INFORMATION TO USERS

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

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

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

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

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

CHEMOMETRIC ANALYSIS IN QUANTITATIVE NEAR-INFRARED SPECTROSCOPY

by

ZHIJUN JIANG

A dissertation submitted to the Graduate Faculty in Chemistry
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

The City University of New York

2002

UMI Number: 3063841

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

This manuscript has been read and accepted by the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

11 September 2002
Date

David C. Lorke
Chair of Examining Committee

9/19/2002
Date

Executive Officer

Prof. Ronald L. Birke

Prof. Nan-Loh Yang

The City University of New York

ABSTRACT

CHEMOMETRIC ANALYSIS IN QUANTITATIVE NEAR-INFRARED SPECTROSCOPY

ZHIJUN JIANG

Advisor: Professor David C. Locke

City University of New York

This study describes chemometric methods and their use in quantitative near-infrared (NIR) spectroscopy, and NIR instrumentation. The particular application studied was the NIR analysis of nicotine chewing gum containing nicotine at the 0.4% level using both transmittance and reflectance near-infrared spectrophotometers. The gum tablets were scanned in a sample holder specially designed for this study and a specific HPLC method was developed to determine the true potencies of the nicotine tablets after they were scanned on the NIR spectrophotometer. The calibration set was constructed from using the NIR and HPLC data, which was used to predict the nicotine content of unknown samples. The potencies of the unknown samples were further determined by the HPLC method and the HPLC results were compared to those obtained from NIR. Different calibration techniques, including simple linear regression, multiple linear regression, and partial least squares regression, were investigated and their effects were discussed. In addition, the techniques for detection of outliers and different mathematical sample pre-treatments were explored.

Acknowledgements

First I would like to thank my advisor, Professor David C. Locke, who generously advised me on a variety of academic issues as I was working on my Ph.D. degree. I am very proud of finishing my Ph.D. study under his supervision. His special guidance and patience are always appreciated.

Second I would like to thank the members of the thesis committee: Prof. Ronald L. Birke and Prof. Nan-Loh Yang. Their valuable guidance and suggestions made this thesis a success.

My next thanks goes out to Ms. Yin Luo for her special support during the dissertation preparation which motivated me in many ways and made everything interesting.

I would also like to thank Dr. Frederic Long who provided much help on the preparation of this thesis, Foss NIRSystems and Mr. Daniel A. Sanborn who allowed me to use their instrument and some of the training materials.

I greatly appreciate the outstanding support from GlaxoSmithKline Co. and Barr Laboratories, Inc., which made this research possible.

I am very grateful to my parents, brothers, and sisters for their love, encouragement, and support.

I am also grateful to Ms. Limin Zhang for her exceptional support that made the completion of this work a joy.

Finally, I also want to thank my six-year old son, James Jiang, who always shows a strong interest in science, which inspired me to complete this doctoral research.

Table of Contents

1. INTRODUCTION 1

2. NEAR-INFRARED THEORY 3

2.1 NEAR-INFRARED SPECTRUM 3
2.1.1 VIBRATIONAL TRANSITIONS 4
2.1.2 FUNDAMENTAL FREQUENCIES 6
2.1.3 ANHARMONICITY AND OVERTONES 6
2.1.4 COMBINATION BANDS 8

3. NEAR-INFRARED INSTRUMENTATION 10

3.1 TRANSMITTANCE NEAR-INFRARED 10
3.2 REFLECTANCE NEAR-INFRARED 11
3.3 COMBINATION OF TRANSMITTANCE AND REFLECTANCE NIR 12
3.4 SOURCE TYPE 13
3.5 DETECTOR TYPE 14
3.6 INSTRUMENT SPECIFICATIONS 14
3.6.1 WAVELENGTH RANGE 14
3.6.2 SPECTRAL BANDWIDTH 14
3.6.3 WAVELENGTH ACCURACY 15
3.6.4 PHOTOMETRIC NOISE 15
3.6.5 DYNAMIC RANGE 15
3.6.6 BASELINE SHIFT 15
3.6.7 STRAY LIGHT 16
3.6.8 SCAN RATE 16
3.6.9 DATA INTERVAL 16
3.7 INSTRUMENT QUALIFICATION 16

3.7.1 DESIGN QUALIFICATION (DQ) 17
3.7.2 INSTALLATION QUALIFICATION (IQ) 17
3.7.3 OPERATIONAL QUALIFICATION (OQ) 17
3.7.4 PERFORMANCE QUALIFICATION (PQ) 17
3.7.5 MAINTENANCE QUALIFICATION (MQ) 17

4. THEORY OF DIFFUSE REFLECTANCE 18

5. EXPERIMENTAL 19

5.1 NICOTINE AND GUM TABLET 19
5.2 NICOTINE TABLET PREPARATION AND COLLECTION 21
5.3 TRANSMITTANCE NIR 22
5.3.1 EQUIPMENT 22
5.3.2 INSTRUMENT PERFORMANCE CHECK 25
5.3.2.1 Photometric Noise 25
5.3.2.2 Wavelength Accuracy/Precision 26
5.3.3 SAMPLE PRESENTATION 26
5.3.4 SAMPLE ACQUISITION 27
5.4 REFLECTANCE NIR 28
5.4.1 EQUIPMENT 28
5.4.2 SAMPLES 30
5.4.3 INSTRUMENT PERFORMANCE CHECKS 31
5.4.4 SAMPLE PRESENTATION 32
5.4.5 SAMPLE ACQUISITION 32
5.5 HPLC ANALYSIS 32
5.5.1 EQUIPMENT 32
5.5.2 SAMPLES 33
5.5.3 REAGENTS 34
5.5.4 PREPARATION OF MOBILE PHASE 34
5.5.5 WATER CONTENT OF REFERENCE STANDARD 35
5.5.6 PREPARATION OF STANDARD SOLUTION 35
5.5.7 PREPARATION OF SAMPLE SOLUTION 35

5.5.8 CHROMATOGRAPHIC PARAMETERS 36
5.5.9 SYSTEM SUITABILITY 36
5.5.10 CALCULATION (MG/TABLET) 36

6. DATA ANALYSIS – HPLC 37

6.1 HPLC METHOD VALIDATION 37
6.2 RESULTS 41

7. DATA ANALYSIS – NIR SPECTRA 43

8. DATA ANALYSIS – QUANTITATIVE NEAR INFRARED 53

8.1 SIMPLE LINEAR REGRESSION (SLR) 54
8.2 MULTIPLE LINEAR REGRESSION (MLR) 55
8.3 PARTIAL LEAST SQUARES REGRESSION (PLSR) 56
8.3.1 PRINCIPAL COMPONENT ANALYSIS 57
8.3.2 PARTIAL LEAST SQUARES REGRESSION 60

9. DATA ANALYSIS – TRANSMITTANCE NIR 62

9.1 SAMPLE SELECTION 63
9.1.1 RANDOM SELECTION 64
9.1.2 MAHALANOBIS DISTANCE IN PRINCIPAL COMPONENT SPACE 65
9.1.3 MAXIMUM DISTANCE IN WAVELENGTH SPACE 67
9.1.4 DISCUSSION 68
9.2 PARTIAL LEAST SQUARES REGRESSION 69
9.2.1 EFFECT OF SAMPLE PRESENTATION 72
9.2.1.1 One-way Scan 72
9.2.1.2 Four-way Scan 74
9.2.1.3 Discussion 76
9.2.2 EFFECT OF WAVELENGTH RANGE 76
9.2.2.1 Range of 900 nm – 1300 nm 77

9.2.2.2	Range 850 nm – 1650 nm	83
9.2.2.3	Discussion	88
9.2.3	DIFFERENT MATHEMATICAL PRETREATMENTS	89
9.2.3.1	Raw Data	89
9.2.3.2	First Derivative	91
9.2.3.3	Second Derivative	93
9.2.3.4	Standard Normal Variate	94
9.2.3.5	Detrend	96
9.2.3.6	Savitzky-Golay	98
9.2.3.7	Multiplicative Scatter Correction	99
9.2.3.8	Discussion	101
9.2.4	SELECTION OF FACTOR NUMBERS	102
9.2.4.1	Calibration with Two, Seven, and Twelve Factors	102
9.2.4.2	Discussion	104
9.2.5	SELECTION OF SEGMENT SIZE	105
9.2.5.1	Segment Size = 10	105
9.2.5.2	Segment Size = 20	107
9.2.5.3	Segment Size = 30	108
9.2.5.4	Discussion	110
9.2.6	REGRESSION VALIDATION	110
9.2.7	PREDICTION OF UNKNOWN SAMPLES	113
9.2.8	DISCUSSION	115
9.3	**SIMPLE LINEAR REGRESSION**	**116**
9.3.1	REGRESSION DEVELOPMENT	116
9.3.2	REGRESSION VALIDATION	120
9.3.3	PREDICTION OF UNKNOWN SAMPLES	122
9.3.4	DISCUSSION	122
9.4	**MULTIPLE LINEAR REGRESSION**	**123**
9.4.1	REGRESSION DEVELOPMENT	123
9.4.2	REGRESSION VALIDATION	127
9.4.3	PREDICTION OF UNKNOWN SAMPLES	129
9.4.4	DISCUSSION	129
9.5	**COMPARISON OF SLR, MLR, AND PLSR**	**129**

10. DATA ANALYSIS – REFLECTANCE NIR **130**

10.1 SAMPLE PRESENTATION **131**
10.2 REGRESSION DEVELOPMENT **131**
10.2.1 RANGE 750 NM – 2250 NM 131
10.2.2 RANGE 750 NM – 1050 NM PLUS 1150 NM – 2250 NM 136
10.2.3 RANGE 900 NM – 1300 NM 139
10.3 REGRESSION VALIDATION **142**
10.4 PREDICTION OF UNKNOWN SAMPLES **144**
10.5 DISCUSSION **145**

11. COMPARISON OF TRANSMITTANCE AND REFLECTANCE NIR **146**

12. SUMMARY **147**

13. GLOSSARY **150**

14. APPENDIX **151**

15. REFERENCES **153**

List of Tables

Table 1. Basic Characteristic Wavelengths in NIR Region 9
Table 2. The Specifications of InTact® Analyzer 23
Table 3. Results of the Photometric Noise Test 25
Table 4. Results of the Wavelength Accuracy/Precision Test 26
Table 5. The Specifications of Rapid-Content® Analyzer 28
Table 6. Samples Analyzed in Reflectance NIR 31
Table 7. Equipment Used in HPLC Analysis 33
Table 8. Samples Run in HPLC Analysis 33
Table 9. Reagents Prepared and Used in HPLC Analysis 34
Table 10. Parameters Used in HPLC Analysis 36
Table 11. Results of the HPLC Method Validation 38
Table 12. HPLC Results for Tablets Scanned Using Transmittance NIR 42
Table 13. HPLC Results for Tablets Scanned Using Reflectance NIR 43
Table 14. Statistical Results, PLSR Calibration, One-way Scan 72
Table 15. NIR vs. HPLC Data, PLSR Calibration, One-way Scan 73
Table 16. Statistical Results, PLSR Calibration, Four-way Scan 74
Table 17. NIR vs. HPLC Data, PLSR Calibration, Four-way Scan 75
Table 18. F-test of One-way and Four-way Scans: Two-Sample for Variances 76
Table 19. Statistical Results, PLSR Calibration, 900 nm – 1300 nm 79
Table 20. NIR vs. HPLC Data, PLSR Calibration, 900 nm – 1300 nm 80
Table 21. Statistical Results, PLSR Calibration, 850 nm – 1650 nm 85
Table 22. NIR vs. HPLC Data, PLSR Calibration, 850 nm – 1650 nm 86
Table 23. Statistical Results, PLSR Calibration, Raw Data 90
Table 24. Statistical Results, PLSR Calibration, 1st Derivative 92
Table 25. Statistical Results, PLSR Calibration, 2nd Derivative 93
Table 26. Statistical Results, PLSR Calibration, SNV 95
Table 27. Statistical Results, PLSR Calibration, Detrend 97
Table 28. Statistical Results, PLSR Calibration, S-G 98
Table 29. Statistical Results, PLSR Calibration, MSC 100

Table 30. Statistical Results, PLSR Calibration, Factor = 2, 7, and 12 102
Table 31. Statistical Results, PLSR Calibration, s=10..... 106
Table 32. Statistical Results, PLSR Calibration, s=20..... 107
Table 33. Statistical Results, PLSR Calibration, s=30..... 109
Table 34. NIR vs. HPLC Data, PLSR Validation, s=20, Factor=7, 2nd Derivative.. 111
Table 35. NIR Predicted Results of Unknown Samples by Different Methods 114
Table 36. t-Test of Method 1 vs. HPLC Data for Comparison of Means* 115
Table 37. F-test of Method 1 vs. HPLC Data for Comparison of SD..... 115
Table 38. NIR vs. HPLC Data, SLR Calibration 118
Table 39. NIR vs. HPLC Data, SLR Validation 120
Table 40. NIR vs. HPLC Data, SLR Prediction..... 122
Table 41. NIR vs. HPLC Data, MLR Calibration..... 125
Table 42. NIR vs. HPLC Data, MLR Validation..... 127
Table 43. NIR vs. HPLC Data, MLR Prediction 129
Table 44. NIR Predicted Results of Unknowns by Different Regressions 130
Table 45. Statistical Results, R-PLSR Calibration, 750 nm – 2250 nm 133
Table 46. NIR vs. HPLC, R-PLSR Calibration, 750 nm – 2250 nm 134
Table 47. Statistical Results, 750 nm – 1050 nm & 1150 nm – 2250 nm 137
Table 48. NIR vs. HPLC Data, 750 nm – 1050 nm & 1150 nm – 2250 nm..... 138
Table 49. Statistical Results, R-PLSR Calibration, 900 nm – 1300 nm 139
Table 50. NIR vs. HPLC Data, R-PLSR Calibration, 900 nm – 1300 nm..... 140
Table 51. NIR vs. HPLC Data, R-PLSR Validation, 900 nm – 1300 nm..... 143
Table 52. Reflectance NIR Predicted Results (mg/tablet) 145
Table 53. Predicted Results of Transmittance and Reflectance NIR..... 146
Table 54. Critical Value of the Student t-test (two-tailed)..... 151
Table 55. Critical Values of F for a One-tailed Test (P = 0.05)..... 152

List of Figures

Figure 1. Spectral Regions Used in Different Techniques 4
Figure 2. Vibrational Transition, $\Delta\upsilon=1$ 7
Figure 3. The Harmonic and Anharmonic Potentials for a Diatomic Molecule. 7
Figure 4. Diagram of Basic Transmittance NIR Spectrophotometer 11
Figure 5. Diagram of Basic Reflectance NIR Spectrophotometer 12
Figure 6. Diagram of Basic NIR Spectrophotometer Using both Transmittance and Reflectance Detectors 13
Figure 7. Nicotine Degradation Pathways and Its Impurities 20
Figure 8. InTact® Analyzers Model AP-1600-II 23
Figure 9. InTact® Analyzer Transmittance Design 24
Figure 10. InTact® Analyzer Signal Detection 24
Figure 11. Sample Holder Specially Made for Nicotine Tablet 27
Figure 12. Rapid-Content ® Analyzers Model AP-1365-II 29
Figure 13. Detection of Rapid-Content® Analyzer 29
Figure 14. Rapid-Content® Analyzer Signal Detection 30
Figure 15. UV Spectrum of Nicotine in Mobile Phase 39
Figure 16. Linearity Plot of Nicotine Standard Solution at 260 nm 39
Figure 17. HPLC Chromatogram of Nicotine Standard Solution at 260 nm 40
Figure 18. HPLC Chromatogram of Selectivity Solution at 260 nm 40
Figure 19. Nicotine Peak Purity Plot of Selectivity Solution 41
Figure 20. R-NIR Spectrum of Nicotine Polacrilex (active) 44
Figure 21. R-NIR Spectrum of Gum Base (excipient) 44
Figure 22. R-NIR Spectrum of Sorbitol (excipient) 45
Figure 23. R-NIR Spectrum of Mannitol (excipient) 45
Figure 24. R-NIR Spectrum of Quiniline Yellow (excipient) 46
Figure 25. R-NIR Spectrum of Sodium Carbonate (excipient) 46
Figure 26. R-NIR Spectrum of Talc (excipient) 47
Figure 27. R-NIR Spectra of All Tablet Ingredients (7) 47

Figure 28. R-NIR Spectrum of Cotinine (impurity) 48
Figure 29. R-NIR Spectrum of Myosmine (impurity) 48
Figure 30. R-NIR Spectrum of Nornicotine (impurity) 49
Figure 31. R-NIR Spectrum of 1'R,2S-Nicotine-N-oxide (impurity) 49
Figure 32. R-NIR Spectrum of 1'R,2S-Nicotine-N'-oxide (impurity) 50
Figure 33. R-NIR Spectrum of Anatabine (impurity) 50
Figure 34. R-NIR Spectrum of Anabasine (impurity) 51
Figure 35. R-NIR Spectra of All Nicotine Impurities (7) 51
Figure 36. R-NIR Spectra of All Nicotine Gum Samples (48) 52
Figure 37. R-NIR Spectra of All Ingredients, Impurities, and Gum Samples 52
Figure 38. T-NIR Spectra of Tablets in Range of 600 nm – 1900 nm 62
Figure 39. T-NIR Spectra of all Tablets in Range of 900 nm – 1300 nm 63
Figure 40. The 2nd Derivative of T-NIR Spectra of all Tablets 64
Figure 41. Sample Clusters of Mahalanobis Distance in 3-Dimensions 66
Figure 42. Mahalanobis Distance vs. Frequencies of all Gum Samples 66
Figure 43. Maximum Distance in Wavelength Space of All Tablets 68
Figure 44. Loading Plot, PLSR Calibration, 900 nm – 1300 nm 77
Figure 45. Correlation Coefficient Plot, PLSR Calibration, 900 nm – 1300 nm 78
Figure 46. PRESS Plot, PLSR Calibration, 900 nm – 1300 nm 78
Figure 47. NIR vs. HPLC Plot, PLSR Calibration, 900 nm – 1300 nm 81
Figure 48. NIR Residual Plot, PLSR Calibration, 900 nm – 1300 nm 82
Figure 49. UCL and LCL, PLSR Calibration, 900 nm – 1300 nm 82
Figure 50. Loading Plot, PLSR Calibration, 850 nm – 1650 nm 83
Figure 51. Correlation Plot, PLSR Calibration, 850 nm – 1650 nm 84
Figure 52. PRESS Plot, PLSR Calibration, 850 nm – 1650 nm 84
Figure 53. NIR vs. HPLC Plot, PLSR Calibration, 850 nm – 1650 nm 87
Figure 54. NIR Residual Plot, PLSR Calibration, 850 nm – 1650 nm 87
Figure 55. UCL/LCL Plot, PLSR Calibration, 850 nm – 1650 nm 88
Figure 56. NIR vs. HPLC Plot, PLSR Calibration, Raw Data 90
Figure 57. NIR vs. HPLC Plot, PLSR Calibration, 1st Derivative 92
Figure 58. NIR vs. HPLC Plot, PLSR Calibration, 2nd Derivative 94

Figure 59. NIR vs. HPLC Plot, PLSR Calibration, SNV 96
Figure 60. NIR vs. HPLC Plot, PLSR Calibration, Detrend 97
Figure 61. NIR vs. HPLC Plot, PLSR Calibration, S-G 99
Figure 62. NIR vs. HPLC Plot, PLSR Calibration, MSC 101
Figure 63. NIR vs. HPLC Plot, PLSR Calibration, Factor=2 103
Figure 64. NIR vs. HPLC Plot, PLSR Calibration, Factor=7 103
Figure 65. NIR vs. HPLC Plot, PLSR Calibration, Factor=12 104
Figure 66. Loading Plot, PLSR Calibration, s=10 105
Figure 67. NIR vs. HPLC Plot, PLSR Calibration, s=10 106
Figure 68. Loading Plot, PLSR Calibration, s = 20 107
Figure 69. NIR vs. HPLC Plot, PLSR Calibration, s=20 108
Figure 70. Loading Plot, PLSR Calibration, s=30 109
Figure 71. NIR vs. HPLC Plot, PLSR Calibration, s=30 110
Figure 72. NIR vs. HPLC Plot, PLSR Validation, 2^{nd} Derivative 112
Figure 73. NIR Residual Plot, PLSR Validation, 2^{nd} derivative 112
Figure 74. Correlation Coefficient and Sensitivity Plot, SLR Calibration 117
Figure 75. NIR vs. HPLC Plot, SLR Calibration 119
Figure 76. NIR Residual Plot, SLR Calibration 119
Figure 77. NIR vs. HPLC Plot, SLR Validation 121
Figure 78. NIR Residual Plot, SLR Validation 121
Figure 79. Correlation & Sensitivity Plot, MLR Calibration, 1204 nm 123
Figure 80. Correlation and Sensitivity Plot, MLR Calibration, 1156 nm 124
Figure 81. Correlation and Sensitivity Plot, MLR Calibration, 1038 nm 124
Figure 82. NIR vs. HPLC Plot, MLR Calibration 126
Figure 83. NIR Residual Plot, MLR Calibration 126
Figure 84. NIR vs. HPLC Plot, MLR Validation 128
Figure 85. NIR Residuals Plot, MLR Validation 128
Figure 86. Loading Plot, R-PLSR Calibration, 750 nm – 2250 nm 132
Figure 87. Coefficient Plot R-PLSR Calibration, 750 nm – 2250 nm 132
Figure 88. PRESS Plot R-PLSR Calibration, 750 nm – 2250 nm 133
Figure 89. NIR vs. HPLC Plot, R-PLSR Calibration, 750nm – 2250 nm 135

Figure 90. NIR Residual Plot, R-PLSR Calibration, 750nm – 2250 nm 135
Figure 91. PRESS Plot, 750 nm – 1050 nm & 1150 nm – 2250 nm..... 136
Figure 92. NIR vs. HPLC Plot, 750 nm – 1050 nm & 1150 nm – 2250 nm..... 136
Figure 93. NIR Residual Plot, 750 nm – 1050 nm & 1150 nm – 2250 nm..... 137
Figure 94. PRESS Plot, R-PLSR Calibration, 900 nm – 1300 nm 141
Figure 95. NIR vs. HPLC Plot, R-PLSR Calibration, 900 nm – 1300 nm 141
Figure 96. NIR Residual Plot, R-PLSR Calibration, 900 nm – 1300 nm 142
Figure 97. NIR vs. HPLC Plot, R-PLSR Validation, 900 nm – 1300 nm..... 143
Figure 98. NIR Residuals Plot, R-PLSR Validation, 900 nm – 1300 nm..... 144

1. INTRODUCTION

In the field of analytical chemistry, there is always a demand for fast, easy, and economic methods. Accordingly, near-infrared (NIR) spectroscopy has become one of the most powerful techniques in analytical chemistry[1,2], especially when it is utilized for quantitative analysis.

F. E. Fowle[3] first used quantitative NIR to determine the moisture in the atmosphere at Mount Wilson observatory in 1912. In 1931, Kubelka and Munk[4] published their work on the diffuse scattering of light in both transmittance and reflectance modes that made the quantitative NIR for solids possible. In 1933, Hotelling's[5] publication on principal component analysis (PCA) provided the mathematical approach for data representation and separation in multidimensional space. In 1968, Ben-Gera and Norris[6] published their initial work on using multiple linear regression (MLR) to solve the problem of the calibration for agricultural products.

Tremendous advances have been made recently in the use of NIR spectroscopy for quantitative analyses thanks to the advances in instrumentation, software, and sample handing. Today, quantitative NIR with both transmittance and reflectance detectors is widely used for off-line and on-line analysis in agriculture, food, beverages, wool, textiles, biomedicals, petrochemicals, polymers, and pharmaceuticals. The pharmaceutical industry is utilizing NIR methods to monitor the manufacturing process, from raw material to finished product, which reduces the cost of manufacturing and improves the quality and safety of the drug products. However, quantitative NIR has not widely been used for low potency pharmaceutical dosage forms.

Chemometric methods for quantitative transmittance and reflectance NIR spectrophotometry were applied in this research to the analysis of nicotine-containing chewing gum. Different sample selection (or outliers detection) methods (random selection, Mahalanobis distance in principal component space selection, and maximum distance in wavelength space selection), different mathematical pre-treatments (1^{st} derivative, 2^{nd} derivative, Standard Normal Variate (SNV), Detrend, Savitzky-Golay, and Multiplicative Scatter Correction (MSC)), and different regression methods (simple linear regression, multiple linear regression, and partial least squares regression) were studied. In addition, the sample presentation and the selection of analytical wavelength were investigated.

Nicotine chewing gum is a therapeutic product that contains a small amount of nicotine, which is designed to help smokers quit by quelling their desire for nicotine, the principal alkaloid in tobacco that accounts for the widespread human use of its products. Smoking is rapidly becoming the world's number one cause of death. In the United States, smoking kills more people than car accidents, AIDS, alcohol, illegal drugs, homicides, and fires combined[7].

For safety[8] and regulatory reasons[9], all pharmaceutical products, including nicotine chewing gum, have to be monitored for quality and purity both upon QA release and during stability-testing. Traditionally nicotine is analyzed by titration[10,11], UV absorption[12], paper chromatography[13], TLC[14], GC[15,16,17,18,19], GC-MS[20,21,22,23] HPLC[24,25,26,27,28,29,30], LC-MS[31,32,33], and CE[34,35,36]. Due to the characteristics of the gum base, it is necessary to use potentially harmful organic solvents such as THF, for sample preparation prior to analysis. These solvents may not only cause health, safety, and environmental

problems but also increase the cost to the manufacturers, which is passed on to consumers.

NIR has been used to analyze nicotine in tobacco leaf for many years[37,38,39], but has not been utilized to analyze the nicotine in pharmaceutical dosage forms because of the unique properties of the chewing gum – it is a low potency product with a flexible surface, made using a complicated formulation of ingredients with very fine particles.

