Inc.: New York, 1973, pp 1-85.

145. Malem, F.; Mandler, D., *Self-assembled monolayers in electroanalytical chemistry: application of w-mercapto carboxylic acid monolayers for the electrochemical detection of dopamine in the presence of a high concentration of ascorbic acid*, Analytical Chemistry **1993**, *a65*, 3f7-41.

146. Guo, G.; Anazi, J.-I.; Osa, T., *Modification of a glassy carbon electrode with diols for the suppression of electrode fouling in biological fluids*, Chemical & Pharmaceutical Bulletin **1996**, *44*, 860-862.

147. Rice, R. J.; Pontikos, N. M.; McCreery, R. L., *Quantitative correlations of heterogenous electron-transfer kinetics with surface properties of glassy carbon electrodes*, Journal of the American Chemical Society **1990**, *112*, 4617-4622.

148. Liu, Y.-C.; McCreery, R. L., *Reactions of organic monolayers on carbon surfaces observed with unenhanced Raman spectroscopy*, Journal of the American Chemical Society **1995**, *117*, 11254-11259.

149. Kepley, L. J.; Bard, A. J., *Ellipsometric, electrochemical, and elemental characterization of the surface phase produced on glassy carbon electrodes by electrochemical activation*, Analytical Chemistry **1988**, *60*, 1459-1467.

150. Bowling, R.; Packard, R. T.; McCreery, R. L., *Mechanism of electrochemical activation of carbon electrodes: Role of graphite lattice defects*, Langmuir **1989**, *5*, 683-688.

151. McDermott, M. T.; McCreery, R. L., *Scanning tunneling microscopy of ordered graphite and glassy carbon surfaces: Electronic control of quinone adsorption*, Langmuir **1994**, *10*, 4307-4314.

152. Fagan, D. T.; Hu, I.-F.; Kuwana, T., *Vacuum heat treatment for activation of glassy carbon electrodes*, Analytical Chemistry **1985**, *57*, 2759-2763.