2. NEAR-INFRARED THEORY

2.1 Near-Infrared Spectrum

In order to absorb infrared (IR) radiation, a molecule must undergo a net change in dipole moment as a consequence of its vibrational or rotational motion. Only under these circumstances can the alternating electrical field of the radiation interact with the molecule and cause changes in the amplitude of one of its motions. Vibrational modes which do not involve a change in dipole moment are infrared-inactive. The IR region ranges from about 780 nm to 1000 μm and the NIR is in the range of 780 nm to 2500 nm where absorptions correspond to overtones and combinations of fundamental vibrational transitions that occur in the mid-infrared (MIR) region. Figure 1 shows the spectral regions used in the different techniques[40].

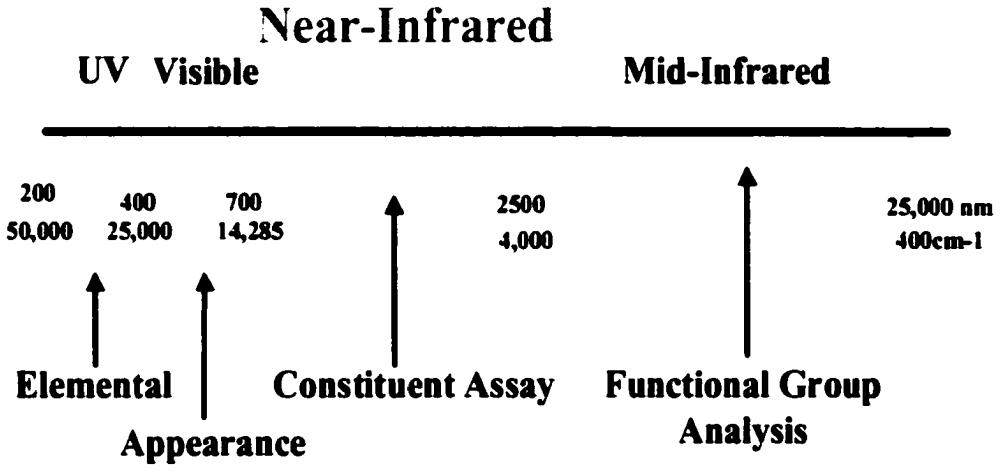

Figure 1. Spectral Regions Used in Different Techniques

2.1.1 Vibrational Transitions

Vibrational spectroscopy is based on the concept that molecules vibrate with certain frequencies which can be described by physics[41,42] and quantum mechanics[43,44]. When the molecules absorb light of a particular frequency, they are excited to a higher energy level.

For diatomic molecules, the time-independent Schroedinger equation can be solved using the vibrational Hamiltonian.

$$\frac{-h^2\partial^2\psi(x)}{2m\partial(x)} + V(x)\psi(x) = E\psi(x) \qquad \textit{Equation 1}$$

where

h = Planck constant

m = Mass

$\psi(x)$ = Wave function

V(x) = Potential energy

E = Energy of the system

A simplified solution of the equation gives

$$\boxed{E\upsilon = (\upsilon + 1/2)h\nu} \quad (\upsilon = 0, 1, 2, \ldots) \qquad \textit{Equation 2}$$

In the case of polyatomic molecules, the energy levels become quite numerous. To a first approximation, such a molecule can be treated as a series of diatomic, independent, harmonic oscillators. The equation for this case can be expressed as

$$\boxed{E(\upsilon_1, \upsilon_2, \upsilon_3 \ldots) = \sum_{i+1}^{3N-6} (\upsilon_i + 1/2)h\nu} \qquad \textit{Equation 3}$$

where $\upsilon_1, \upsilon_2, \upsilon_3, \ldots = 0, 1, 2, \ldots$

There are (3N-6) vibrational modes for non-linear molecules and (3N-5) modes for linear molecules. Each of the (3N-6) or (3N-5) vibrational modes is called a normal mode. The observed number of absorption bands may be less than predicted if some vibrations are infrared-inactive, degenerate, very weak, not resolved instrumentally, or the vibrations occur outside the range of the spectrophotometer. The number of observed absorption bands may be greater than predicted if overtones and combination bands occur.

2.1.2 Fundamental Frequencies

At room temperature, most molecules are at their rest energy levels and vibrate in the least energetic state allowed by quantum mechanics, $\upsilon=0$. The lowest or fundamental frequencies of any two atoms connected by a chemical bond may be roughly calculated by assuming that the band energies arise from the vibration of a diatomic harmonic oscillator and obey Hooke's Law:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{\kappa}{\mu}} \qquad \textit{Equation 4}$$

where

ν = vibrational frequency

κ = classical force constant

μ = reduced mass of the two atoms

2.1.3 Anharmonicity and Overtones

The selection rules predict that the fundamental absorption will occur with $\Delta\upsilon = \pm 1$ (υ from 0 to 1, Figure 2) and other transitions, such as overtones and combination bands are not allowed. However, overtones and combination bands do appear as weak bands due to anharmonicity as shown in Figure 3.

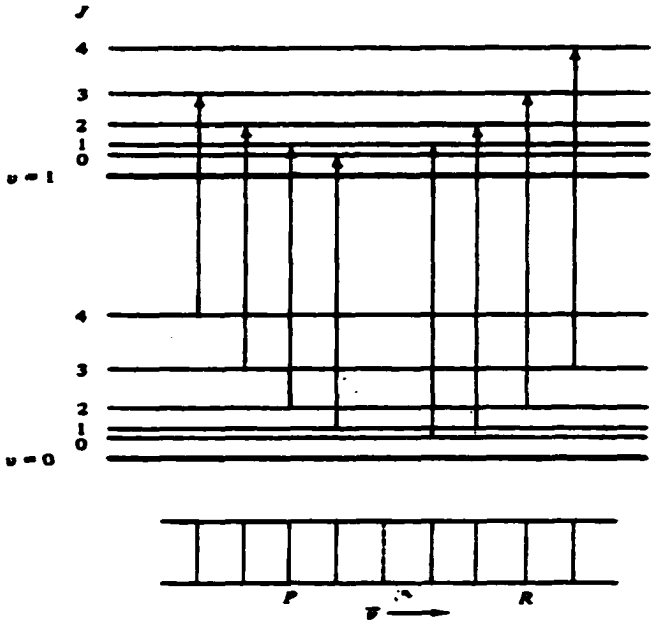

Figure 2. Vibrational Transition, Δυ=1

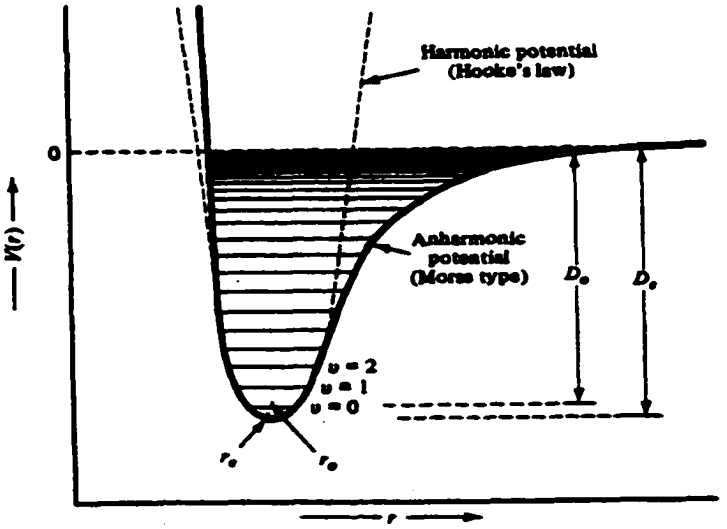

Figure 3. The Harmonic and Anharmonic Potentials for a Diatomic Molecule.

The energy levels in the anharmonic oscillator are not equal. The levels are slightly closer as the energy increases. This phenomenon can be seen in the equation:

$$E\upsilon = (\upsilon + 1/2)h\omega_e - (\upsilon + 1/2)^2\omega_e\chi_e + \ldots \qquad \textit{Equation 5}$$

where

$\omega_e = (1/2\pi)(K_e/\mu)^{1/2}$ is a vibrational frequency

$\omega_e\chi_e$ = anharmonicity constant

K_e = harmonic force constant

μ_e = reduced mass of the two atoms

At higher quantum numbers, ΔE becomes smaller and the selection rule is not rigorously followed. As a result, transitions of $\Delta v = \pm 2$ or ± 3 are observed. Such transitions are responsible for the appearance of first or second overtone lines at frequencies approximately two or three times the frequency of the fundamental line. The intensity of overtone absorption is about 10 to 1000 times weaker than fundamental bands since they are forbidden quantum mechanically, i.e. less probable.

The majority of overtone peaks seen in a NIR spectrum arise from the X-H stretching modes (C-H, N-H, O-H, and S-H).

2.1.4 Combination Bands

Vibrational spectra are further complicated by the fact that two different vibrations in a molecule can interact to give absorption peaks with frequencies that are approximately the sums or differences of their fundamental frequencies when the absorption of a photon excites two vibrational modes simultaneously. The intensities of combination and

difference peaks are generally low. If the absorption frequencies of the two independent vibration are ν_1 and ν_2, a combination band may be found at about frequency $(\nu_1 + \nu_2)$. Combination bands may be observed in NIR region. Table 1 shows some of the most often observed wavelengths in NIR region[45,46] might applicable to this research.

Table 1. Basic Characteristic Wavelengths in NIR Region

Wavelength (nm)	Bond Vibration	Structure
890	C-H stretch 3rd overtone	-CH3
910	C-H stretch 3rd overtone	-CH2
930	C-H stretch 3rd overtone	-CH
1143	C-H stretch 2nd overtone	-ArCH
1170	C-H stretch 2nd overtone	.HC=CH
1195	C-H stretch 2nd overtone	-CH3
1215	C-H stretch 2nd overtone	-CH2
1225	C-H stretch 2nd overtone	-CH
1360	C-H combination	-CH3
1395	C-H combination	-CH2
1415	C-H combination	-CH
1417	C-H combination	-ArCH
1420	O-H 1st overtone	H2O
1440	C-H combination	-CH2
1446	C-H combination	Aromatic
1620	C-H stretch 1st overtone	=CH2
1685	C-H stretch 1st overtone	Aromatic
1695	C-H stretch 1st overtone	-CH3
1705	C-H stretch 1st overtone	-CH2
1725	C-H stretch 1st overtone	-CH
1780	C-H stretch 1st overtone	Cellulose
1940	O-H combinations	H2O
2180	C-C combinations	C-C
2300	C-H+C-H combinations	-CH3
2350	C-H+C-H combinations	-CH2
2420	C-H+C-H combinations	-CH

Note: The wavelengths are approximated numbers and vibration frequencies vary from molecule to molecule.

3. NEAR-INFRARED INSTRUMENTATION

Karl Norris, considered by many to be the "father" of NIR, demonstrated in the late 1960s the potential value of the NIR region for quantitative work by making measurements on agricultural products. Since the first commercial NIR was introduced in 1971 by Dickey-John at Illinois State Fair[47], the instrumentation has been improved dramatically for quantitative and qualitative analysis. Modern NIR instruments consist of light source, monochromator, sample device, detector, and data processor. Based on the position of the detector, NIR instrumentation can be divided into transmittance NIR and reflectance NIR. A detailed discussion of the history of NIR instrumentation was published[48] and the technologies used in NIR instrumentation were reviewed[49,50].

3.1 Transmittance Near-Infrared

In transmittance measurements, the entire sample in the light path is measured, which reduces errors due to sample non-homogeneity. Transmittance techniques are most useful for measuring translucent samples and large particles. Small particle samples cause greater scattering, decrease the light penetrating through the sample, and produce a weaker signal-to-noise ratio. In transmittance, higher frequencies (800 nm – 1400 nm) are most commonly used due to the greater depth of penetration into the sample. Higher frequencies are more susceptible to front surface scattering than lower frequencies energy. Therefore, transmittance measurements must be optimized for the frequency used for measurement, front surface scatter, and pathlength of the sample. Figure 4 is a schematic diagram of the basic transmittance NIR spectrophotometer.

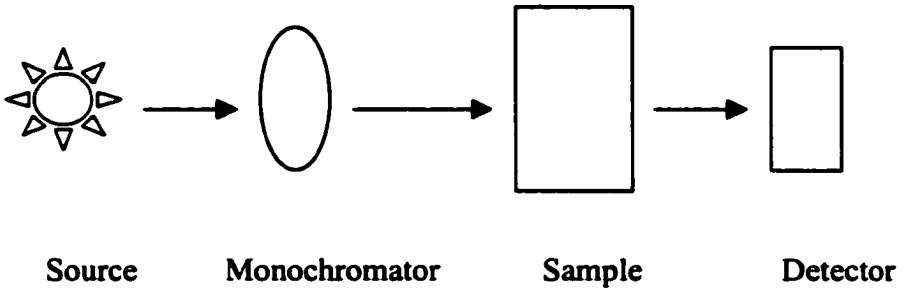

Figure 4. Diagram of Basic Transmittance NIR Spectrophotometer

The detectors for transmitted light are placed at right angle to the light path (diameter <1 mm) and normally a single detector is used.

3.2 Reflectance Near-Infrared

The reflectance measurements penetrate only about 1-4 mm into the front surface of the samples. This small penetration of energy into a sample brings greater variation when measuring non-homogeneous samples than the transmittance technique. In cases where small particles are involved, reflectance measurements provide the greatest bulk chemical information. Figure 5 is a representation of a basic reflectance NIR spectrophotometer.

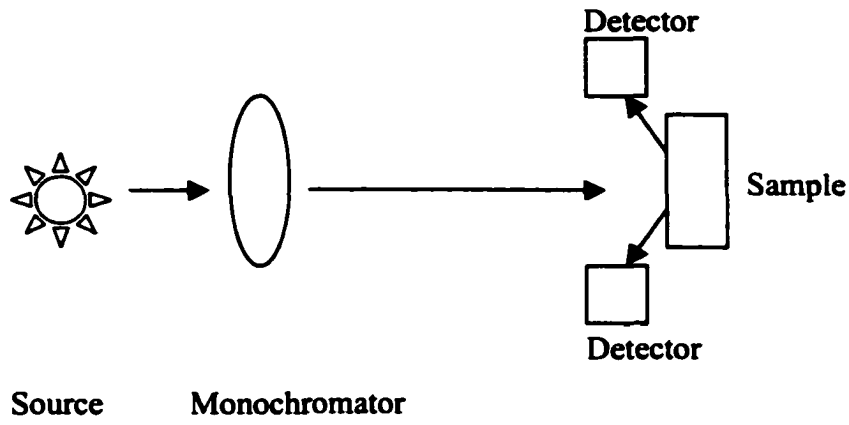

Figure 5. Diagram of Basic Reflectance NIR Spectrophotometer

The detectors for reflected light are placed at 45° and normally two or four detectors are used.

3.3 Combination of Transmittance and Reflectance NIR

As samples increase in absolute reflectivity such as samples with small particles, less bulk information can be measured using either transmittance or diffuse reflectance measurement geometries. To improve the signal-to-noise ratio, the ideal instrument would have both transmittance and reflectance capabilities. Figure 6 is a diagram of basic NIR spectrophotometer utilizing both transmittance and reflectance detectors.

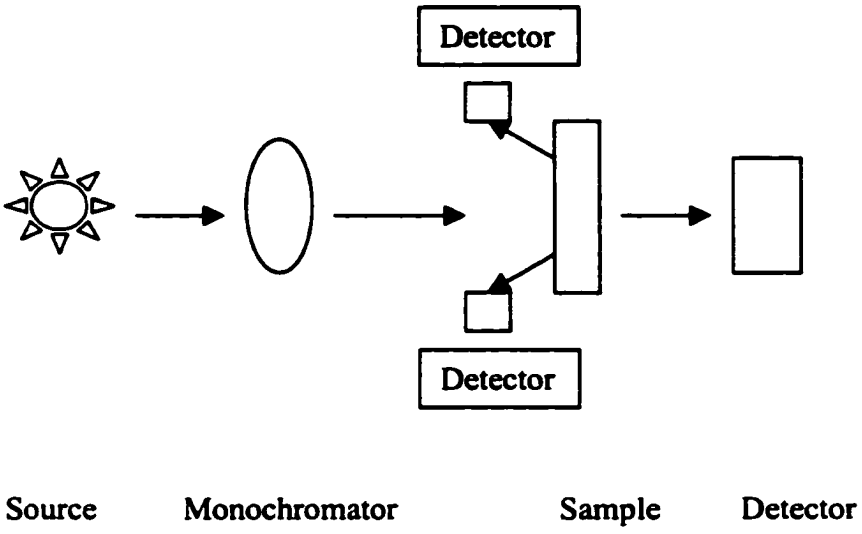

Figure 6. Diagram of Basic NIR Spectrophotometer Using both Transmittance and Reflectance Detectors

The advantage of the transmittance detector for non-homogeneous samples and the reflectance detector of greater signal-to-noise ratio provides the unique advantage of an NIR system with both transmittance and reflectance detectors.

3.4 Source Type

IR sources consist of an inert solid (such as Nernst glower or globar) that is heated electrically to produce continuous radiation similar that of a blackbody. The most common source of near-infrared is the tungsten-halogen filament lamp. Near-Infrared emitting diodes (NIR-Eds) are also used in some NIR instruments.

3.5 Detector Type

Photoconducting detectors are widely used in NIR instrumentation with semiconductors such as silicon, InAs, InSb, PbS, PbSe, and InGaAs. Lead sulphide (PbS) detectors are used mainly for measurements in the 1100 nm – 2500 nm region, PbS "sandwiched" with silicon photodiodes are most often used for visible-near-infrared work in the region of 400 nm – 2500 nm, and InGaAs detector is often used in the range of 600 nm – 1900 nm. Absorption of radiation by these materials promotes nonconducting valence electrons to a higher-energy conducting state, thus decreasing the electrical resistance of the semiconductor. Typically, a photoconductor is placed in series with a voltage source and load resistor, and the voltage drop across the load resistor serves as a measure of the power of the beam of radiation.

3.6 Instrument Specifications

Instrument specifications give the function and quality of the instrument (ref. 43). It is important to select an instrument with suitable specifications for specific applications.

3.6.1 Wavelength Range

The total useful wavelength range of the instrument, which is determined by the lamp source, optical design, and detector. The typical wavelength range of NIR instrument is about 700 nm – 3000 nm.

3.6.2 Spectral Bandwidth

The bandwidth can be defined as the full width at maximum of the bandshape of monochromatic radiation passing through a monochromator. Bandwidth affects the sharpness of the spectral peak.

The smaller the bandwidth, the higher the resolution. Modern NIR instrument has a typical bandwidth of about 10 nm.

3.6.3 Wavelength Accuracy

Wavelength accuracy is defined as the difference between the measured and reported wavelengths of a wavelength standard, which is important to identify and quantify a constituent correctly. Standard reference materials are available from the National Institute for Standards and Technology (NIST) for wavelength accuracy measurement. A quality NIR instrument typically has wavelength accuracy of less than 1 nm.

3.6.4 Photometric Noise

Photometric noise is the actual root mean square deviation for standard scans within a specified wavelength range at a specified absorbance level. Typically, photometric noise is about 0.02 mAU for wavelength of 700 nm – 3000 nm.

3.6.5 Dynamic Range

Dynamic range is the range of an instrument from the lowest detectable level to the highest concentration level that produces reproducible results. Dynamic range is different from linear range that is specific to product.

3.6.6 Baseline Shift

Baseline shift (or absorbance stability) is the change in photometric value of a spectrometer's baselines at a specific wavelength with respect to time. Normally, it is determined during instrument

qualification by scanning a specific reference material. Normally the value is about 0.05 mAU/hr

3.6.7 Stray Light

Stray light is the major cause of nonlinearity for most NIR instruments and is defined as the total light power at wavelengths other than the wavelength of interest that reaches the detector.

3.6.8 Scan Rate

Scan speed is defined as the number of scans (full range) that can be performed in a given time unit. Typical scan rate of a modern NIR instrument is more than 1 scan per second.

3.6.9 Data Interval

Data interval is the time over which data is captured. The smaller the interval, the greater the data/information available for processing. Normally, the data interval is about 2 nm or less.

3.7 Instrument Qualification

Like any other analytical instrument, the NIR spectrophotometer must be qualified prior to use to ensure the performance is meeting all expectations[51,52]. Instrument qualification normally includes Design Qualification (DQ), Installation Qualification (IQ), Operational Qualification (OQ), Performance Qualification (PQ), and Maintenance Qualification (MQ).

3.7.1 Design Qualification (DQ)

The design qualification should document that the quality of the instrument has been built into the design of the application, the manufacturer compliants with appropriate quality systems, and the instrument meets a set of specifications.

3.7.2 Installation Qualification (IQ)

The installation qualification documents the receipt of the instrument, the selection of the location, and the appropriate installation of the equipment by qualified personnel.

3.7.3 Operational Qualification (OQ)

Operational qualification documents the functional test to confirm the performance specifications.

3.7.4 Performance Qualification (PQ)

The performance qualification provides the documentation that the performance of the instrument meets certain requirements before analyzing the samples. The performance verification normally is performed on a daily (working day) basis.

3.7.5 Maintenance Qualification (MQ)

The maintenance qualification documents the service and re-qualification performed on the instrument and periodic preventive maintenance.

4. THEORY OF DIFFUSE REFLECTANCE

The use of NIR diffuse reflection for the quantitative analysis of many products and commodities is now becoming widely accepted. Many diffuse reflectance theories have been developed[53,54,55,56,57,58]. Lambert[59] first provided a mathematical description of diffuse reflection in 1760 and Mie[60] developed a more general theory in 1908. However, the Kubelka-Munk theory[61,62] is the most widely accepted.

Kubelka and Munk derived a mathematical equation to describe diffuse reflection

$$\frac{A}{S} = \frac{(1-R)^2}{2R} = F(R) \qquad \textit{Equation 6}$$

where $F(R)$ is known as K-M function, R is the diffuse reflectance, A is the absorption coefficient, and S is the scattering coefficient.

It may be stated that R, the diffuse reflectance, a function of the ratio A/S, is proportional to the absorbing species added to the reflecting sample medium. The relationship is based on the assumption that the diffuse reflectance of an incident beam of radiation is directly proportional to the quantity of absorbing species interacting with the incident beam, and so R depends on analyte concentration.

For quantitative analysis, the above equation can be used to determine the concentration c of an absorbing analyte:

$$\frac{A}{S} = \frac{(1-R)^2}{2R} = \frac{ac}{S} \qquad \textit{Equation 7}$$

where a is the absorptivity and c is the analyte concentration.

In practical NIR measurement, a relative reflectance, R', rather than an absolute diffuse reflectance, R, is usually measured. The relationship exists at every wavelength measured.

$$R' = \frac{Intensity\ of\ sample\ reflec.}{Intensity\ of\ reference\ reflec.} \qquad Equation\ 8$$

It is generally accepted that the K-M equation, like Beer's law, is a limiting equation and should only apply for weakly absorbing bands when the product absorptivity and concentration is low. As described above, for organic materials, absorptions in the NIR are due to vibrational overtones and combination bands. The absorptivities of these bands are much weaker (10^{-1} to 10^{-3}) than the absorptivities of the corresponding fundamental vibrations. Thus most organic analytes can be considered to be weakly absorbing in the NIR region even without dilution. However, for most NIR analyses, the analyte is usually not isolated from other components, but is surrounded by a matrix that may be not only complex but also may absorbs the incident radiation at least as strongly as the analyte at the analytical wavelengths. Therefore, it is expected that unless a proper referencing method is used, absorption by the matrix surrounding the analyte may cause deviations from the K-M equation.

5. EXPERIMENTAL

5.1 Nicotine and Gum Tablet

Nicotine, 3-(1-methyl-2-pyrrolidinyl) pyridine, a tertiary amine composed of a pyridine and pyrrolidine ring, is the principal alkaloid of the cultivated tobacco species[63,64]. It is an unstable compound and degrades

mainly into cotinine, myosmine, nornicotine, 1'R,2S-nicotine-*N*-oxide, and 1'R,2S-nicotine-*N'*-oxide, especially on exposure to air and light[65]. The commercially available nicotine is a byproduct of the tobacco industry and contains anatabine and anabasine as additional impurities[66]. Nicotine has a molecular formula $C_{10}H_{14}N_2$ and molecular weight 162.23 dalton. It has pK_b values of 6.16 and 10.96 at 15°C[67].

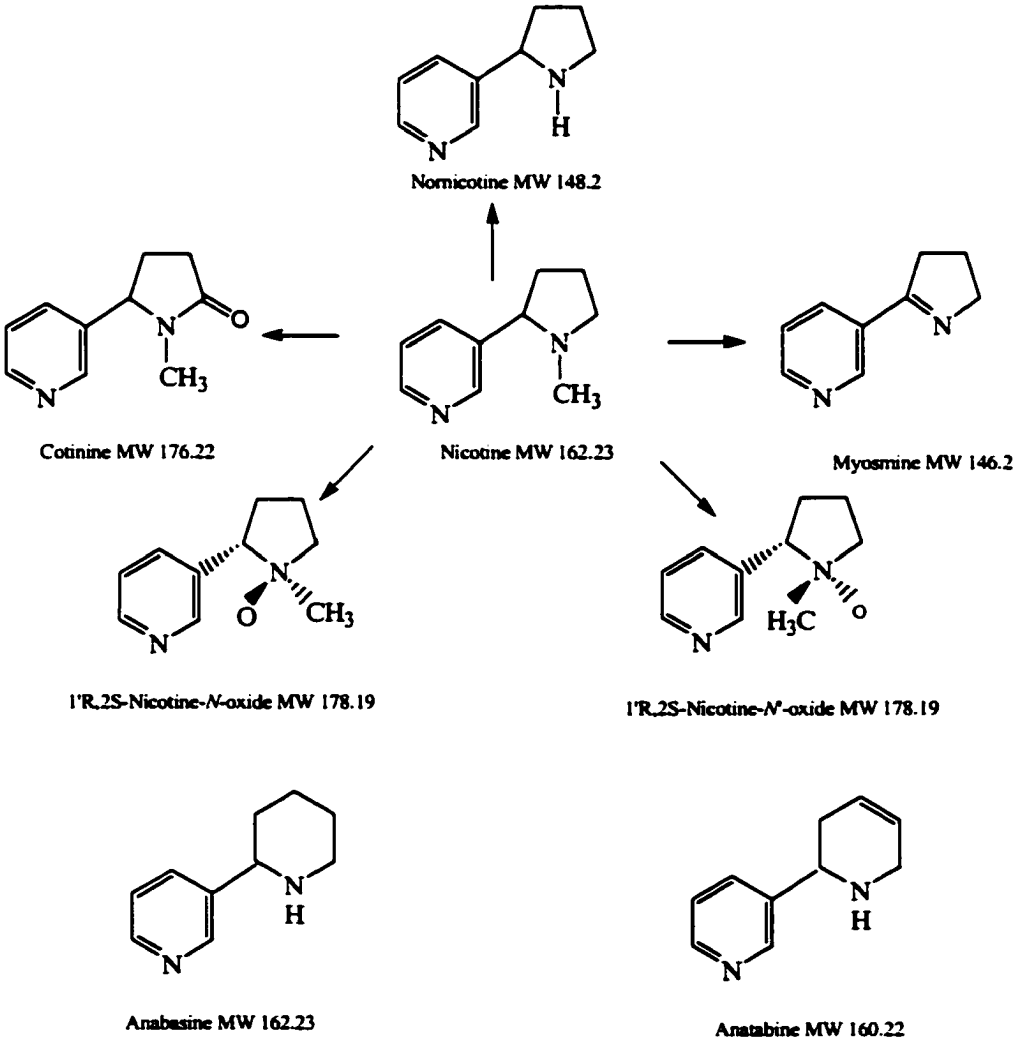

Figure 7. Nicotine Degradation Pathways and Its Impurities

5.2 Nicotine Tablet Preparation and Collection

Sample collection is an important step in quantitative method development and validation. All the samples collected for method development and validation must have the same level of authentication.

The preparation of a calibration set involves the following important considerations. The samples should cover the complete range of component concentration as evenly distributed as possible. Random sample selection will allow the mathematical model to most closely fit the middle concentration samples leaving few extremely high or low concentration level samples, which would influence the slope and intercept inordinately. An even concentration distribution allows the model to minimize the residuals at the extremes and at the center with relatively equal weighing.

To obtain a robust calibration, different batches should be selected over a period of time to cover the possible changes in constituent concentration, process, temperature, particle size, moisture, residual solvents, degradation products, time, operator, and purity. The number of the batches required depends on the complexity of the application.

For this research, five batches of nicotine tablet (Nicotine Tablets, 4 mg/tablet, Batch 1, 2, 3, 4, and 5. Manufactured by GlaxoSmithKline Co.), manufactured on different dates with different lots of excipients were selected. The concentration ranged from about 3.6 mg/tablet to about 4.4 mg/tablet. Batch 1 was stressed by storage at 40°C and 75% relative humidity for three months to produce different levels of water and degradation products in the samples, to cover all realistic variations.

Twelve tablets were randomly selected and analyzed from each batch, a total of 60 (5x12) tablets. The 60 samples were divided into three sets: calibration set (36 samples of batches 1-4), validation set (12 samples of batch 1-4), and prediction set (12 samples of batch 5). The samples used for calibration and validation were from same batches while the samples used in prediction set were from an independent batch.

In the calibration set, the NIR data were plotted against the true potencies of the gum samples obtained from HPLC analysis and an regression plot was obtained.

In the validation set, the values of the samples (from the same population as the calibration set) predicted by the calibration plot were compared to that obtained from the HPLC analysis.

In the prediction set, the samples (from a independent batch) were treated as unknowns and the potencies were determined by the calibration plot. The determined values by NIR then were compared to those obtained from the HPLC analysis.

The above procedure was used for transmittance and reflectance NIR, respectively.

5.3 Transmittance NIR

5.3.1 Equipment

The InTact® Analyzer (Foss NIRSystems[68], Silver Spring, Maryland) had the specifications given in Table 2:

Table 2. The Specifications of InTact® Analyzer

Source	Tungsten Halogen Filament Lamp
Detector	Indium-Gallium-Arsenide (InGaAs)
Wavelength Range	600-1900 nm
Photometric Range	3.0 AU
RMS Noise	<20 μAU
Data Interval	2.0 nm
Scan Speed	1.8 scans/sec
Wavelength Accuracy	0.3 nm
Wavelength Precision	0.01 nm
Spectral Bandwidth	9.5 ± 1 nm
Operating Temperature	15-32°C

The InTact® Analyzer, its optical design, and detector are shown in Figures 8, 9, and 10, respectively.

Figure 8. InTact® Analyzers Model AP-1600-II

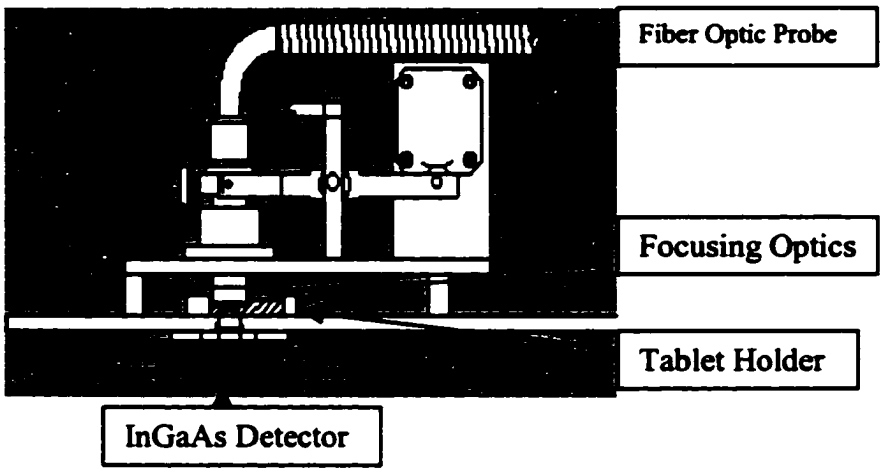

Figure 9. InTact® Analyzer Transmittance Design

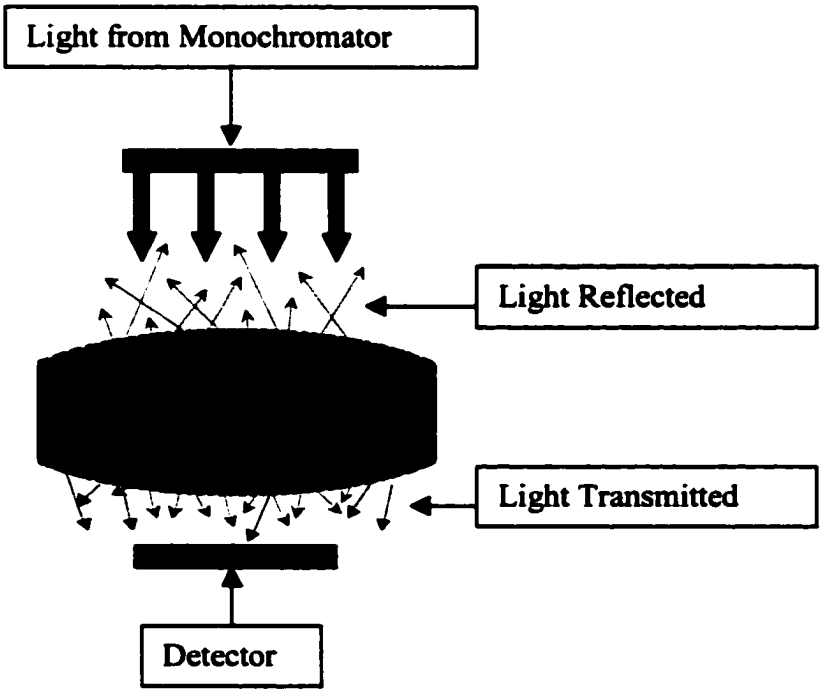

Figure 10. InTact® Analyzer Signal Detection

5.3.2 Instrument Performance Check

The transmittance NIR spectrophotometer was checked for photometric noise and wavelength accuracy/precision prior to use to ensure the reliability of the results.

5.3.2.1 Photometric Noise

The photometric noise was measured by scanning the reference material (Coors ceramic plate) twice, first as a reference and then again as a sample. This test was repeated 10 consecutive times and the mean of all results was calculated and reported. Table 3 is the summary of the photometric noise test. All results met the instrument specifications.

Table 3. Results of the Photometric Noise Test

Scan Range 850-1100 nm (mA)			
Test	Specification	Actual	Valid
P-P	0.200	0.055	Yes
RMS	0.025	0.011	Yes
Bias	0.100	0.001	Yes
Scan Range 1100-1650 nm (mA)			
Test	Specification	Actual	Valid
P-P	0.150	0.030	Yes
RMS	0.020	0.004	Yes
Bias	0.100	0.008	Yes

5.3.2.2 Wavelength Accuracy/Precision

An internal polystyrene film reference was subjected to 10 replicate scans. Table 4 is the summary of the wavelength accuracy/precision test. All results met the instrument specifications.

Table 4. Results of the Wavelength Accuracy/Precision Test.

Wavelength (nm)	Specification (nm)	Accuracy (nm)	Precision <0.010 nm	Valid
804.54	804.40±0.30	0.14	0.001	Yes
877.91	878.10±0.30	-0.19	0.001	Yes
1143.62	1143.63±0.30	-0.01	0.001	Yes
1680.90	1680.90±0.30	0.00	0.002	Yes

5.3.3 Sample Presentation

The consistency of the sample presentation is another important factor. To study the effect of the gum surface, each tablet was scanned in four different orientations by rotating the sample 180° with respect to its x and y axes. A special sample holder was made to fit the tablet sample to reduce stray light. Figure 11 is the picture of the sample holder made for this research.

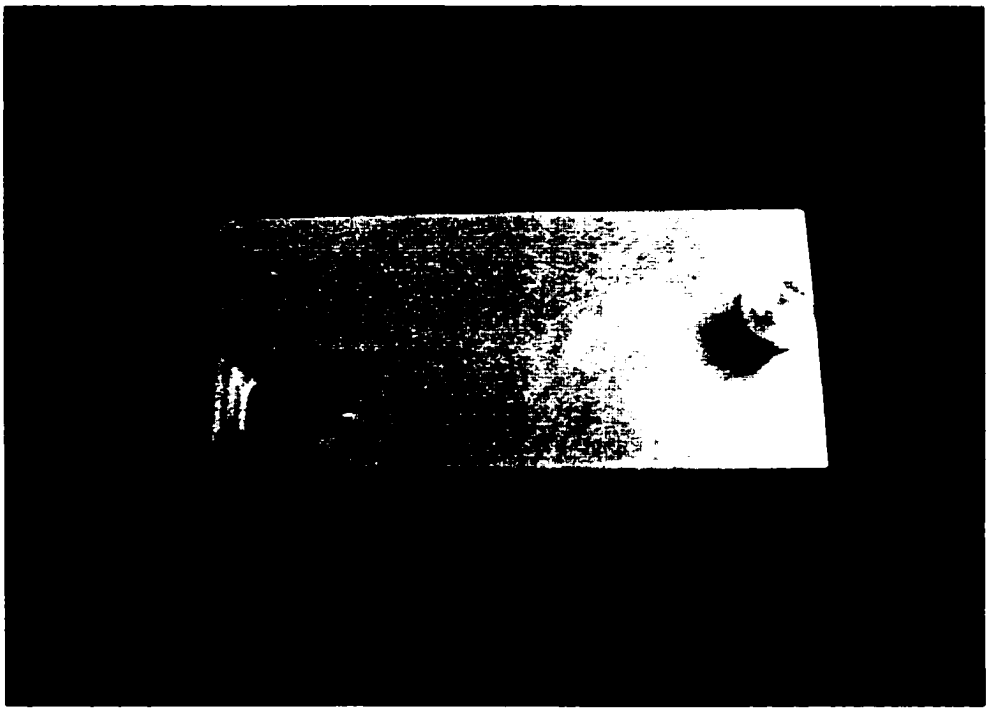

Figure 11. Sample Holder Specially Made for Nicotine Tablet

5.3.4 Sample Acquisition

Each sample was scanned 32 times at a scan rate of 1.8 scans/sec in the range of 600 nm – 1900 nm.

5.4 Reflectance NIR

5.4.1 Equipment

The Rapid-Content® Analyzer (AP-1365-II, Foss NIRSystems[69], Silver Spring, Maryland) had the specifications given in Table 5.

Table 5. The Specifications of Rapid-Content® Analyzer

Source	Tungsten Halogen Filament Lamp
Detector	Silicon for 400-1100 nm Lead Sulfide for 1100-2500 nm
Wavelength Range	400-2500 nm
Photometric Range	3.0 AU (1100-2500 nm)
RMS Noise	<0.2 mAU (400-700 nm) <0.04 mAU (700-2500 nm)
Data Interval	2.0 nm
Scan Speed	1.8 scans/sec
Wavelength Accuracy	0.3 nm
Wavelength Precision	0.01 nm
Spectral Bandwidth	10 ± 1 nm
Operating Temperature	15-33°C

The Rapid-Content® Analyzer, the detection, and reflectance light are shown in Figures 12, 13, and 14, respectively.

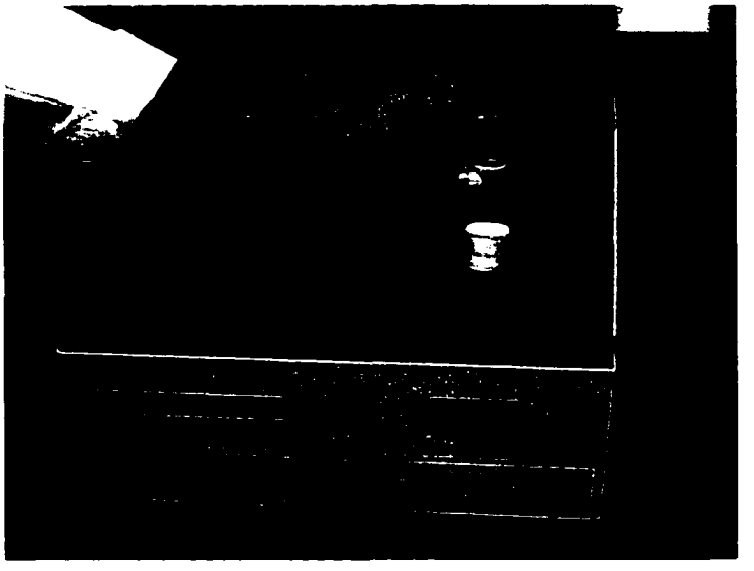

Figure 12. Rapid-Content ® Analyzers Model AP-1365-II

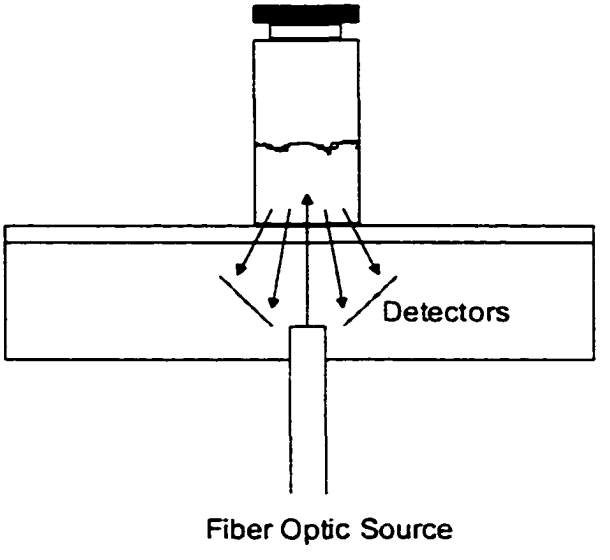

Figure 13. Detection of Rapid-Content® Analyzer

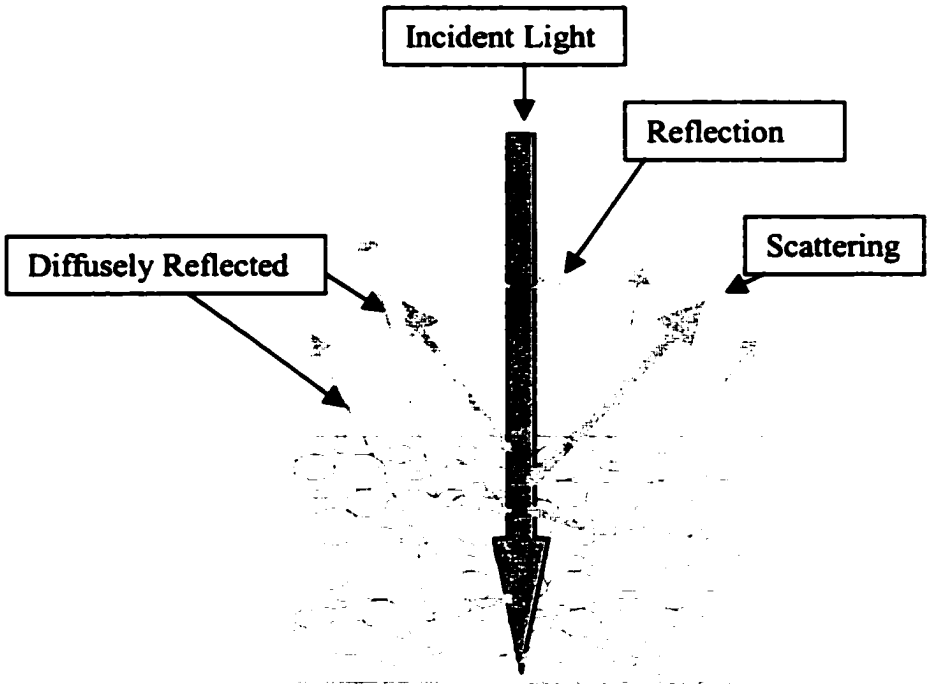

Figure 14. Rapid-Content® Analyzer Signal Detection

5.4.2 Samples

The samples run in the reflectance NIR mode are listed in Table 6.

Table 6. Samples Analyzed in Reflectance NIR

No.	Sample Name
1	Nicotine Tablet, 4 mg/tablet, Batch 1, 2, 3, 4, and 5, GlaxoSmithKline (GSK) Co. (Parsippany, NJ).
2	Nicotine Polacrilex (active pharmaceutical ingredient), GSK (Arvine, Scotland).
3	Gum Base (excipient), Dandy/Fertin (Vejle, Denmark).
4	Sorbitol (excipient), EM Science (Cincinnati, OH).
5	Mannitol (excipient), Atlas Chemical Industries, Ltd. (Brooklyn, NY).
6	Talc (excipient), J. T. Baker Chemical Co. (Phillipsburg, NJ).
7	Quiniline Yellow (excipient), Warner Jerkinson (South Plainfields, NJ).
8	Sodium Carbonate (excipient), J. T. Baker Chemical Co.
9	Cotinine (impurity), Sigma (St. Louis, MO).
10	Myosmine (impurity), Sigma (St. Louis, MO).
11	Nornicotine (impurity), Toronto Research Cooperative (TRC) (Toronto, Canada).
12	1'R,2S-Nicotine-*N*-oxide (impurity), TRC.
13	1'R,2S-Nicotine-*N'*-oxide (impurity), TRC.
14	Anatabine (impurity), TRC.
15	Anabasine (impurity), Lancaster Synthesis (Windham, NH).

5.4.3 Instrument Performance Checks

The instrument passed the performance checks prior to use. The results are similar to those in Section 7.3.3.

5.4.4 Sample Presentation

The gum samples were analyzed directly on the Rapid-Content® Analyzer and powder ingredients and impurities were analyzed in NIR-inactive glass vials.

5.4.5 Sample Acquisition

Each sample was scanned 32 times at a scan rate of 1.8 scans/sec in the range of 400 nm – 2500 nm.

5.5 HPLC Analysis

After the samples scanned with transmittance and reflectance NIR spectrophotometers, respectively, they were analyzed by a validated HPLC procedure. In the HPLC procedure, nicotine was extracted from Nicotine Tablets using tetrahydrofuran (THF) and 0.1 N hydrochloric acid (HCl) with aid of mechanically shaking. The sample solution was centrifuged then analyzed against nicotine bitartrate reference standard on HPLC using reversed phase chromatography. The nicotine peak was identified by comparing its retention time and UV spectra to those of nicotine bitartrate reference standard.

5.5.1 Equipment

The equipment used in HPLC analysis are listed in Table 7.

Table 7. Equipment Used in HPLC Analysis

No.	Name
1	Balance, AE240, Mettler (Columbus, OH).
2	KF Titrator, 703 Ti Stand, Brinkmann (Westbury, NY).
3	Centrifuge, Bio-fuge 15R, Heraeus (Heraeusstreet 12-14 63450 Hanau, Germany).
4	Shaker, HS 501 digital, IKA Work (Wilmington, NC).
5	HPLC column, RP-18, 4.6 x 50 mm, 3.5 μm, Waters Xterra (Milford, MA).
6	HPLC system, Waters Alliance equipped with PDA detector (Milford, MA).

5.5.2 Samples

The Samples run in HPLC analysis are listed in Table 8.

Table 8. Samples Run in HPLC Analysis

No.	Sample Name
1	Nicotine Tablet, 4 mg/tablet, Batch 1, 2, 3, 4, and 5.
2	Cotinine
3	Myosmine
4	Nornicotine
5	1'R,2S-Nicotine-*N*-oxide
6	1'R,2S-Nicotine-*N'*-oxide
7	Anatabine
8	Anabasine

5.5.3 Reagents

The reagents prepared and used are listed in Table 9.

Table 9. Reagents Prepared and Used in HPLC Analysis

No.	Name
1	Water, USP grade, in-house.
2	Acetonitrile, HPLC grade, J. T. Baker (Phillipsburg, NJ).
3	Tetrahydrofuran, ACS reagent grade, J. T. Baker.
4	Ammonium phosphate, monobasic, ACS reagent grade, J. T. Baker.
5	Ammonium hydroxide solution (about 28-30%), ACS reagent grade, EM Science (Cincinnati, OH).
6	**Diluent:** 10.0 g of potassium phosphate, dibasic, was dissolved in 1000 mL of water and the pH was adjusted to 11.0 with potassium hydroxide solution. The concentration was about 73 m*M*.
7	1 ***M*** **Ammonium Hydroxide:** concentrated ammonium hydroxide solution, 69 mL, was diluted to 1000 mL with water, and mixed well
8	Nicotine bitartrate, dihydrate, USP Reference Standard (Rockville, MD).
9	Karl Fisher Reagent, Comp 5, J. T. Baker.

5.5.4 Preparation of Mobile Phase

Ammonium phosphate, monobasic, 5.7 g, was dissolved in 450 mL of DI water. About 425 mL of 1M ammonium hydroxide solution and 125 mL of acetonitrile were added and mixed well. the mobile phase was filtered through a 0.45-micron filter and degassed with helium for 10 minutes before use.

5.5.5 Water Content of Reference Standard

About 100 mg of Nicotine Bitartrate Reference Standard was accurately weighed into a Karl-Fisher (KF) titrator and the solution was titrated to the coulometric end point with KF reagent (Comp 5, J. T. Baker). The water content of the reference standard was determined to be 7.4% which was within the specification. The theoretical water content for the dihydrate is 7.2%.

5.5.6 Preparation of Standard Solution

Approximately 24 mg of nicotine bitartrate dihydrate reference standard was accurately weighed and quantitatively transferred into a 100-mL volumetric flask. Exactly 30.0 mL of 0.1 N HCl and 20.0 mL of THF were added. The solution was mixed well and labeled as "Stock-Std".

Exactly 5.0 mL of "Stock-Std" was transferred into a 100-mL volumetric flask, diluted to volume with diluent, mixed well, and labeled as "Working-Std". This solution contained about 8 μg/mL of nicotine.

5.5.7 Preparation of Sample Solution

Each tablet was accurately weighed (for information only) and individually transferred into separate glass bottles (about 100 mL). Exactly 20.0 mL of THF was added and the sample was shaken for 30 minutes to dissolve. Exactly 30.0 mL of 0.1 N HCl was added and the sample was shaken for additional 5 minutes. A portion of the sample was centrifuged at 3000 rpm for 15 minutes and 5.0 mL of the supernatant

was pipetted into a 50-mL volumetric flask, diluted to volume with diluent, and mixed well.

5.5.8 Chromatographic Parameters

The HPLC parameters used are listed in Table 10.

Table 10. Parameters Used in HPLC Analysis

Column	Waters Xterra RP-18, 4.6 x 50 mm, 3.5 μ
Column temperature	35.0 ± 2.0 °C
Flow rate	1.0 mL/min
Wavelength	260 nm
Injection volume	50 μL
Integration mode	Peak area
Run time	About 10 minutes

5.5.9 System Suitability

Five replicate injections of the working calibration standard were made and the relative standard deviation (RSD) was calculated. The %RSD of the five consecutive injections was less than 1.0%.

5.5.10 Calculation (mg/tablet)

$$\frac{A_{spl}}{A_{std}} \times P \times \frac{162.23}{462.41} \times \frac{100 - H_2O\%}{100} \times \frac{W_{std}}{50} \times \frac{5}{100} \times \frac{50}{1} \times \frac{50}{5} \quad \textit{Eq. 9}$$

where:

A_{spl} is the area response of nicotine from the sample solution;

A_{std} is the average area response of the nicotine bracketing standards;

W_{std} is the weight of nicotine bitartrate dihydrate standard, mg;

P is a purity factor of nicotine bitartrate standard, anhydrous basis;

162.23 is the molecular weight of nicotine;

462.41 is the molecular weight of nicotine bitartrate, and

$H_2O\%$ is the water content of nicotine bitartrate dihydrate determined by KF.

6. DATA ANALYSIS – HPLC

Each tablet sample solution was injected into the HPLC system in duplicate and the mean result was used for the calculation.

6.1 HPLC Method Validation

The HPLC method used to determine the nicotine content for the gum product was fully validated for linearity, accuracy, precision, specificity, sample solution stability, and robustness. Table 11 summarizes the method validation results of the HPLC procedure.

Table 11. Results of the HPLC Method Validation

Study	Results
Linearity and range (4 – 12 μg/mL)	• Correlation Coefficient: 1.0000. • Relative response factor: 100.0% to 100.0%. • RSD: 0.1% to 0.4%.
Accuracy (4 – 12 μg/mL)	• Recovery: 99.0% to 100.0%.
Precision	• RSD: 1.0%.
Specificity	• Nicotine peak from all the stressed samples was pure. • Known degradation products, diluent, placebo, and blank didn't interfere with nicotine quantitation.
Solution Stability	• Solutions were stable for 3 days under ambient conditions.
Robustness	• Different batch of column, slightly modified mobile phases.

Figures 15-19 show the nicotine UV spectrum, standard linearity plot, nicotine standard HPLC chromatogram, selectivity solution (containing nicotine and its all impurities: cotinine, myosmine, nornicotine, 1'R,2S-Nicotine-N-oxide, 1'R2S-Nicotine-N'-oxide, anabasine, and anatabine) HPLC chromatogram, and nicotine peak purity plot, respectively.

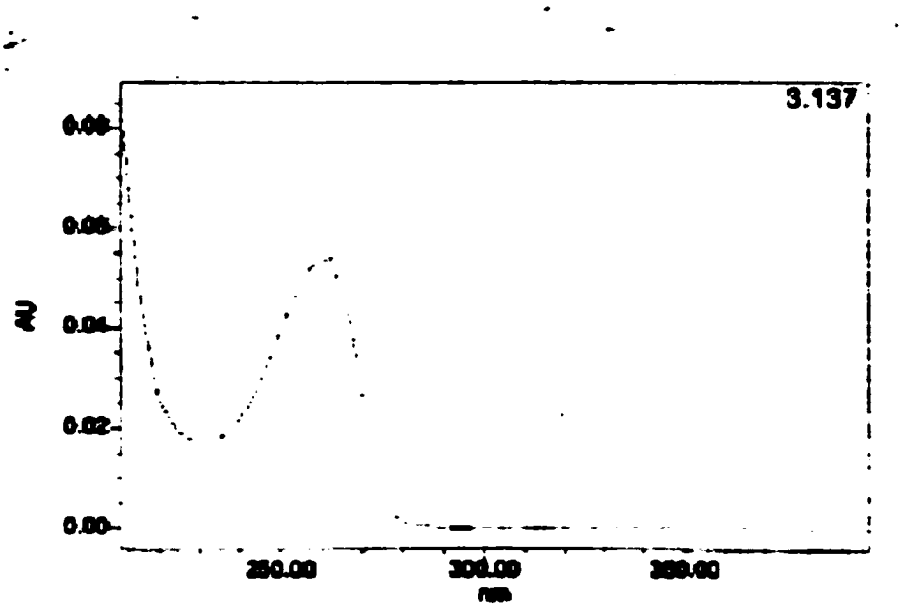

Figure 15. UV Spectrum of Nicotine in Mobile Phase

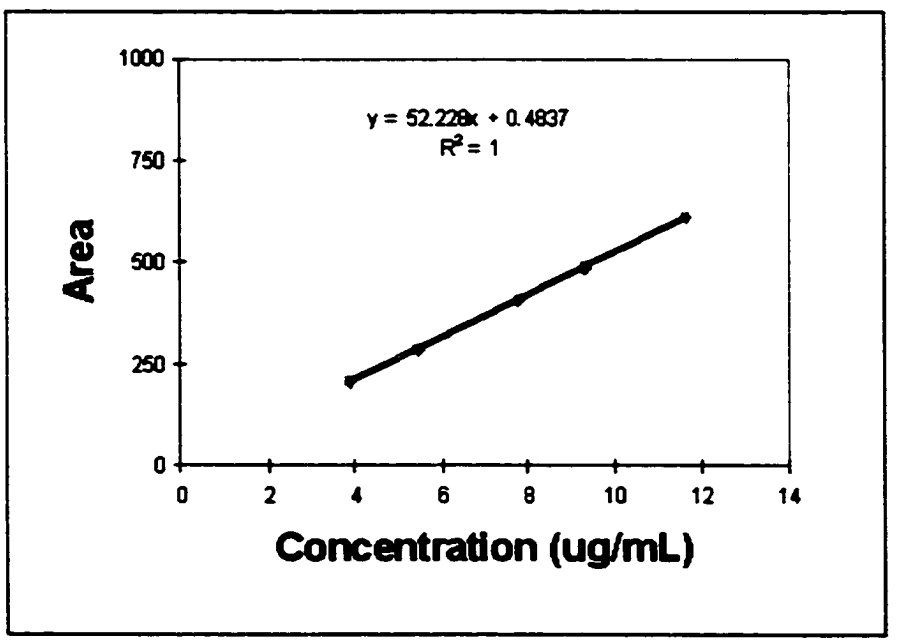

Figure 16. Linearity Plot of Nicotine Standard Solution at 260 nm

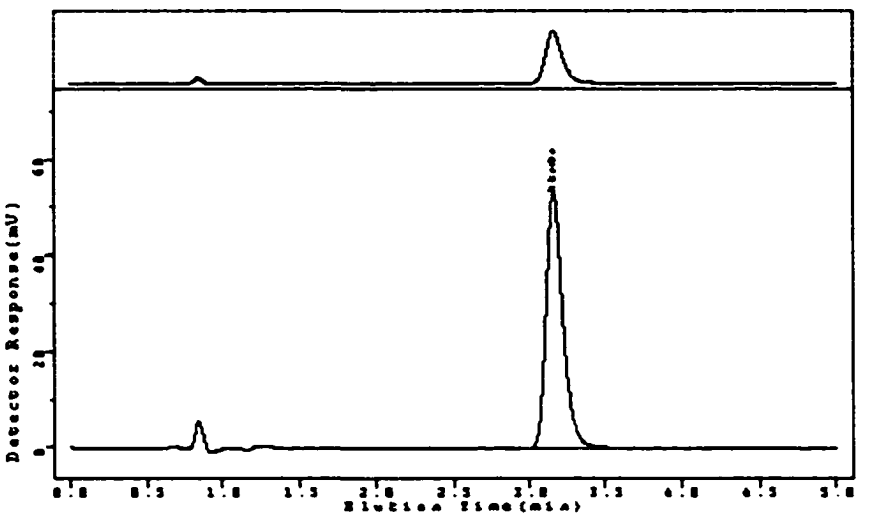

Figure 17. HPLC Chromatogram of Nicotine Standard Solution at 260 nm

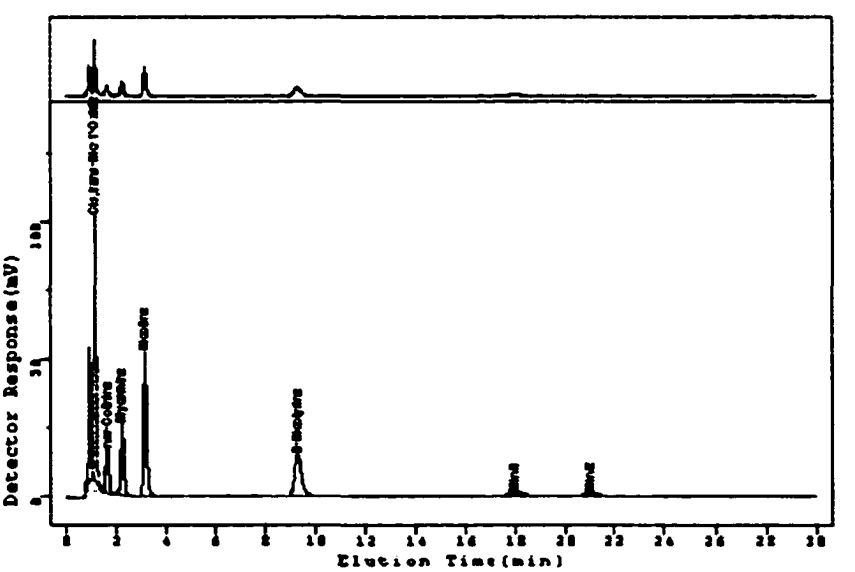

Figure 18. HPLC Chromatogram of Selectivity Solution at 260 nm

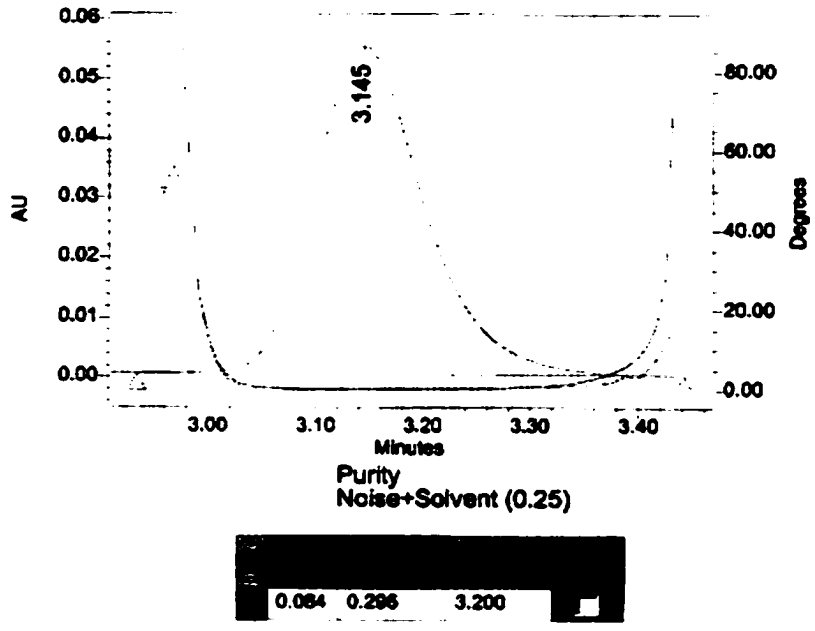

Figure 19. Nicotine Peak Purity Plot of Selectivity Solution

The nicotine peak purity curve is below the purity threshold curve and the purity angle (obtained from Waters PDA 996 detector) of 0.084 is smaller than the purity threshold of 0.296, which indicate that no impurities or degradants (with different UV spectrum) co-eluted with nicotine at about 3 minutes (details in: Waters 996 PDA Detector Operator's Guide, Milford, MA). The validation data and figures demonstrated that the HPLC procedure is linear, specific, precise, accurate, and robust.

6.2 Results

Tables 12 and 13 summarize the HPLC results of gum samples previously scanned by transmittance and reflectance NIR, respectively.

Table 12. HPLC Results for Tablets Scanned Using Transmittance NIR

Sample N0.	Batch 1	Batch 2	Batch 3	Batch 4
1	3.47	4.05	4.07	4.52
2	3.44	3.94	4.44	4.53
3	3.47	3.85	4.20	4.45
4	3.22	4.02	4.32	4.43
5	3.38	3.95	4.02	4.50
6	3.41	3.97	4.10	4.54
7	3.35	3.96	4.10	4.38
8	3.25	4.04	4.04	4.51
9	3.23	3.92	4.23	4.53
10	3.24	3.97	4.05	4.53
11	3.41	3.97	4.14	4.44
12	3.43	3.92	4.31	4.51
Min	3.22	3.85	4.02	4.38
Max	3.47	4.05	4.44	4.54
Mean	3.36	3.96	4.17	4.49
RSD%	2.9	1.4	3.2	1.1

Table 13. HPLC Results for Tablets Scanned Using Reflectance NIR

Sample N0.	Batch 1	Batch 2	Batch 3	Batch 4
1	2.88	3.90	4.20	4.53
2	3.42	3.79	4.20	4.64
3	3.47	3.84	4.18	4.60
4	3.49	3.92	4.23	4.57
5	3.49	3.86	4.25	4.61
6	3.54	3.89	4.31	4.63
7	3.41	4.01	4.21	4.53
8	3.51	3.80	4.22	4.78
9	3.39	3.98	4.33	4.60
10	3.38	3.97	4.27	4.58
11	3.32	3.94	4.28	4.53
12	3.37	4.03	4.21	4.56
Min	2.88	3.79	4.18	4.53
Max	3.54	4.03	4.33	4.78
Mean	3.39	3.91	4.24	4.60
RSD%	5.1	2.0	1.1	1.5

7. DATA ANALYSIS – NIR SPECTRA

The NIR spectra of all gum samples, the seven ingredients, and seven impurities and degradation products were collected in the range of 600 nm-2500 nm using the reflectance NIR (R-NIR) spectrophotometer. Figures 20-27 show the NIR spectra of all excipients, Figures 28-35 show all nicotine impurities, and Figure 36 shows all nicotine gum samples, respectively. Figure 37 shows all spectra together for excipients, impurities/degradation products, and gum samples, where the gum samples displayed different NIR spectra from the ingredients.

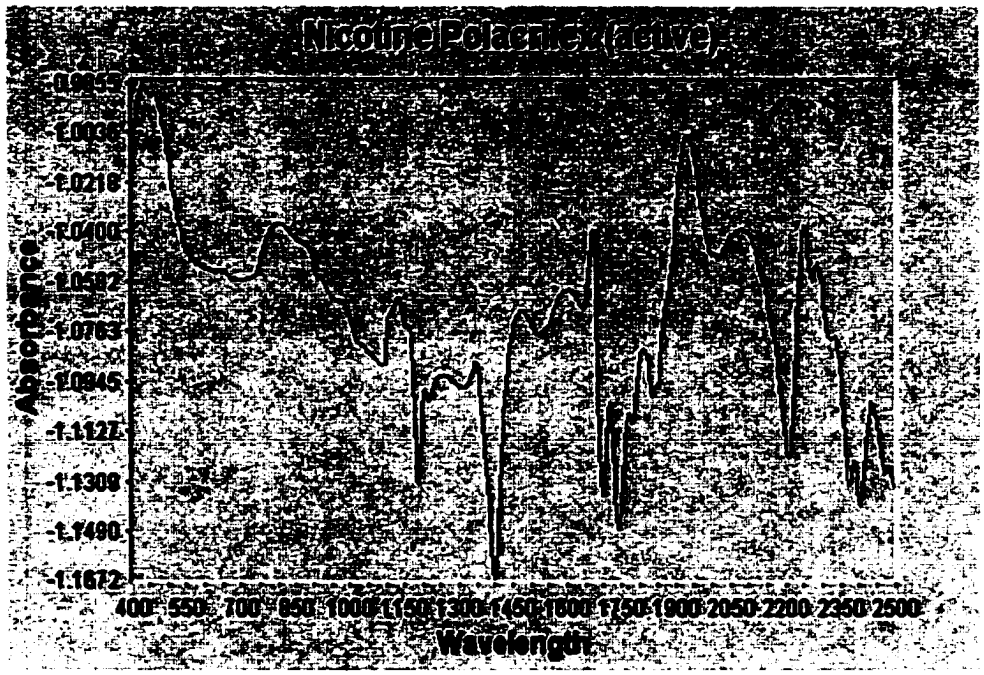

Figure 20. R-NIR Spectrum of Nicotine Polacrilex (active)

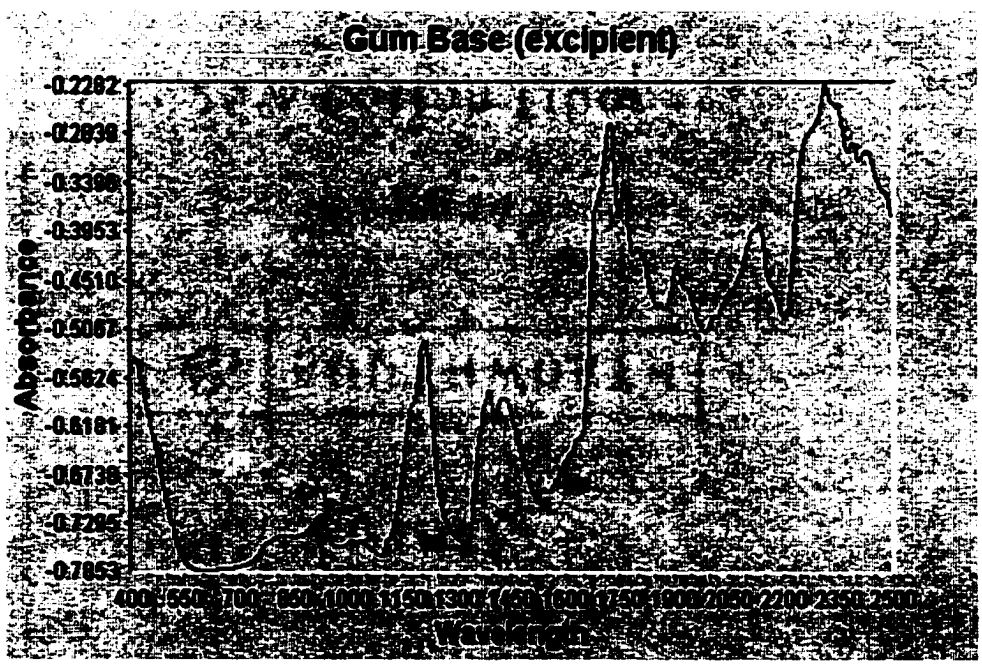

Figure 21. R-NIR Spectrum of Gum Base (excipient)

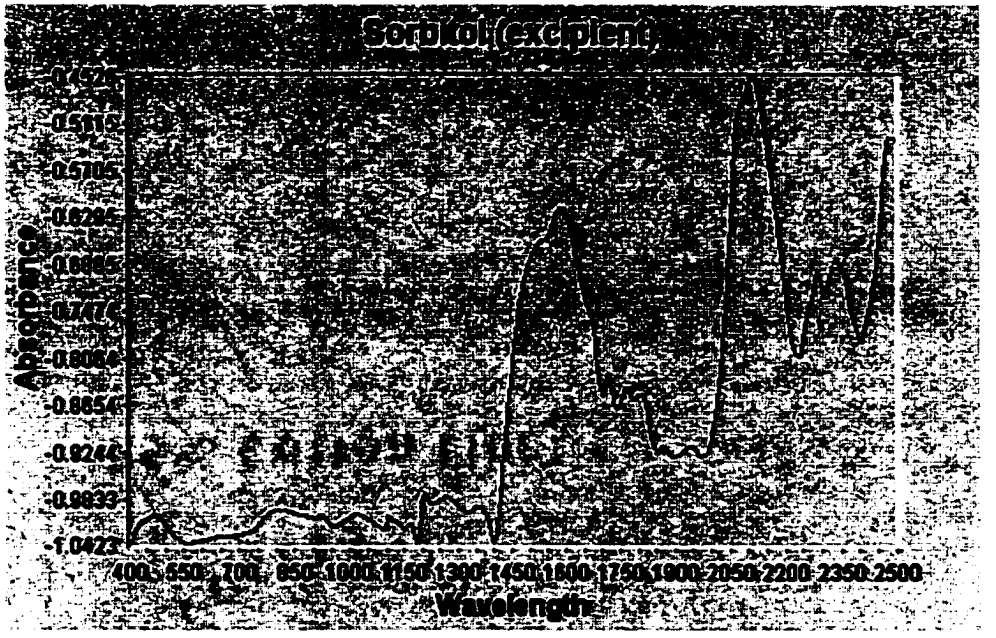

Figure 22. R-NIR Spectrum of Sorbitol (excipient)

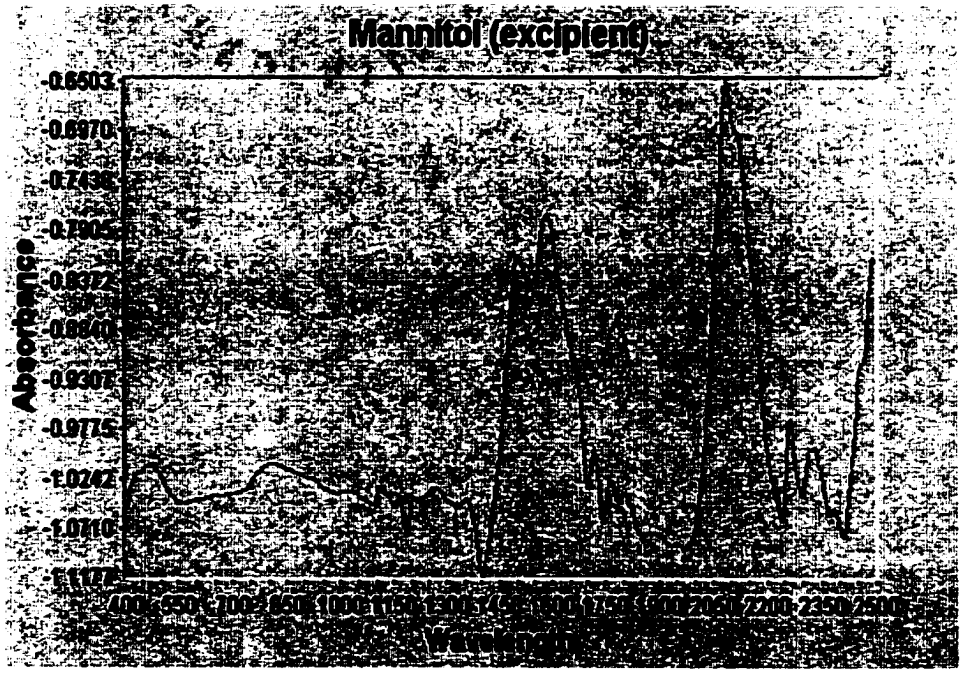

Figure 23. R-NIR Spectrum of Mannitol (excipient)

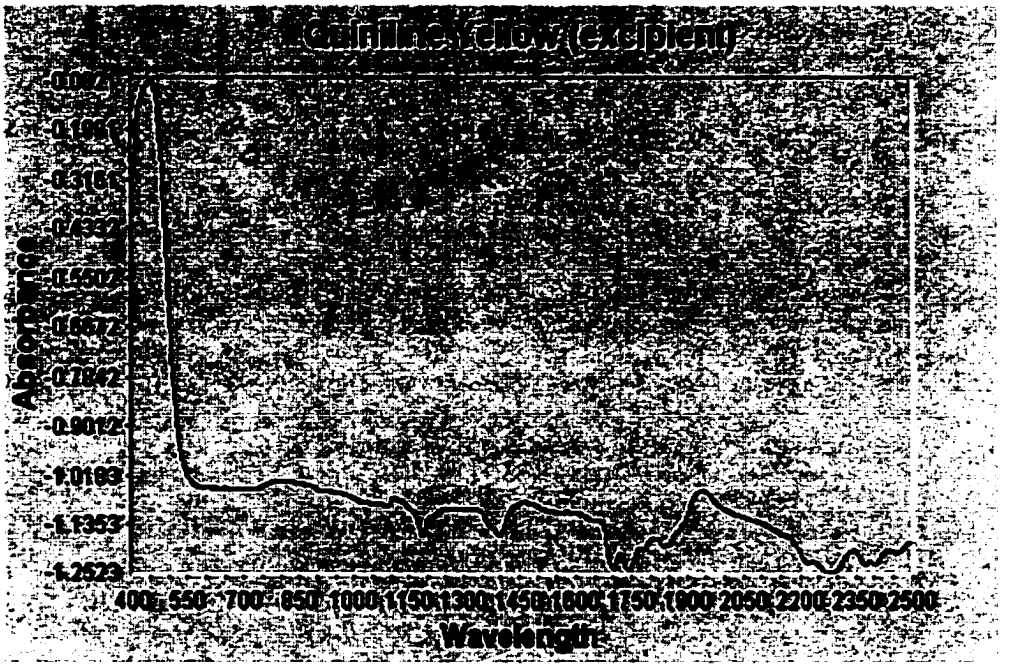

Figure 24. R-NIR Spectrum of Quiniline Yellow (excipient)

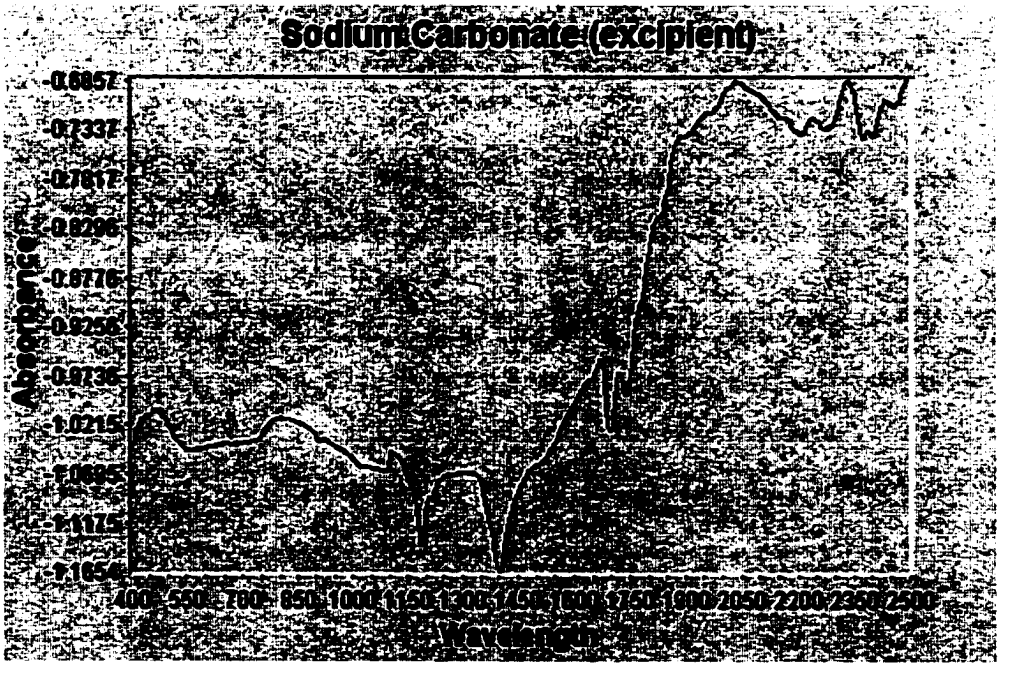

Figure 25. R-NIR Spectrum of Sodium Carbonate (excipient)

Figure 26. R-NIR Spectrum of Talc (excipient)

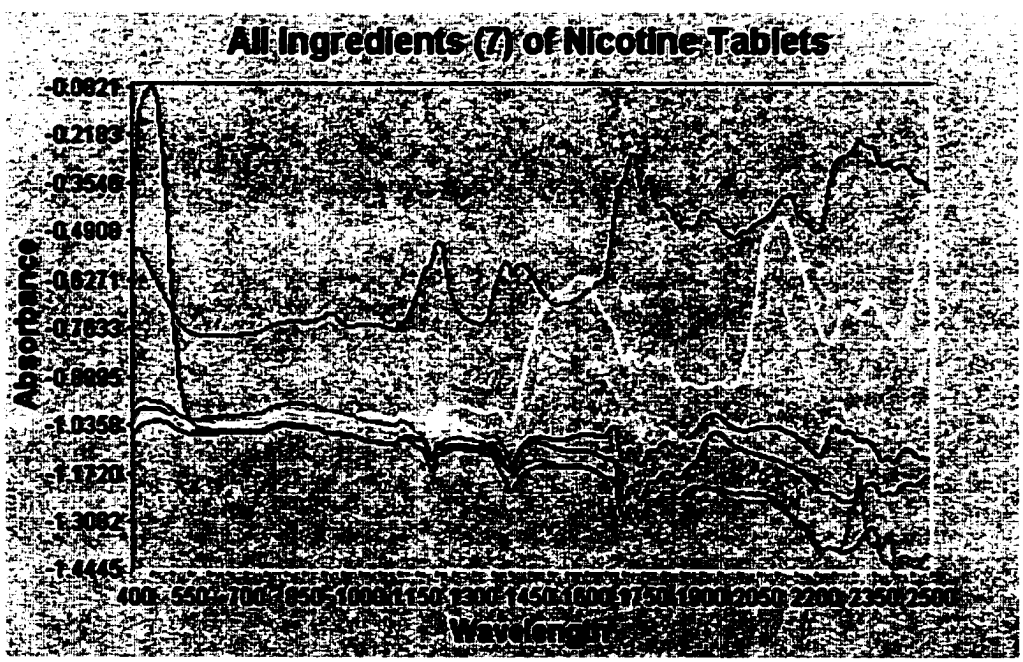

Figure 27. R-NIR Spectra of All Tablet Ingredients (7)

Figure 28. R-NIR Spectrum of Cotinine (impurity)

Figure 29. R-NIR Spectrum of Myosmine (impurity)

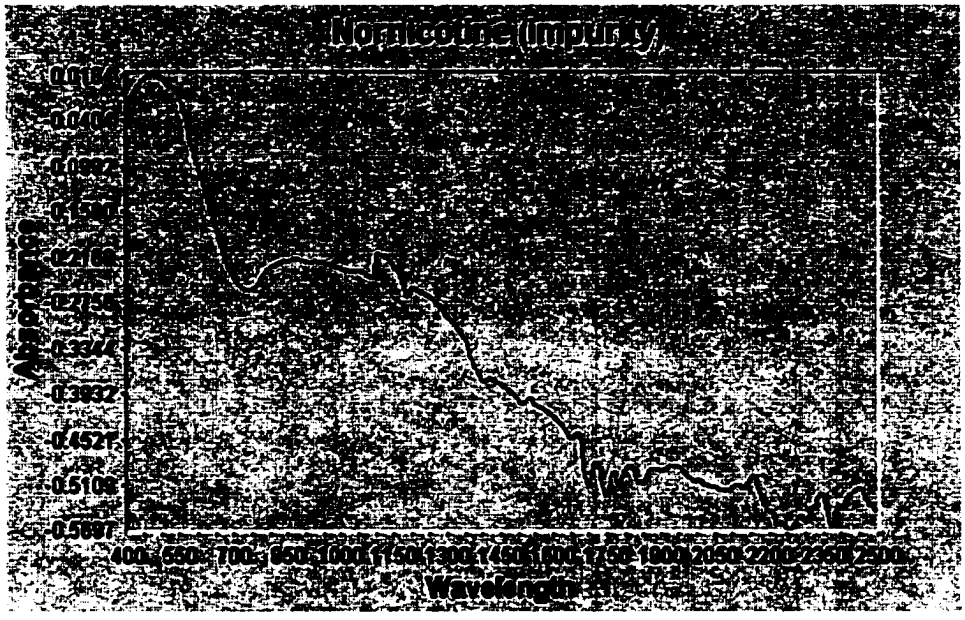

Figure 30. R-NIR Spectrum of Nornicotine (impurity)

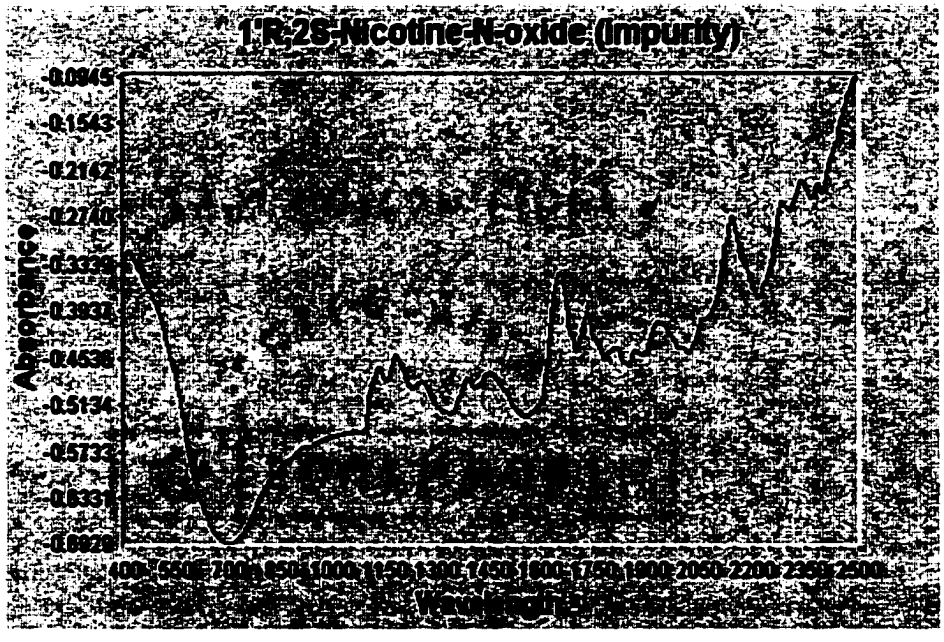

Figure 31. R-NIR Spectrum of 1'R,2S-Nicotine-N-oxide (impurity)

Figure 32. R-NIR Spectrum of 1 'R,2S-Nicotine-N'-oxide (impurity)

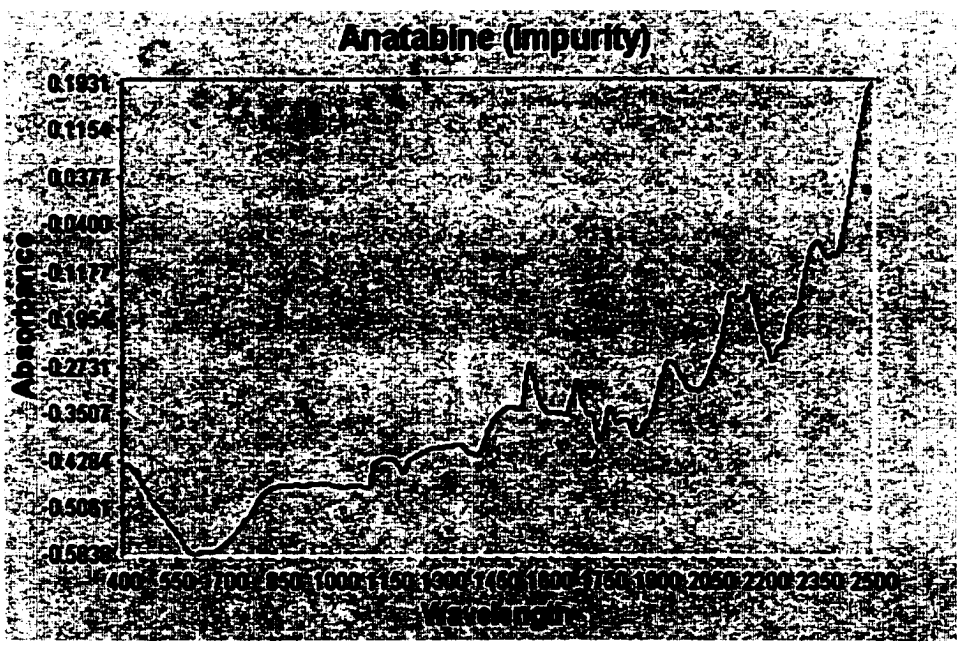

Figure 33. R-NIR Spectrum of Anatabine (impurity)

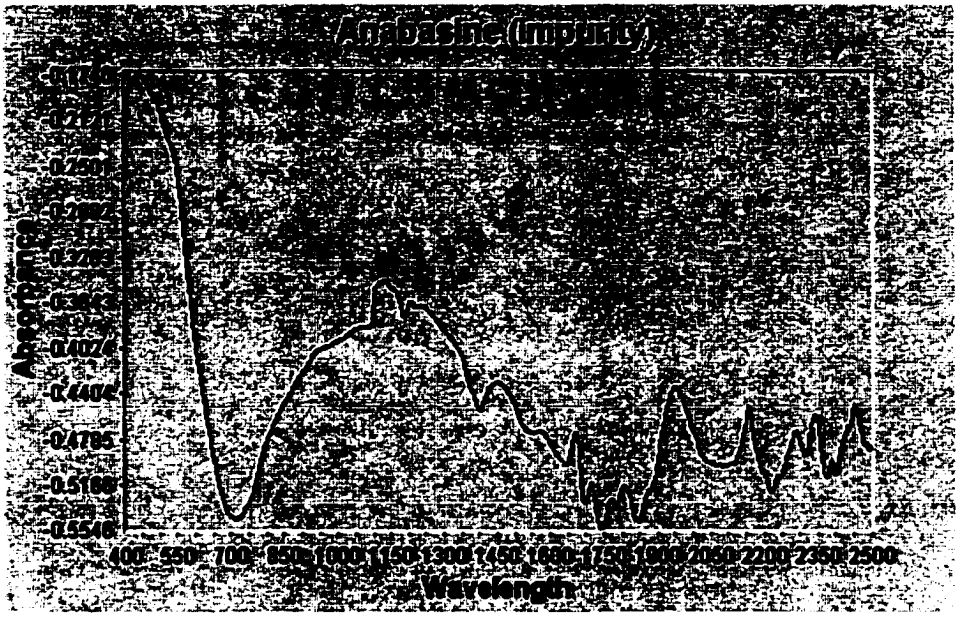

Figure 34. R-NIR Spectrum of Anabasine (impurity)

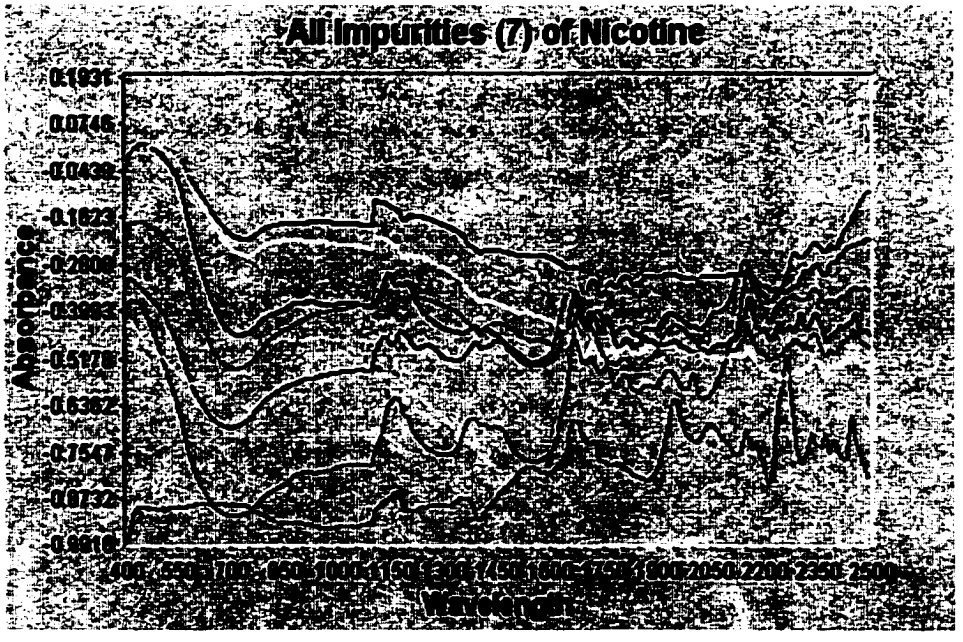

Figure 35. R-NIR Spectra of All Nicotine Impurities (7)

Figure 36. R-NIR Spectra of All Nicotine Gum Samples (48)

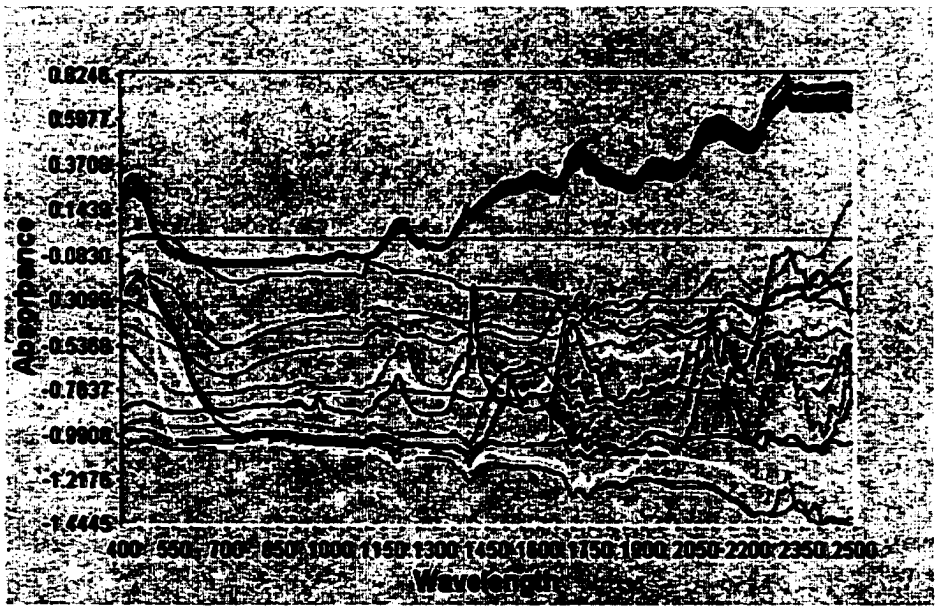

Figure 37. R-NIR Spectra of All Ingredients, Impurities, and Gum Samples

According to the characteristic wavelengths listed in Table 1, the weak peaks at abut 920 nm represent the 3rd overtones of CH vibrations corresponding to =CH, –CH2, and –CH3; the peaks at about 1200 nm represent the second overtones of CH vibrations corresponding to =CH, –CH2, and –CH; the peaks at about 1700 nm represent the first overtones of CH vibrations corresponding to =CH, –CH2, and –CH3; the peaks at about 1950 nm are combinations of O-H (H_2O) vibrations; and the peak at about 2150 nm are the combinations of C-H vibrations.

8. DATA ANALYSIS – QUANTITATIVE NEAR INFRARED

When light interacts with the sample, the amount of reflected or transmitted energy received at the detector is dependent on both the chemical (molecular absorbance) and physical (scattering/reflective) properties of the sample as well as the optical measurement geometry.

There are two kinds of data analyses that can be achieved with NIR spectra. The first is quantitative analysis in which NIR spectra are correlated to a quantified chemical component of interest. Quantitative NIR has several advantages. It is simple, fast, non-destructive, and information rich. More than one constituent can be determined from the same scan and the technology is also applicable to on-line analysis. The second is qualitative analysis[70], which involves the classification of sample properties into two or more groups according to the differences between their NIR spectra. The sample consists of a large number of molecules that absorb/reflect energy differently at different wavelengths creating a pattern of peaks and valleys. However, most NIR bands of natural materials are broad and overlapping and are not readily assigned to a single constituent. Furthermore, scattering effects in reflectance measurements may cause further uncertainties. Therefore, a statistical

approach to calibration development is needed in order to handle these complexities.

Chemometrics[71] is the use of statistical and mathematical techniques to analyze chemical data. Simple statistical analysis involves the analysis of data sets with only one variable, for example a set of assay values obtained by HPLC. The statistics describing this kind of data is called univariate or univariable. Spectroscopic measurements by their fundamental nature yield data with many variables. The statistical tools used to describe this kind of measurements are multivariate statistics.

Modern multivariate statistics were first developed for social science applications by Hotelling in 1933. These approaches were first applied to chemical and spectroscopic data by Karl Norris in the 1960's. With the development of affordable high-speed desktop computers, NIR spectroscopy has become an important area of analytical chemistry. Of particular importance for the NIR spectroscopy of pharmaceutical products was the development of partial least squares regression (PLSR) methods by Joreskog and Worl[72]. Most quantitative NIR analyses of pharmaceuticals are now done using PLSR[73,74].

8.1 Simple Linear Regression (SLR)

The simplest method of calibration is based on a single independent variable (wavelength). The assumption is that constituent values c (usually concentrations) can be expressed as a linear function of absorbance A at some wavelength i:

$$\boxed{c = k(0) + k(1) \cdot A_i} \qquad \textit{Equation 10}$$

This expression contrasts with the Beer-Lambert Law, which gives absorbance as a function of concentration, and hence is referred to as "inverted" Beer's law. Inverted Least Squares (ILS) is presented graphically by a least squares fit of a straight line to data points in the plot of concentration versus absorbance. The least squares approach fits the line so as to minimize the sum of squares of deviations between data points and the calibration line; it yields two calibration constants: intercept *k(0)* and slope *k(1)*. The wavelength at which the calibration is performed is usually selected because of high correlation between concentration and absorbance and least sensitivity to the concentration. Sensitivity in NIR is the slope of the calibration line for concentration of analyte (y axis), versus change in optical response (x axis), for samples of varying concentration. If the sensitivity is very high, even small changes in absorbance, e.g. those caused by noise, will register as a large change in predicted concentration. On the other hand, low sensitivity indicates a robust calibration that gives stable and reliable predictions. Of course, low sensitivity is useful only when correlation is high, for low sensitivity with low correlation would have little predictive value. Superimposed plots of correlation and sensitivity values across the full spectral range are helpful in visualizing which wavelength is best for calibration. Thus, that wavelength is best which has a correlation close to 1, as well as low sensitivity.

8.2 Multiple Linear Regression (MLR)

Multiple Linear Regression[75,76] (MLR) is an extension of the simple linear regression. This method uses information at more than one wavelength to create a calibration equation:

$$c = k(0) + \sum_{i=1}^{n} [k(i) \cdot A_i] \qquad \textit{Equation 11}$$

where *k(0)* is the intercept, *k(i)* is the slope at wavelength *i*, and A_i is the absorbance at wavelength *i*.

When excipients in the matrix interfere with the constituent absorbance, simple linear regression is no longer valid to obtain a meaningful calibration. MLR is useful when the information at a single wavelength does not yield a model that performs suitably well. It can overcome the interference caused by other constituents in that matrix and provide a valid regression. When interference requires the use of multiple wavelengths, the calibration plot is multi-dimensional and graphical presentation of the calibration is not possible.

Overfitting (including too many wavelengths in the calibration equation) and colinearity (wavelengths that are not independent of each other) are two common problems in MLR, especially as spectral overlap increases. While such a model may describe the calibration set very well, it may be very sensitive to noise or systematic errors in the calibration data which may not be representative of samples in general. Consequently, MLR models with colinearity among analysis wavelengths may not provide reliable analysis of real samples.

8.3 Partial Least Squares Regression (PLSR)

PLSR has been employed since the early 1980s[77] and is a multivariate technique showing promise in calibration where samples are complex and where low signal-to-noise relationships exist between analyte and spectral features. It is often used together with principal component analysis (PCA) and has wide-ranging applications in a variety of spectroscopic and scientific situations[78,79,80]. PLSR in multivariate calibration is covered in detail in reference[81].

8.3.1 Principal Component Analysis

Principal component analysis (PCA) can be traced as far back as 1901[82] and was developed into its modern concept by Hotelling in 1933. The American Society For Testing and Materials defines PCA as "a mathematical procedure for resolving sets of data into orthogonal components whose linear combinations approximate the original data to any desired degree of accuracy. As successive components are calculated, each component accounts for the maximum possible amount of residual variance in the set of data. In spectroscopy, the data are usually spectra, and the numbers of components is smaller than or equal to the number of variables or the number of spectra, whichever is less"[83].

The goal of PCA is to use a small number of principal components (PCs) (also called factors) to represent the variations presented by many variables in the original raw data. The mathematical expression is[84]:

$$[R] = [U][S][V]^T \qquad \textit{Equation 12}$$

where *[R]* is the response matrix, *[U]* is the score matrix, $[V]^T$ is a transpose of the loading matrix *[V]*, and *[S]* is a diagonal matrix (has zero everywhere except on the diagonal). The rows of *[R]* contain the measurements vectors for individual samples and the column contain the variables (wavelengths). This matrix has the dimensions n-sample x n-variable. The *[U]* matrix contains the coordinates of the samples along the new PC axes. The *[V]* matrix contains the information about how the original measurements are related to the new PCs. The *[S]* matrix contains the information about the amount of variance each PC describes. All PCs have the following properties:

1. The first PC explains the maximum amount of variation possible in the data set in one direction. In another word, it is the direction that describes the maximum spread of data points. The percent of the total variation in the data set described by any PC can be precisely calculated.
2. Each sample has coordinates with respective to the new PC axes just like it has coordinates in the original row space. The coordinates of the samples relative to PC axes are typically called "scores".
3. Each PC is constructed from combinations of the original measurement variables and can more efficiently describe the variation in the data set. The extent to which a measurement variable contributes to a PC depends on the relative orientation in space of the PC and variable axes. In mathematical terms, the contribution of each axis to a PC is the cosine of the angle between the variable axis and the PC axis. The cosine values are often called "loading", ranging from –1 to 1.
4. Noise can be filtered out from a data set by excluding the nonsignificant PCs because nonsignificant PCs typically describe more noise than sample signal.
5. PCs are orthogonal. The maximum number of PCs is the smaller of the number of samples or variables. Determining the number of PCs to use is the main challenge of principal component analysis.

Principal components are considered superior to the optical data that are directly produced by the NIR instrument since noise is rejected from the initial principal components produced. These initial principal components represent true sources of variation of the spectra, presumably caused by real physical phenomena.

PCA has two principal advantages: (a) Lack of requirement for wavelength selection: One of the main advantages of using the principal component approach to calibration is that there is no need to perform a wavelength search since all the wavelengths are included and (b) robustness: the orthogonality of the principal components combined with the fact that collectively they account for the maximum amount of variance in the data makes the use of principal components an inherently robust method of calibration. Calibration based on principal components prevent the problems associated with calibrations based on the highly correlated optical data.

PC can also be used to calculate Mahalanobis distance[85] and Euclidean distance. The Mahalanobis distance is the distance between a spectrum and the center of the distribution of a set of spectra represented by a covariance matrix *[C]* as:

$$D_M^2 = (A - \mu)^T [C^{-1}](A - \mu) \qquad \textit{Equation 13}$$

where μ denotes the mean spectrum of the distribution. The spectrum A may or may not belong to the calibration set. Calculating the Mahalanobis distance for a spectrum in this way has disadvantages. An alternative, simpler approach calculates Mahalanobis distance from primary principal component scores (s) after secondary PC's have been rejected from the model. The Mahalanobis distance in PC space is written as:

$$D_{M,i}^2 = (n-1)^2 \sum_{j=1}^{k} [s_{ij}^2 \sum_{i=1}^{n} s_{ij}^2] \qquad \textit{Equation 14}$$

where k is the number of PC's used, i and j are indices, and n is the number of samples in the calibration set.

In addition, PC can be used to calculate the Euclidean distance (ref. 85). The Euclidean distance is the distance between two objects (e.g. spectra) x and y with n coordinates (PC scores) and is calculated according to formula:

$$D_E = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad \textit{Equation 15}$$

8.3.2 Partial Least Squares Regression

Partial Least Squares regression (PLSR) is a method that allows use of many wavelengths – broad segments or an entire spectrum – while avoiding the problem of colinearity that besets MLR. It is the method most commonly used to quantify constituents in complex systems such as slurries and tablets. PLSR has several critical advantages of traditional methods for quantification of spectra. A very good discussion of PLSR can be found in reference[86].

PLSR is an inverse method and solves inversion problem using principal component analysis – replacing the original variables with linear combinations of the PCs or factors. The mathematical expression of inverse method is:

$$c_v = [R]b_v \qquad \textit{Equation 16}$$

where c_v is the vector contains the concentrations, *[R]* is the response matrix, and b_v is the vector contains regression coefficients (or weights).

With the known c_v and *[R]* from the calibration set, Equation 16 can be solved as:

$$\boxed{b_v = ([R]^T[R])^{-1}[R]^T c_v = [R]^* c_v} \quad \textit{Equation 17}$$

where *[R]** is called the pseudo-inverse of *[R]*.

Therefore Equation 16 can be used to estimate the concentrations of unknown samples, c_v, with known *[R]* and b_v. However the $[R]^T[R]$ matrix in Equation 17 is not invertible because of the redundancy in the variables. PCA eliminates the redundancy by constructing a new matrix *[U]* with columns that are linear combinations of the original columns in *[R]*. Equation 17 now can be expressed as:

$$\boxed{b_v = ([U]^T[U])^{-1}[U]^T c_v = [U]^* c_v} \quad \textit{Equation 18}$$

where $[U]^T[U]$ is invertible.

Matrix *[U]* can be found by solving Equation 12 as:

$$\boxed{[U] = [R][V][S]^{-1}} \quad \textit{Equation 19}$$

With the *[V]*, *[S]*, and b_v obtained from the calibration set, the concentration of unknown can be predicted as:

$$\boxed{c_v = [R]([V][S]^{-1}b_v)} \quad \textit{Equation 20}$$

where *[V]* contains information from both response matrix *[R]* and concentration vector c_v.

Because PLSR uses a range of wavelengths for quantitation, it can analyze multiple components simultaneously, reduce the standard deviation in a factor of $n^{1/2}$, identify the interference, and detect the small changes in the responses.

9. DATA ANALYSIS – TRANSMITTANCE NIR

Figure 38 shows the transmittance NIR (T-NIR) spectra of all nicotine tablets used in calibration and validation sets (total 48 samples) in the wavelength range of 600 nm – 1900 nm.

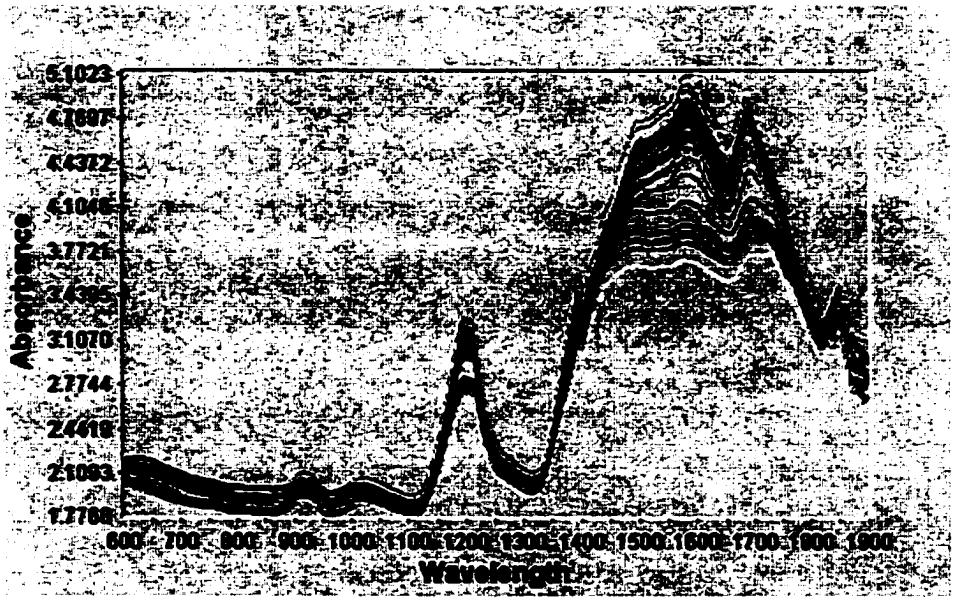

Figure 38. T-NIR Spectra of Tablets in Range of 600 nm – 1900 nm

Because the fiber optic produces significant noise above 1400 nm, and the water content of the gum sample adversely affects quantitation near 1400 nm (the 1st overtones O-H vibrations), the wavelength range of 900 nm – 1300 nm was selected for analysis. Figure 39 shows the NIR

spectra of all gum samples (48) in the range of 900 nm –1300 nm. The peaks at about 1200 nm are corresponding to the second overtones of the =CH, -CH2, and –CH3 vibrations.

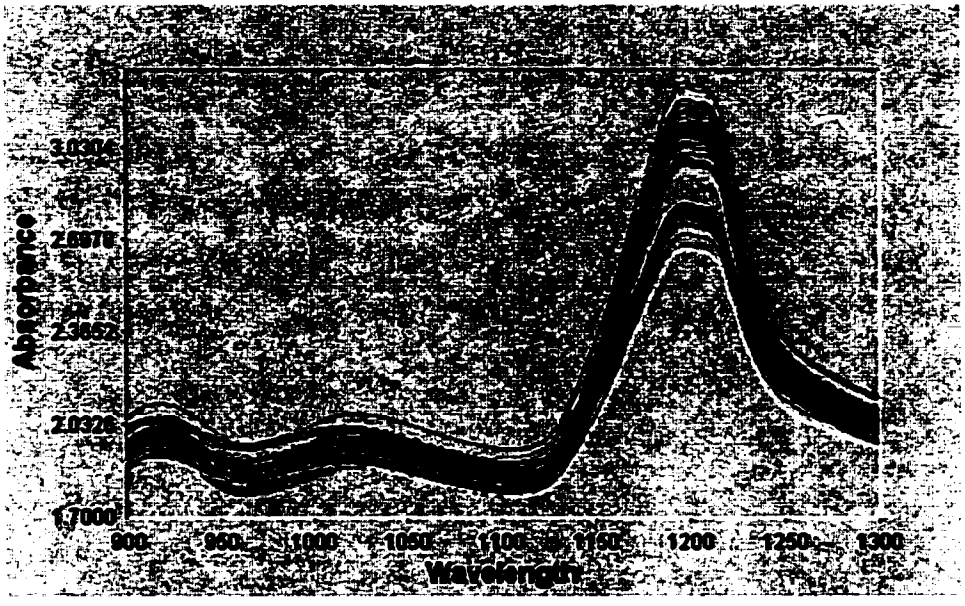

Figure 39. T-NIR Spectra of all Tablets in Range of 900 nm – 1300 nm

9.1 Sample Selection

The selection process divides the samples into a calibration set (also known as training set), validation set (also known as redundant samples), and rejection set (also known as outliers). Redundant samples are defined as the samples carry no distinguished information and locate in the similar space to others. The true spectral outliers are considered to be samples whose spectral characteristics are not represented within a specified sample set. Outliers are not considered to be part of the group that is designed to be used as a calibration set. The criterion often given representing outlier selection is a sample spectrum with a distance of greater than three Mahalanobis distances from the center of the data.

To reduce the spectral complexity and environmental effects, it is often necessary to mathematically treat the raw data, but care must be taken to not lose important information during the treatment. Figure 40 is the 2nd derivative of the NIR spectra of all nicotine gum samples (48) in the range of 900 nm – 1300 nm.

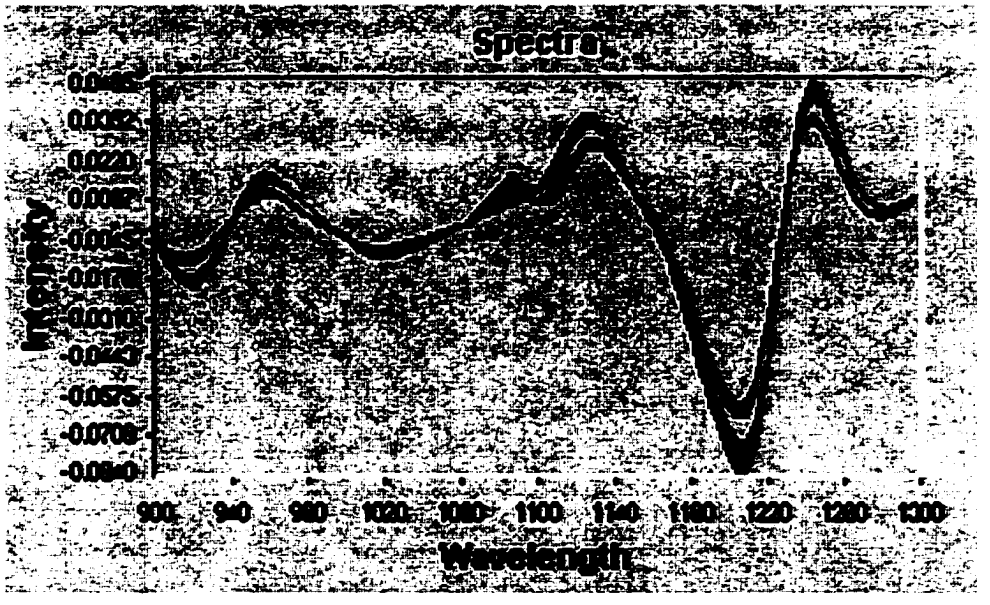

Figure 40. The 2nd Derivative of T-NIR Spectra of all Tablets

Three sample selection methods (random, Mahalanobis Distance in Principal Component Space, and Maximum Distance in Wavelength Space) were investigated and results are discussed below. The second derivative mathematical treatment was used in the last two selection methods.

9.1.1 Random Selection

In the Random Selection method, samples were selected randomly for the calibration and validation sets. No rejection set is created and no mathematical treatment is necessary.

9.1.2 Mahalanobis Distance in Principal Component Space

The selection process first calculates a principal component. All samples with Mahalanobis distances from the center of the distribution greater than the threshold, 3.0, were flagged as outliers. Samples located in high-density regions of the population (nearest neighbors) were identified as redundant in such a way that the Euclidean distances between sample PCA scores in the calibration set was greater than the threshold, 0.6. The values, 3.0 and 0.6, are not statistically meaningful and have been established using scale factor by the instrument manufacturer.

Figures 41 and 42 present the plots of the sample clusters (Mahalanobis distances of the samples in the 3-dimensional plot) and Mahalanobis distance vs. sample frequencies. The darker color represents the samples used in the calibration set and the lighter color represents the samples used in the validation set. No samples were identified as outliers and all the tablet spectra were used in calibration set or validation set, respectively.

Figure 41. Sample Clusters of Mahalanobis Distance in 3-Dimensions

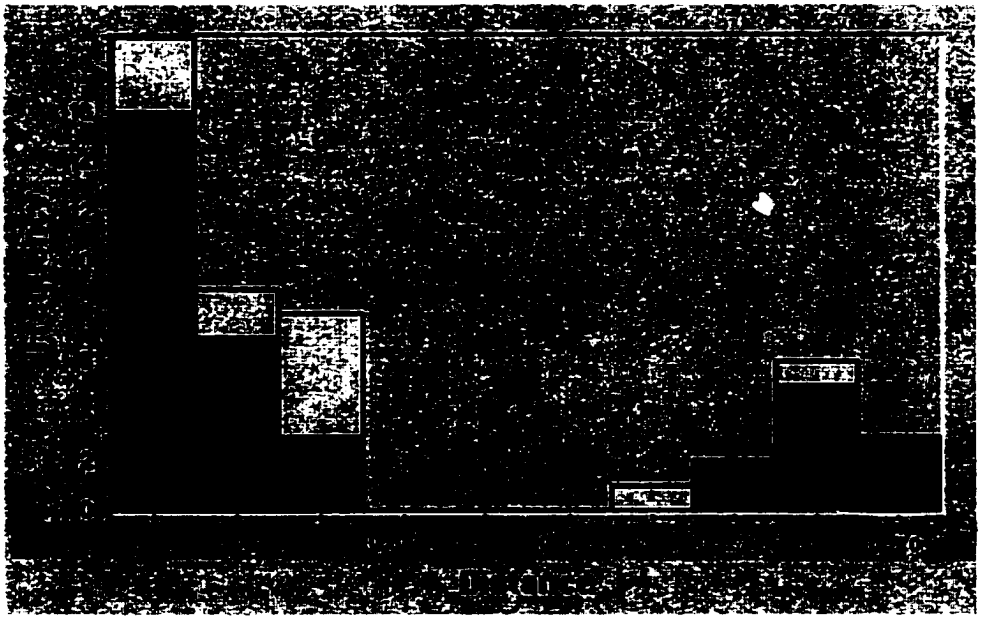

Figure 42. Mahalanobis Distance vs. Frequencies of all Gum Samples

9.1.3 Maximum Distance in Wavelength Space

Selection by wavelength distance uses a maximum distance concept (maximum conformity index) to identify outliers. Samples with the maximum distance from the mean product spectrum greater than the threshold, 3.0, were placed in the outlier set. Euclidean distance in the wavelength space was used to select the redundant samples.

To calculate maximum distance (distance in this context is in terms of spectral value, the absorbance) of a spectrum y to a mean product spectrum x, the inflated standard deviation spectrum from the product set was calculated to accommodate the sometimes limited number of samples available for identification:

$$s_i^d = [1 + \frac{1}{\sqrt{2(n-1)}}][\frac{\sum_j (x_i - \bar{x}_j)^2}{n-1}]^{1/2} \qquad \textit{Equation 21}$$

where the index i runs over all wavelengths, and index j runs over all spectra in the set.

Then, the maximum distance is calculated from:

$$D_x = \max \left\{ abs\,[\frac{y_i - \bar{x}_i}{s_i^d}] \right\}_{over\ all\ i} \qquad \textit{Equation 22}$$

Maximum distance can be interpreted as the maximum deviation from the product mean spectrum expressed in units of inflated standard deviation.

Figure 43 is the plot of maximum distance of all gum samples. The darker color represents the samples used in the calibration set and the lighter color represents the samples used in the validation set. No outliers were detected.

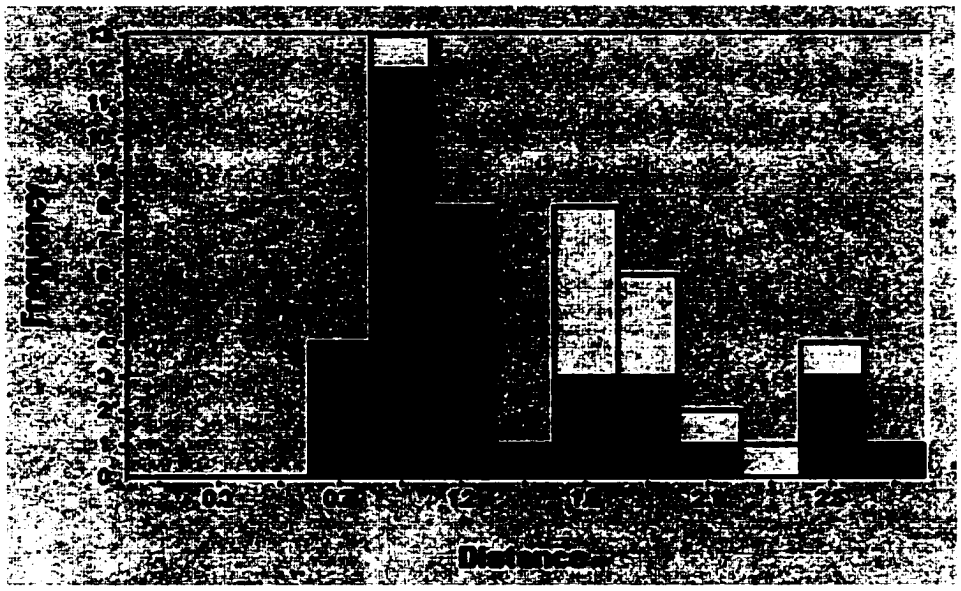

Figure 43. Maximum Distance in Wavelength Space of All Tablets

9.1.4 Discussion

Different selection methods can be used to identify outliers and redundant samples for different applications.

In general, random selection is used if no samples should be treated as outliers and all the samples should be included in the calibration or validation sets. On the other hand, Mahalanobis distance and maximum distance can be used to identify the outliers and redundant samples. The Mahalanobis distance method with principal component analysis is more sophisticated and may detect outliers which cannot be

detected using the maximum distance method; therefore, it was used in rest of this research.

In addition, different mathematical methods besides the 2nd derivative can be used to treat the raw spectra prior to sample selection process to remove the noise effect, etc. Further discussion on different mathematical pretreatments is presented in Section 9.2.3.

9.2 Partial Least Squares Regression

Deciding how many factors to be used is an important part of the PLSR calibration. With too few factors, the calibration includes too little information and gives correspondingly high prediction errors. When too many factors are used, the model overfits the calibration data (noise or systematic errors unique to the calibration are included in the model), resulting in a model that is not robust or stable. The PRESS (Prediction Residual Error – Sum of Squares)[87], SEC (Standard Error of Calibration), SEP (Standard Error of Prediction), F-test for regression, and coefficient of multiple determination (R^2) are used to determine the appropriate factors to be used.

Standard Error of Calibration (SEC) is a statistical parameter that indicates the upper limit of accuracy because the SEC value can not be better than the standard deviation of the lab value that obtained from the primary method such as HPLC. When the calibration equation is applied to the calibration set itself, the SEC is calculated from residuals f as:

$$SEC = \sqrt{\frac{\sum f_i^2}{N - M - 1}} \qquad \textit{Equation 23}$$

where f_i is the residuals obtained from the calibration samples (fi = lab value – estimated value from a given regression line), N is the number of samples, and M is the number of wavelengths or factors.

F value (F-test) for regression can be expressed as[88]

$$F = \frac{R^2(N - M - 1)}{M(1 - R^2)} \qquad \textit{Equation 24}$$

where N is the number of samples, M is the number of wavelengths or factors, and R^2 is the multiple correlation coefficient. The F value is a useful estimate of goodness of fit of spectral and constituent data. It can be also used as a tool for evaluation of how many wavelengths or factors should be used in an equation, and for determining which samples to eliminate as outliers from the calibration set.

F increases as the equation begins to model. With R^2 held constant, the F value increases as the number of samples increases. As the number of wavelengths used within the regression equation decreases, F tends to increase. Deleting an unimportant wavelength from an equation will cause the F for regression to increase.

The F can also be useful in recognizing suspected outliers within a calibration sample set; if the F value decreases when a sample is deleted, the sample was not an outlier. This situation is the result of the sample not affecting the overall fit of the calibration line to the data while at the same time decreasing the number of samples *(N)*. Conversely, if deleting a single sample increases the overall F for regression, the sample is considered a suspected outlier.

Multiple correlation coefficient (R^2) is a measure of how well the spectral data fit the constituent values for a non-linear relationship. This statistical quantity, also called multiple correlation coefficient, is equal to 0 when spectral response is unrelated to constituent data (the relationship is statistically random). A value of 1 signifies that the constituent values fit spectral data perfectly and all residuals are equal to zero. $R^2 = 1.00$ indicates the calibration equation models 100% of the variation within the data. It can be proved that $R^2 = r^2$ (r is correlation coefficient) when the regression is linear. The R^2 is calculated from:

$$R^2 = \frac{\sum_{i=1}^{N} (x_i' - \bar{x})^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \qquad \textit{Equation 25}$$

where x' is the estimated x value by a given regression line.

The PRESS for each factor was calculated, and the factor having the minimum PRESS value was then recommended. However, the loading plot, F value, and SEC were also used to help determining the number of the factors used in the calibration set since using only PRESS to select the number of factors may mislead the calibration development. The correlation coefficient plot was presented to help to determine the wavelength range that might be used.

The optimum factors to be used should have the least PRESS with fewer factors, larger F value, small SEC (but not significantly smaller than the SD of HPLC data). Each curve in the loading plot represents a primary PC (or factor) therefore the loading plot can be useful to help estimating the number of factors which should be used.

9.2.1 Effect of Sample Presentation

A wavelength range of 900 nm – 1300 nm and mathematical pretreatment of the 2nd derivative with segment size 20 were used in this section.

9.2.1.1 One-way Scan

The gum samples were scanned with one presentation. Tables 14 and 15 are the summary of statistical results and NIR vs. HPLC data of the calibration set, respectively.

Table 14. Statistical Results, PLSR Calibration, One-way Scan

Factor	R^2	SEC	PRESS	F-value
1	0.8351	0.1831	1.2389	172.1
2	0.8849	0.1552	0.9533	126.9
3	0.9481	0.1059	0.6551	194.8
4	0.9600	0.0945	0.5353	185.8
5	0.9684	0.0853	0.3572	184.0
6	0.9771	0.0739	0.2342	206.2
7*	0.9822	0.0662	0.1994	220.9
8	0.9841	0.0638	0.2712	208.7
9	0.9857	0.0617	0.3194	198.8
10	0.9883	0.0568	0.2752	211.8
11	0.9903	0.0527	0.3330	223.9
12	0.9908	0.0525	0.4318	207.3
13	0.9913	0.0523	0.4525	192.8
14	0.9923	0.0504	0.5619	193.1
15	0.9928	0.0498	0.5636	184.2

**The number of factors used in the calibration set.*

Table 15. NIR vs. HPLC Data, PLSR Calibration, One-way Scan

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-01	3.45	3.47	-0.02
2	1-02	3.38	3.44	-0.06
3	1-03	3.46	3.47	-0.01
4	1-04	3.27	3.22	0.05
5	1-05	3.48	3.38	0.10
6	1-06	3.37	3.41	-0.04
7	1-07	3.37	3.35	0.02
8	1-08	3.32	3.25	0.07
9	1-09	3.16	3.23	-0.07
10	1-10	3.38	3.24	0.14
11	1-11	3.31	3.41	-0.10
12	1-12	3.37	3.43	-0.06
13	2-01	3.94	4.05	-0.11
14	2-02	3.89	3.94	-0.05
15	2-05	4.00	3.95	0.05
16	2-06	4.04	3.97	0.07
17	2-07	3.96	3.96	0.00
18	2-09	3.96	3.92	0.04
19	2-11	3.97	3.97	0.00
20	3-03	4.17	4.20	-0.03
21	3-04	4.24	4.32	-0.08
22	3-05	4.07	4.02	0.05
23	3-06	4.11	4.10	0.01
24	3-07	4.10	4.10	0.00
25	3-08	4.11	4.04	0.07
26	3-09	4.24	4.23	0.01
27	3-10	4.05	4.05	0.00
28	3-11	4.18	4.14	0.04
29	3-12	4.29	4.31	-0.02
30	4-01	4.53	4.52	0.01
31	4-02	4.45	4.53	-0.08
32	4-03	4.47	4.45	0.02
33	4-04	4.52	4.43	0.09
34	4-05	4.47	4.50	-0.03
35	4-09	4.53	4.53	0.00
36	4-10	4.47	4.53	-0.06
Mean	N/A	3.92	3.92	SAR: 1.66
SD	N/A	0.44	0.44	

N/A: Not Applicable or Not Available in this research.
SAR: Sum of Absolute Residuals.

9.2.1.2 Four-way Scan

Each gum sample was scanned with four presentations by rotating it 180° respect to x and y axis, respectively. Tables 16 and 17 are the summary of statistical results and NIR vs. HPLC data of the calibration set, respectively.

Table 16. Statistical Results, PLSR Calibration, Four-way Scan

Factor	R^2	SEC	PRESS	F-value
1	0.7495	0.2098	1.5310	101.7
2	0.8850	0.1443	0.8088	127.0
3	0.9438	0.1024	0.6069	179.1
4	0.9520	0.0961	0.4773	153.8
5	0.9686	0.0791	0.3938	185.1
6	0.9716	0.0765	0.3161	165.1
7*	0.9806	0.0643	0.2325	202.3
8	0.9836	0.0602	0.2559	202.8
9	0.9863	0.0562	0.4690	207.5
10	0.9897	0.0497	0.4791	239.4
11	0.9958	0.0324	0.4689	514.4
12	0.9967	0.0293	0.4660	576.7
13	0.9975	0.0260	0.4031	677.3
14	0.9979	0.0247	0.3846	699.8
15	0.9984	0.0221	0.4028	817.2

**Number of factors used in the calibration set.*

Table 17. NIR vs. HPLC Data, PLSR Calibration, Four-way Scan

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-01	3.43	3.47	-0.09
2	1-04	3.29	3.22	-0.03
3	1-05	3.52	3.38	0.06
4	1-06	3.35	3.41	-0.04
5	1-07	3.37	3.35	0.02
6	1-09	3.23	3.23	0.04
7	1-11	3.34	3.41	-0.06
8	1-12	3.38	3.43	0.01
9	2-01	3.96	4.05	0.09
10	2-02	3.91	3.94	0.04
11	2-03	3.94	3.88	0.03
12	2-04	3.98	4.02	-0.04
13	2-06	3.99	3.97	-0.04
14	2-07	4.00	3.96	0.07
15	2-08	3.98	4.04	-0.05
16	2-09	3.93	3.92	0.09
17	2-12	4.01	3.92	0.00
18	3-01	4.11	4.07	-0.01
19	3-02	4.47	4.44	-0.04
20	3-03	4.16	4.20	-0.01
21	3-04	4.28	4.32	-0.10
22	3-05	4.09	4.02	0.02
23	3-06	4.05	4.10	0.07
24	3-08	4.13	4.04	-0.07
25	3-09	4.23	4.23	0.05
26	3-10	4.04	4.05	-0.04
27	3-12	4.27	4.31	0.00
28	4-01	4.51	4.52	0.04
29	4-02	4.43	4.53	-0.04
30	4-03	4.47	4.45	0.07
31	4-04	4.50	4.43	0.14
32	4-06	4.47	4.54	-0.06
33	4-07	4.43	4.38	0.02
34	4-09	4.49	4.53	0.00
35	4-10	4.53	4.53	-0.07
36	4-11	4.48	4.44	-0.05
Mean	N/A	4.02	4.02	SAR: 1.70
SD	N/A	0.41	0.41	

SAR: Sum of Absolute Residuals.

9.2.1.3 Discussion

Since the samples selected by Mahalanobis distance are different in the one-way and four-way scans, direct comparison of the means of two sets of data are not applicable. But the two sets of the NIR results are comparable.

The F-test[89] for variance of one-way and four-way scans is 1.16 which is less than the critical value of 1.76 and indicates the insignificant difference between these two sample presentation methods. Table 18 gives the summary results of the F-test.

Table 18. F-test of One-way and Four-way Scans: Two-Sample for Variances

Statisticals	*Four-way Scan*	*One-way Scan*
Mean	4.02	3.92
Variance	0.167	0.194
Observations	36	36
Degree of Freedom	35	35
F Statistical	N/A	1.16
F Critical one-tail	N/A	1.76

9.2.2 Effect of Wavelength Range

Two different wavelength ranges were used to study how the selected wavelengths affect the calibration/validation results. To keep the other factors constant, the Mahalanobis distance selection method and 2nd derivative pretreatment with segment = 20 were used for both wavelength ranges.

9.2.2.1 Range of 900 nm – 1300 nm

Figures 44-46 are the loading plot, correlation plot, and PRESS plot of the calibration set, respectively. Table 19 is the summary of the statistical results of the calibration set. The loading plot indicates the primary PCs (each curve represents one PC or factor), the correlation plot indicates the good correlation coefficient in the range of 1100 nm – 1250 nm, the PRESS plot results the first minimum PRESS value with 7 factors, and the F value reaches the first maximum with 7 factors. Therefore, 7 factors were used to obtain the calibration equation.

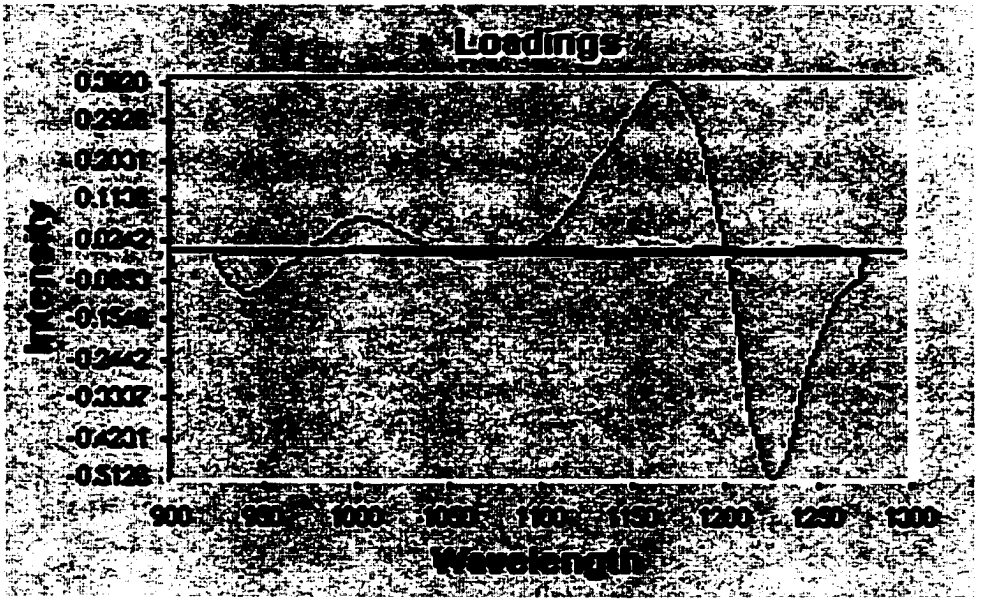

Figure 44. Loading Plot, PLSR Calibration, 900 nm – 1300 nm

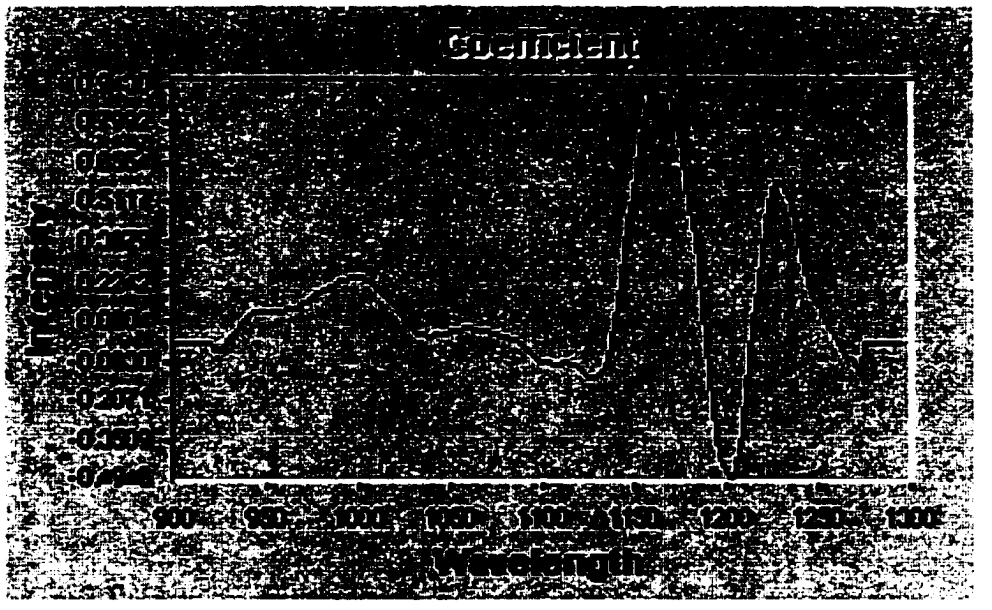

Figure 45. Correlation Coefficient Plot, PLSR Calibration, 900 nm – 1300 nm

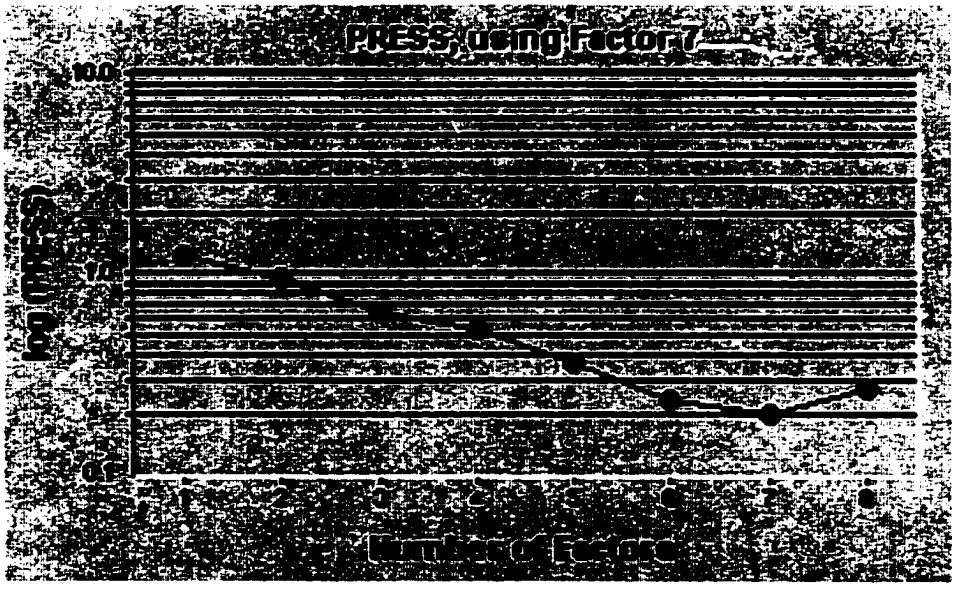

Figure 46. PRESS Plot, PLSR Calibration, 900 nm – 1300 nm

Table 19. Statistical Results, PLSR Calibration, 900 nm – 1300 nm

Factor	R^2	SEC	PRESS	F-value
1	0.8351	0.1831	1.2389	172.1
2	0.8849	0.1552	0.9533	126.9
3	0.9481	0.1059	0.6551	194.8
4	0.9600	0.0945	0.5353	185.8
5	0.9684	0.0853	0.3572	184.0
6	0.9771	0.0739	0.2342	206.2
7*	0.9822	0.0662	0.1994	220.9
8	0.9841	0.0638	0.2712	208.7
9	0.9857	0.0617	0.3194	198.8
10	0.9883	0.0568	0.2752	211.8
11	0.9903	0.0527	0.3330	223.9
12	0.9908	0.0525	0.4318	207.3
13	0.9913	0.0523	0.4525	192.8
14	0.9923	0.0504	0.5619	193.1
15	0.9928	0.0498	0.5636	184.2

**Number of factors used for the calibration set*

In Table 20 are the summary results of the NIR vs. HPLC for the calibration set.

Table 20. NIR vs. HPLC Data, PLSR Calibration, 900 nm – 1300 nm

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-01	3.45	3.47	-0.02
2	1-02	3.38	3.44	-0.06
3	1-03	3.46	3.47	-0.01
4	1-04	3.27	3.22	0.05
5	1-05	3.48	3.38	0.10
6	1-06	3.37	3.41	-0.04
7	1-07	3.37	3.35	0.02
8	1-08	3.32	3.25	0.07
9	1-09	3.16	3.23	-0.07
10	1-10	3.38	3.24	0.14
11	1-11	3.31	3.41	-0.10
12	1-12	3.37	3.43	-0.06
13	2-01	3.94	4.05	-0.11
14	2-02	3.89	3.94	-0.05
15	2-05	4.00	3.95	0.05
16	2-06	4.04	3.97	0.07
17	2-07	3.96	3.96	0.00
18	2-09	3.96	3.92	0.04
19	2-11	3.97	3.97	0.00
20	3-03	4.17	4.20	-0.03
21	3-04	4.24	4.32	-0.08
22	3-05	4.07	4.02	0.05
23	3-06	4.11	4.10	0.01
24	3-07	4.10	4.10	0.00
25	3-08	4.11	4.04	0.07
26	3-09	4.24	4.23	0.01
27	3-10	4.05	4.05	0.00
28	3-11	4.18	4.14	0.04
29	3-12	4.29	4.31	-0.02
30	4-01	4.53	4.52	0.01
31	4-02	4.45	4.53	-0.08
32	4-03	4.47	4.45	0.02
33	4-04	4.52	4.43	0.09
34	4-05	4.47	4.50	-0.03
35	4-09	4.53	4.53	0.00
36	4-10	4.47	4.53	-0.06
Mean	N/A	3.92	3.92	SAR: 1.66
SD	N/A	0.44	0.44	

SAR: Sum of Absolute Residuals.

Figures 47-49 are the plots of NIR vs. HPLC data, the NIR residuals, and the Upper Confidence Limit (UCL)/Low confidence Limit (LCL) (ref. 89) of average differences with 95% confidence, respectively, for the calibration set. When not labeled, the unit in all plots is mg/tablet. The x axis of figures 35 and 41 is the sample number. The calculated data are the NIR data and the lab data are the HPLC data in all relative plots.

Figure 47. NIR vs. HPLC Plot, PLSR Calibration, 900 nm – 1300 nm

Figure 48. NIR Residual Plot, PLSR Calibration, 900 nm – 1300 nm

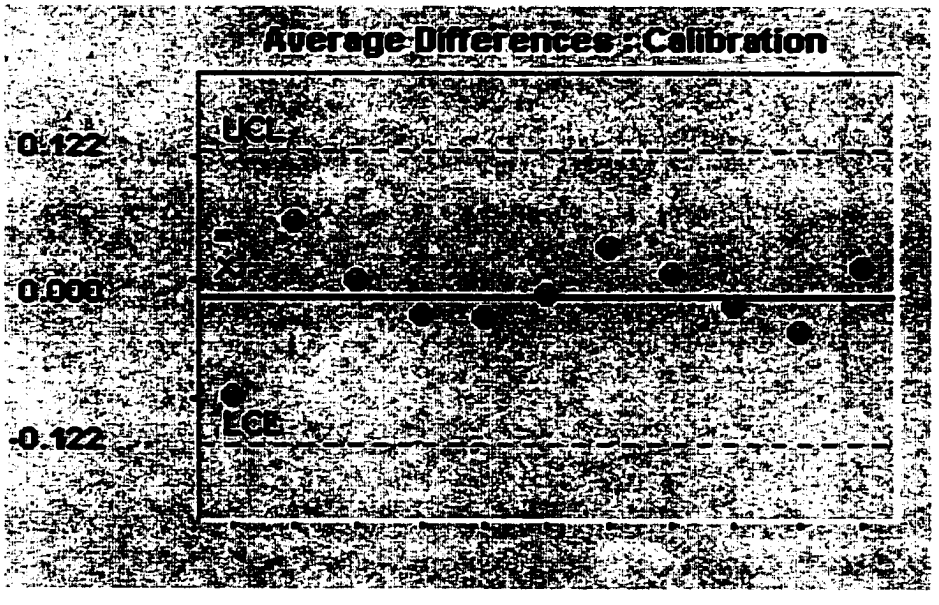

Figure 49. UCL and LCL, PLSR Calibration, 900 nm – 1300 nm

9.2.2.2 Range 850 nm – 1650 nm

Figures 50 – 52 are the loading plot, correlation coefficient plot, and PRESS plot of the calibration set, respectively, for PLSR in the range 850 nm – 1650 nm. Table 21 is the summary of the statistical results for the calibration set. The loading plot indicates the primary PCs and the correlation plot indicates good correlation coefficients in the range 1100 – 1550 nm. The PRESS plot shows the first minimum PRESS value is reached with 4 factors, and the F value reaches the first maximum also with 4 factors. Therefore, 4 factors were used to obtain the calibration equation.

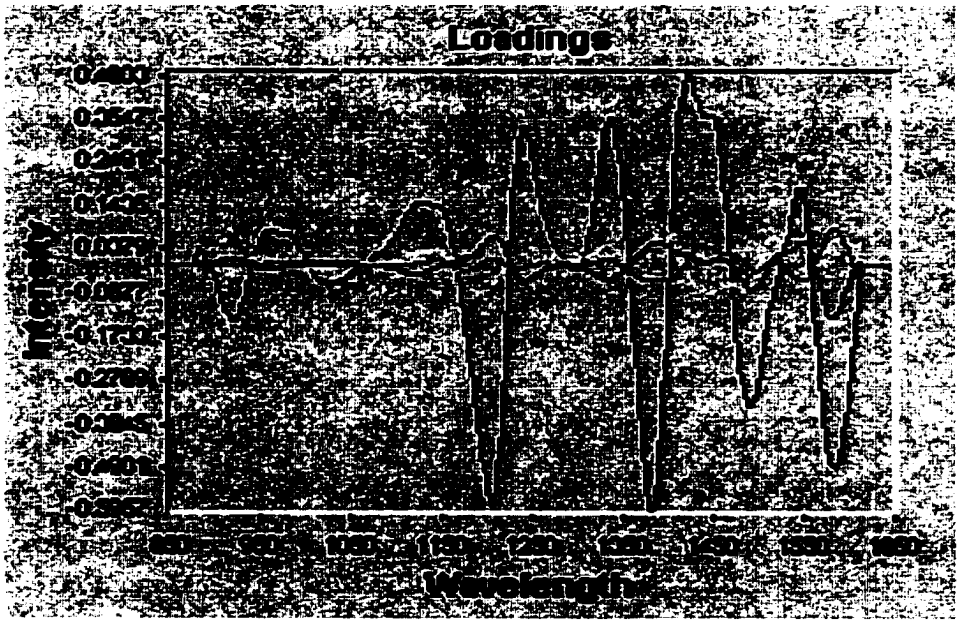

Figure 50. Loading Plot, PLSR Calibration, 850 nm – 1650 nm

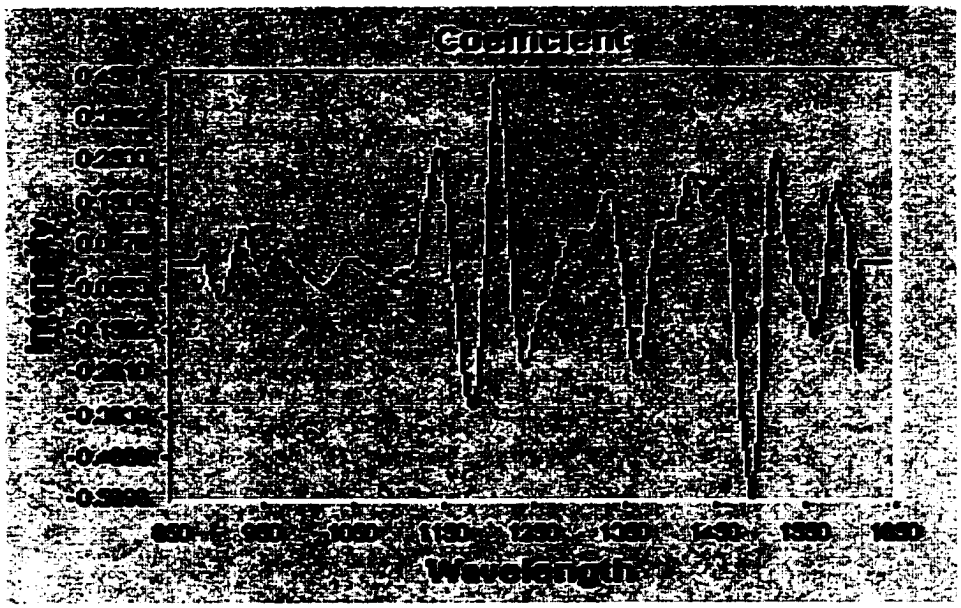

Figure 51. Correlation Plot, PLSR Calibration, 850 nm – 1650 nm

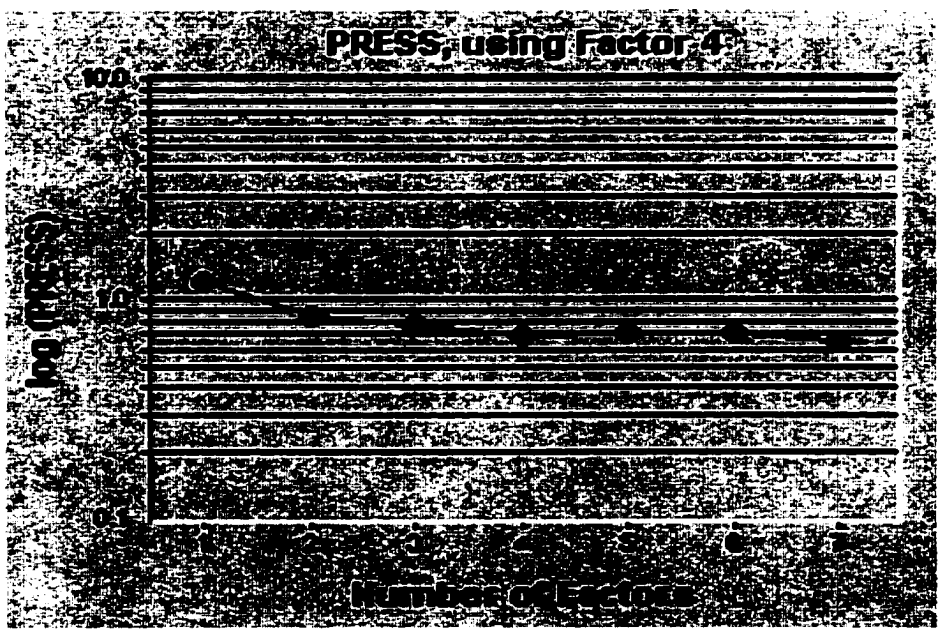

Figure 52. PRESS Plot, PLSR Calibration, 850 nm – 1650 nm

Table 21. Statistical Results, PLSR Calibration, 850 nm – 1650 nm

Factor	R^2	SEC	PRESS	F-value
1	0.8355	0.1828	1.2217	172.7
2	0.8936	0.1492	0.8609	138.6
3	0.9188	0.1324	0.7842	120.7
4*	0.9477	0.1079	0.6909	140.5
5	0.9553	0.1015	0.7012	128.1
6	0.9622	0.0949	0.7064	123.1
7	0.9731	0.0814	0.6394	144.9
8	0.9767	0.0772	0.6327	141.6
9	0.9809	0.0712	0.5477	148.4
10	0.9828	0.0690	0.4633	142.7
11	0.9875	0.0601	0.4988	171.7
12	0.9886	0.0586	0.4720	165.8
13	0.9915	0.0517	0.5643	197.3
14	0.9933	0.0470	0.5871	222.0
15	0.9948	0.0424	0.6025	255.0

**The number of factors used in the calibration set.*

Table 22 gives the results of the NIR vs. HPLC for the calibration set in the range 850 nm – 1650 nm.

Table 22. NIR vs. HPLC Data, PLSR Calibration, 850 nm – 1650 nm

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-01	3.50	3.47	0.03
2	1-02	3.36	3.44	-0.08
3	1-03	3.38	3.47	-0.09
4	1-04	3.42	3.22	0.20
5	1-05	3.42	3.38	0.04
6	1-06	3.27	3.41	-0.14
7	1-07	3.29	3.35	-0.06
8	1-08	3.27	3.25	0.02
9	1-09	3.34	3.23	0.11
10	1-10	3.30	3.24	0.06
11	1-11	3.39	3.41	-0.02
12	1-12	3.31	3.43	-0.12
13	2-01	3.86	4.05	-0.19
14	2-02	4.00	3.94	0.06
15	2-05	4.11	3.95	0.16
16	2-06	3.97	3.97	0.00
17	2-07	4.00	3.96	0.04
18	2-09	3.93	3.92	0.01
19	2-11	4.09	3.97	0.12
20	3-03	4.24	4.20	0.04
21	3-04	4.13	4.32	-0.19
22	3-05	4.08	4.02	0.06
23	3-06	4.11	4.10	0.01
24	3-07	4.15	4.10	0.05
25	3-08	4.18	4.04	0.14
26	3-09	4.40	4.23	0.17
27	3-10	4.10	4.05	0.05
28	3-11	4.18	4.14	0.04
29	3-12	4.25	4.31	-0.06
30	4-01	4.48	4.52	-0.04
31	4-02	4.35	4.53	-0.18
32	4-03	4.47	4.45	0.02
33	4-04	4.45	4.43	0.02
34	4-05	4.35	4.50	-0.15
35	4-09	4.47	4.53	-0.06
36	4-10	4.43	4.53	-0.10
Mean	N/A	3.92	3.92	SAR: 2.93
SD	N/A	0.43	0.44	

SAR: Sum of Absolute Residuals.

Figures 53-55 are the plots of NIR vs. HPLC data, NIR residuals, and UCL/LCL of average difference with 95% confidence, respectively, for the calibration in range of 850 nm – 1650 nm.

Figure 53. NIR vs. HPLC Plot, PLSR Calibration, 850 nm – 1650 nm

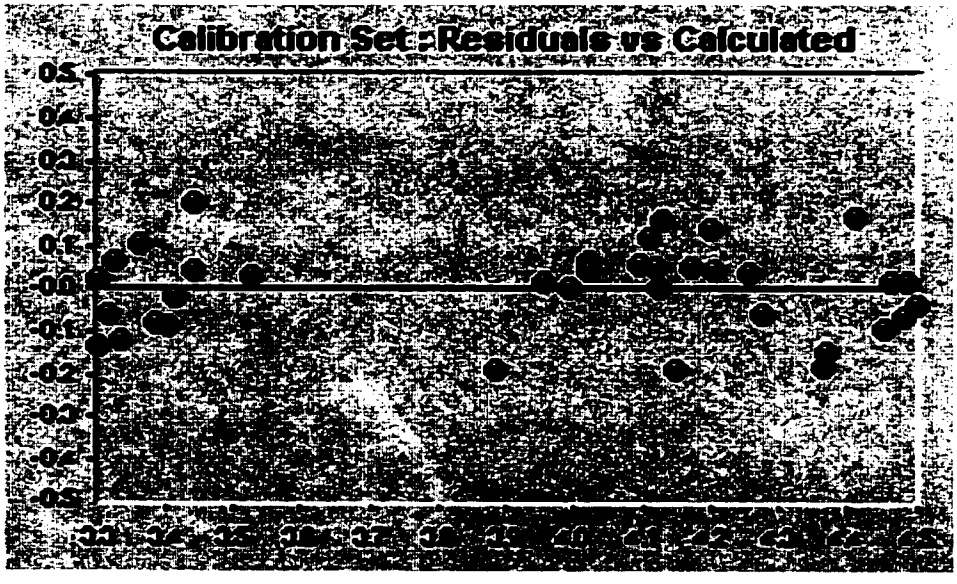

Figure 54. NIR Residual Plot, PLSR Calibration, 850 nm – 1650 nm

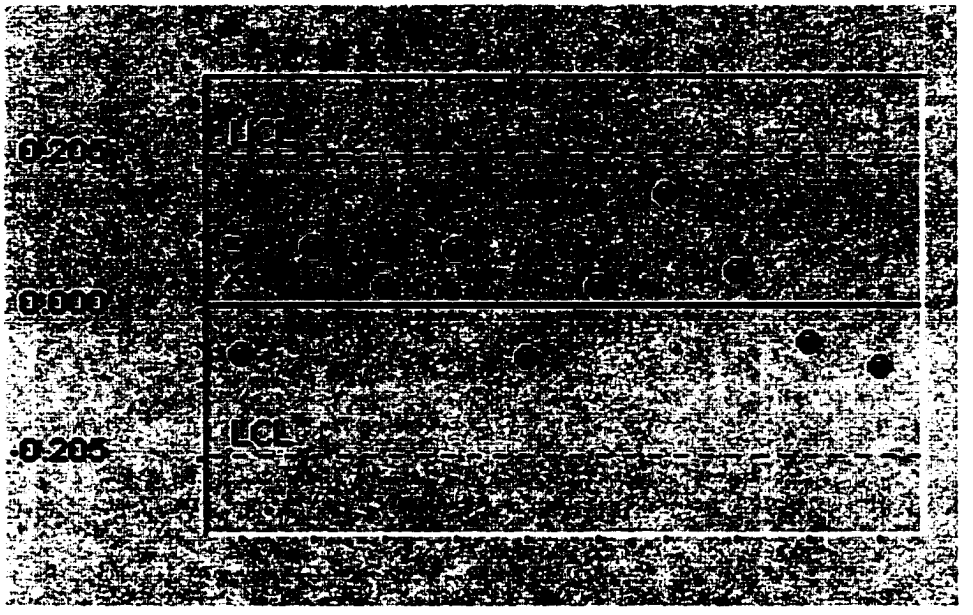

Figure 55. UCL/LCL Plot, PLSR Calibration, 850 nm – 1650 nm

9.2.2.3 Discussion

Because the calibration plot of PLSR is multi dimensional which can not to be presented graphically, the plot of NIR vs. HPLC data was used directly to present the calibration plot.

The R^2 is 0.9822 for the range 900 nm- 1300 nm and 0.9477 for the range 850 nm – 1650 nm. The residual ranges are –0.11 to 0.14 in the range 900 nm – 1300 nm and –0.19 to 0.20 in the range 850 nm – 1650 nm. The PRESS is 0.20 for the range 900 nm – 1300 nm and 0.69 for the range 850 nm – 1650 nm. All these indicate that the wavelength range 900 nm – 1300 nm provides a bit better calibration than the wavelength range 850 nm – 1650 nm, which is not surprising since the water content

of the gum sample interferes with nicotine quantitation at about 1400 nm. It was determined that the water content of the nicotine gum samples ranged from 0.2% to 0.5% which corresponds to 2 – 5 mg/tablet.

In addition, an F-test was performed on these two ranges and it was found to be 1.03 which is below the critical value of 1.76. This indicates the difference between these two wavelength ranges is insignificant.

9.2.3 Different Mathematical Pretreatments

Many things affect a NIR spectra in a real situation, such as instrument noise, baseline drift, stray light, nonlinearity between optical data and analyte concentration (deviation from Beer's law), scattering, nonlinear dispersion, physical property variations in samples, band overlap, band broadening, and temperature fluctuation. Therefore, mathematical pretreatment of the sample spectra is necessary to eliminate or reduce the effects for better calibration, validation, and prediction. To keep other factors constant, the Mahalanobis distance selection method and wavelength range 900 nm – 1300 nm were used for PLSR in this section.

9.2.3.1 Raw Data

The raw sample spectral data were used here without any type of mathematical pretreatment. Table 23 gives the statistical summary and Figure 56 gives the plot of NIR vs. HPLC data for the calibration set.

Table 23. Statistical Results, PLSR Calibration, Raw Data

Factor	R^2	SEC	PRESS	F-value
1	0.7472	0.2267	1.9533	100.5
2	0.8467	0.1792	1.2616	91.1
3	0.8564	0.1761	1.1318	63.6
4	0.9442	0.1115	0.5949	131.1
5*	0.9725	0.0796	0.3244	212.2
6	0.9744	0.0782	0.3184	183.6
7	0.9826	0.0656	0.2470	225.5
8	0.9829	0.0662	0.2612	193.5
9	0.9836	0.0661	0.2848	173.0
10	0.9856	0.0631	0.2945	170.9
11	0.9889	0.0566	0.2984	193.7
12	0.9897	0.0555	0.3463	184.7
13	0.9917	0.0511	0.5201	201.6
14	0.9923	0.0502	0.7628	194.4
15	0.9936	0.0469	0.8326	207.9

*the number of factors used for calibration.

Figure 56. NIR vs. HPLC Plot, PLSR Calibration, Raw Data

9.2.3.2 First Derivative

First derivative is the rate of change (slope) $dA/d\lambda$ which computes the difference between absorbances at adjacent wavelengths.

$$\frac{d}{d\lambda}[f(\lambda)+C]=f'(\lambda) \qquad \textit{Equation 26}$$

The different offset value C associated with each spectrum, $f(\lambda)$, can be eliminated in the first derivative.

One way to calculate the first derivative is as the first order finite-difference derivative. This method requires two values to be specified: the length of the segment (segment size), and the length of the gap between two segments (gap size). A segment size=20 nm and gap size=0 nm were used in this section according to the instrument manufacturer's recommendation. The effect of segment size is discussed in Section 9.2.5.

Table 24 summarizes the statistical results of the 1st derivative mathematical pretreatment with segment size=20 and gap size=0. Figure 57 is the NIR vs. HPLC plot for the calibration set.

Table 24. Statistical Results, PLSR Calibration, 1st Derivative

Factor	R^2	SEC	PRESS	F-value
1	0.8351	0.1831	1.2136	172.1
2	0.8849	0.1552	0.9250	126.9
3	0.9481	0.1059	0.5410	194.8
4	0.9600	0.0945	0.5169	185.8
5	0.9684	0.0853	0.3064	184.0
6	0.9771	0.0739	0.3051	206.2
7*	0.9822	0.0662	0.2424	220.9
8	0.9841	0.0638	0.2709	208.7
9	0.9857	0.0617	0.3320	198.8
10	0.9883	0.0568	0.3093	211.8
11	0.9903	0.0527	0.3273	223.9
12	0.9908	0.0525	0.3370	207.3
13	0.9913	0.0523	0.3309	192.1
14	0.9923	0.0504	0.3641	193.1
15	0.9928	0.0498	0.4709	184.5

*the number of factor used for the calibration.

Figure 57. NIR vs. HPLC Plot, PLSR Calibration, 1st Derivative

9.2.3.3 *Second Derivative*

The Second derivative $d^2A/d\lambda^2$ computes the differences in the spectrum obtained from first derivative and can sharpen bands, increase resolution, and reduce linear background. The second derivative is the mathematical pretreatment used most often.

$$\frac{d^2}{d\lambda^2}[f(\lambda) + (a \cdot \lambda + C)] = f''(\lambda) \qquad \textit{Equation 27}$$

Table 25 summarizes the statistical results of 2nd derivative mathematical pretreatment with segment size=20 and gap size=0. Figure 58 is the NIR vs. HPLC plot for the calibration set.

Table 25. Statistical Results, PLSR Calibration, 2nd Derivative

Factor	R^2	SEC	PRESS	F-value
1	0.8351	0.1831	1.2389	172.1
2	0.8849	0.1552	0.9533	126.9
3	0.9481	0.1059	0.6551	194.8
4	0.9600	0.0945	0.5353	185.8
5	0.9684	0.0853	0.3572	184.0
6	0.9771	0.0739	0.2342	206.2
7*	0.9822	0.0662	0.1994	220.9
8	0.9841	0.0638	0.2712	208.7
9	0.9857	0.0617	0.3194	198.8
10	0.9883	0.0568	0.2752	211.8
11	0.9903	0.0527	0.3330	223.9
12	0.9908	0.0525	0.4318	207.3
13	0.9913	0.0523	0.4525	192.8
14	0.9923	0.0504	0.5619	193.1
15	0.9928	0.0498	0.5636	184.2

**the number of factors used in calibration.*

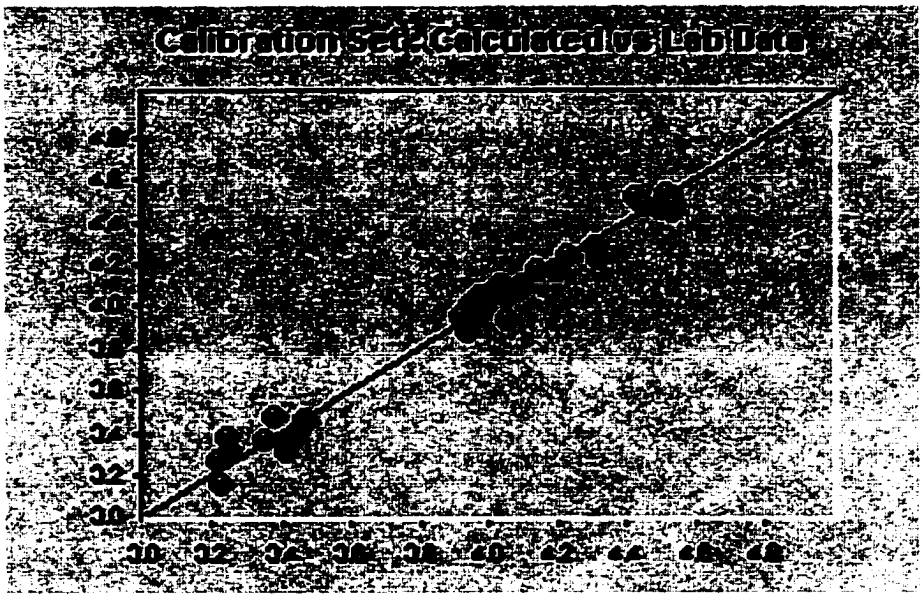

Figure 58. NIR vs. HPLC Plot, PLSR Calibration, 2nd Derivative

9.2.3.4 Standard Normal Variate

Standard Normal Variate (SNV) is a scatter correction method used commonly to normalize spectra when the effective pathlength varies among samples in a data set. Such pathlength variation can occur when measuring the spectra of granular or powdery samples because *(a)* sample presentation in a cell is not perfectly reproducible; *(b)* particle size varies between samples. Mathematically, the spectrum is mean centered and then divided by its standard deviation:

$$S_i^{SNV} = \frac{S_i - \bar{S}}{\sqrt{\frac{\sum_{i=1}^{n}(S_i - \bar{S})^2}{n-1}}}$$

Equation 28

Table 26 summarizes the statistical results of SNV mathematical pretreatment and Figure 59 is the NIR vs. HPLC plot for the calibration set.

Table 26. Statistical Results, PLSR Calibration, SNV

Factor	R^2	SEC	PRESS	F-value
1	0.7863	0.2084	1.6483	125.1
2	0.8304	0.1884	1.7306	80.8
3*	0.9658	0.0860	0.3706	301.0
4	0.9715	0.0797	0.3274	264.3
5	0.9736	0.0780	0.3093	221.4
6	0.9801	0.0689	0.3020	237.9
7	0.9810	0.0685	0.2792	206.5
8	0.9825	0.0670	0.3268	189.2
9	0.9879	0.0567	0.4200	236.0
10	0.9905	0.0512	0.3448	260.9
11	0.9929	0.0452	0.2985	305.2
12	0.9959	0.0352	0.2880	461.8
13	0.9968	0.0315	0.2946	532.7
14	0.9981	0.0250	0.2838	790.3
15	0.9986	0.0219	0.2929	956.3

**the number of factors used in calibration.*

Figure 59. NIR vs. HPLC Plot, PLSR Calibration, SNV

9.2.3.5 Detrend

Detrend is a method that can be used to remove baseline offset, slope, or curvature from a spectrum. This is accomplished by calculating a baseline function as the least squares fit of a polynomial to the sample spectrum, and then subtracting that function from the spectrum.

Table 27 gives the statistical summary of calibration with Detrend (second order function) mathematical pretreatment and Figure 60 is the NIR vs. HPLC plot.

Table 27. Statistical Results, PLSR Calibration, Detrend

Factor	R^2	SEC	PRESS	F-value
1	0.8339	0.1837	1.2258	170.7
2	0.8719	0.1637	1.0398	112.3
3	0.9509	0.1030	0.6265	206.5
4*	0.9775	0.0708	0.5104	337.1
5	0.9784	0.0706	0.2788	271.2
6	0.9816	0.0661	0.2505	258.4
7	0.9850	0.0608	0.2016	262.7
8	0.9878	0.0559	0.2703	272.8
9	0.9896	0.0527	0.2293	273.7
10	0.9924	0.0458	0.2220	326.4
11	0.9944	0.0400	0.2018	390.0
12	0.9956	0.0362	0.2318	438.3
13	0.9970	0.0308	0.2215	557.9
14	0.9980	0.0258	0.2275	740.2
15	0.9991	0.0178	0.2537	1444.9

**the number of factors used in calibration.*

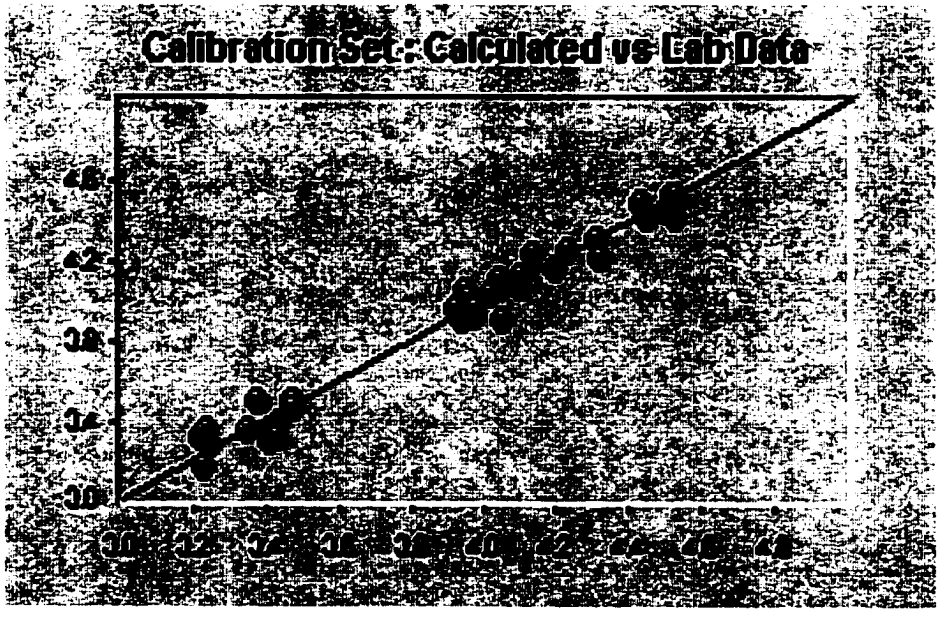

Figure 60. NIR vs. HPLC Plot, PLSR Calibration, Detrend

9.2.3.6 Savitzky-Golay

This method of smoothing and derivative calculation relies on the least squares fit of a polynomial to a spectral segment. Though both the Savitzky-Golay (S-G) and Detrend methods are based on the least squares fit of a polynomial function, they differ in both scope and effect. In the S-G method, a smoothed spectrum or a derivative spectrum of any order can be calculated using coefficients of the polynomials fitted to a spectrum.

Table 28 summarizes the statistical results of calibration with S-G pretreatment (with 7-point smoothing and 2nd derivative) and Figure 61 is the NIR vs. HPLC plot of the calibration.

Table 28. Statistical Results, PLSR Calibration, S-G

Factor	R^2	SEC	PRESS	F-value
1	0.8029	0.2002	1.4788	138.5
2	0.9117	0.1359	0.7451	170.4
3	0.9383	0.1154	0.7799	162.2
4	0.9460	0.1097	0.5887	135.7
5*	0.9657	0.0889	0.4984	169.0
6	0.9694	0.0854	0.4517	153.1
7	0.9847	0.0615	0.4205	257.0
8	0.9942	0.0384	0.3683	581.4
9	0.9965	0.0304	0.2678	826.7
10	0.9976	0.0257	0.2246	1043.3
11	0.9985	0.0206	0.2108	1483.0
12	0.9990	0.0169	0.2031	2004.8
13	0.9996	0.0107	0.2058	4672.7
14	0.9998	0.0080	0.2033	7734.7
15	0.9999	0.0045	0.2064	22594.4

**the number of factors used in calibration.*

Figure 61. NIR vs. HPLC Plot, PLSR Calibration, S-G

9.2.3.7 *Multiplicative Scatter Correction*

The Multiplicative scatter correction (MSC) was developed to reduce the effect of scattered light on diffuse reflectance and transmittance NIR spectra[90]. It is another widely used mathematical pretreatment method, and is a means of removing varying background spectra with nonscattering origins. Consequently, MSC sometimes appears as multiplicative signal correction.

Mathematical pretreatment such as derivatives, SNV, Detrend, and Savitsky-Golay are applied to individual spectra without any preconception of what the resulting spectrum should look like. On the other hand, the MSC method calculates a mean spectrum under the assumption that spectra in the data set are distributed normally. Thus, the mean spectrum is the most probable spectrum.

In MSC, the mean spectrum is calculated from all spectra in a defined data set. Then a least squares linear regression is performed on absorbance values of the sample spectrum versus those at corresponding wavelengths in the mean spectrum. This operation yields a linear equation with a defined intercept and slope. Next, the value of the intercept is subtracted from every data point in the spectrum. Finally, each absorbance value in the resulting spectrum is divided by the value for slope. Using the mean spectrum, the same set of operation is performed on every spectrum in the data set.

Table 29 gives the summary statistical results with MSC pretreatment and Figure 62 is the NIR vs. HPLC plot of the calibration.

Table 29. Statistical Results, PLSR Calibration, MSC

Factor	R^2	SEC	PRESS	F-value
1	0.7962	0.2084	1.5199	125.0
2	0.8325	0.1873	1.3768	82.0
3	0.9657	0.0861	0.3177	299.9
4*	0.9713	0.0800	0.2824	262.2
5	0.9734	0.0783	0.3045	219.2
6	0.9800	0.0691	0.2776	236.5
7	0.9815	0.0675	0.2701	212.4
8	0.9826	0.0667	0.2630	190.6
9	0.9896	0.0526	0.3089	274.3
10	0.9925	0.0455	0.2735	330.8
11	0.9963	0.0327	0.2600	585.0
12	0.9968	0.0312	0.2925	589.4
13	0.9982	0.0237	0.3041	941.0
14	0.9985	0.0224	0.2932	983.5
15	0.9994	0.0145	0.2906	2199.8

**the number of factors used in calibration.*

Figure 62. NIR vs. HPLC Plot, PLSR Calibration, MSC

9.2.3.8 Discussion

Overall, the differences are insignificant with different mathematical pretreatments for the calibration set based on the PRESS value, R^2, and SEC. However, good calibration results do not necessarily mean good results for future prediction. The predictability of the calibrations with different pretreatments are discussed in Section 9.2.7. When necessary, different pretreatments can be used together, but caution should be taken since the more mathematical pretreatments used, the more danger of losing some useful information. In general, the 2^{nd} derivative and MSC are the two pretreatments used most often for quantitation. However, different methods should be evaluated for different applications to obtain the most desirable results.

9.2.4 Selection of Factor Numbers

9.2.4.1 Calibration with Two, Seven, and Twelve Factors

Table 30 gives the statistical results of calibration with factors of 2, 7, and 12. Figures 63-65 are the NIR vs. HPLC plot of calibration with factors of 2, 7, and 12, respectively.

Table 30. Statistical Results, PLSR Calibration, Factor = 2, 7, and 12

Factor	R^2	SEC	PRESS	F-value
1	0.8351	0.1831	1.2389	172.1
2*	0.8849	0.1552	0.9533	126.9
3	0.9481	0.1059	0.6551	194.8
4	0.9600	0.0945	0.5353	185.8
5	0.9684	0.0853	0.3572	184.0
6	0.9771	0.0739	0.2342	206.2
7*	0.9822	0.0662	0.1994	220.9
8	0.9841	0.0638	0.2712	208.7
9	0.9857	0.0617	0.3194	198.8
10	0.9883	0.0568	0.2752	211.8
11	0.9903	0.0527	0.3330	223.9
12*	0.9908	0.0525	0.4318	207.3
13	0.9913	0.0523	0.4525	192.8
14	0.9923	0.0504	0.5619	193.1
15	0.9928	0.0498	0.5636	184.2

**the number of factors used in calibration.*

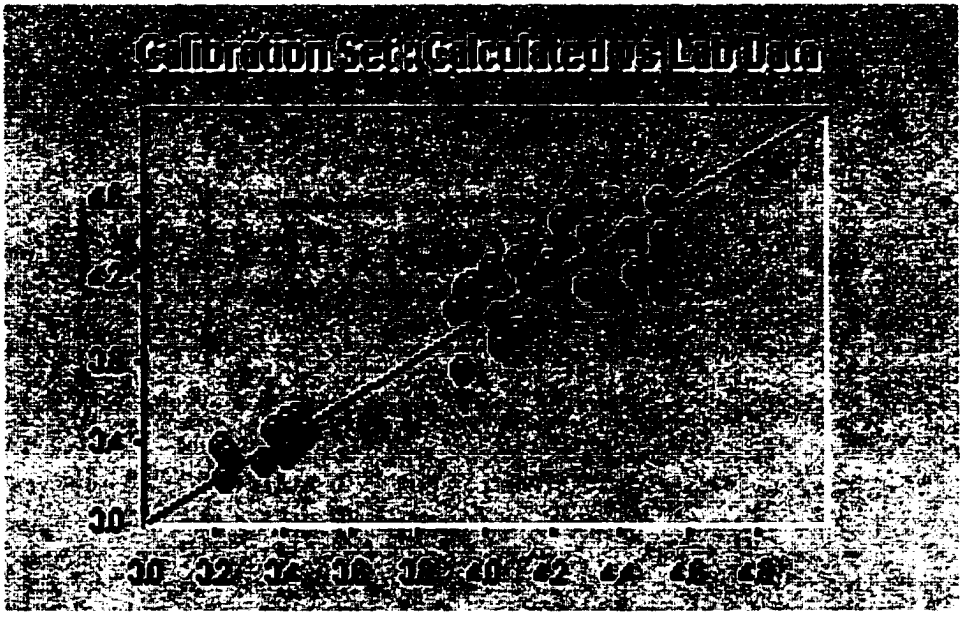

Figure 63. NIR vs. HPLC Plot, PLSR Calibration, Factor=2

Figure 64. NIR vs. HPLC Plot, PLSR Calibration, Factor=7

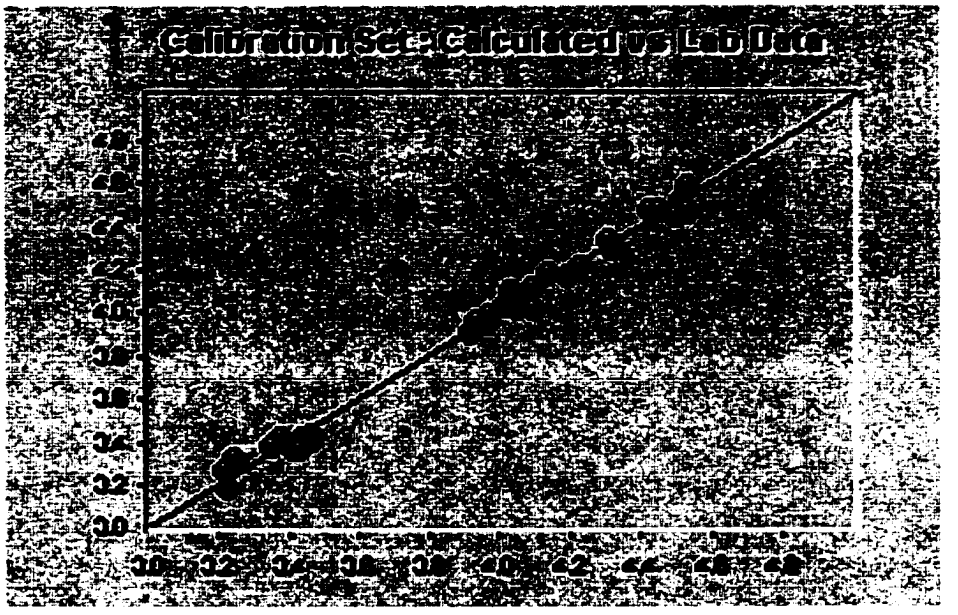

Figure 65. NIR vs. HPLC Plot, PLSR Calibration, Factor=12

9.2.4.2 Discussion

It is obvious that the calibration with two factors produced a low correlation and a high PRESS value. The calibration with twelve factors seems superficially to have the best calibration fit indicated by the high correlation and low SEC, but it results a higher PRESS and SEP values than for seven factors. These indicate the poor robustness and higher prediction error of the calibration equation with two or twelve factors. Again, good calibration data do not necessarily mean good results for future prediction. The predicted results which are discussed in section 9.2.7 below clearly demonstrate that.

9.2.5 Selection of Segment Size

Segment size affects the sharpness of the derivatized spectrum and the outcome of the calibration to a certain extend. When the segment size is too large, some critical information may be averaged; when segment size is too small, some of the non-characteristic spectra may be treated as unique information in the calibration, leading to poor calibration and poor prediction.

9.2.5.1 Segment Size = 10

Figure 66 is the loading plot ; in Table 31 are given the summary statistical results of PLSR calibration; and in Figure 67 is the NIR vs. HPLC plot of the calibration, all with s= 10.

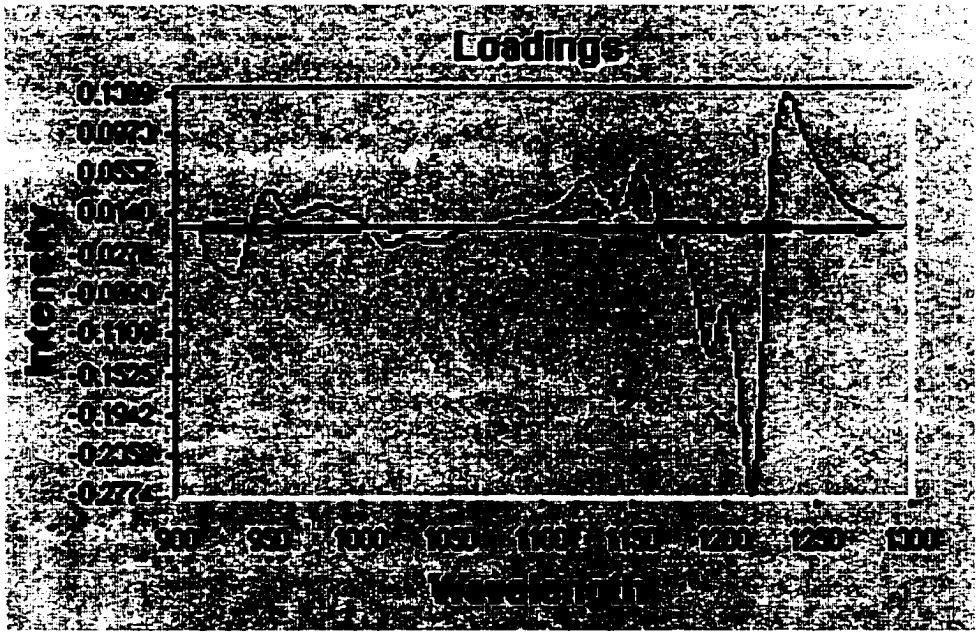

Figure 66. Loading Plot, PLSR Calibration, s=10

Table 31. Statistical Results, PLSR Calibration, s=10

Factor	R^2	SEC	PRESS	F-value
1	0.8120	0.1955	1.4039	146.8
2	0.9070	0.1395	0.7870	160.9
3	0.9429	0.1111	0.9756	176.0
4	0.9642	0.0893	0.7612	208.9
5	0.9705	0.0825	0.4503	197.1
6	0.9716	0.0822	0.3004	165.5
7*	0.9808	0.0688	0.2281	204.7
8	0.9861	0.0595	0.2199	240.3
9	0.9903	0.0508	0.2312	294.7
10	0.9931	0.0438	0.2341	357.8
11	0.9945	0.0397	0.2431	395.4
12	0.9953	0.0377	0.2516	404.2
13	0.9967	0.0320	0.3004	516.1
14	0.9973	0.0298	0.3273	554.1
15	0.9979	0.0272	0.3504	621.1

**the number of factors used in calibration.*

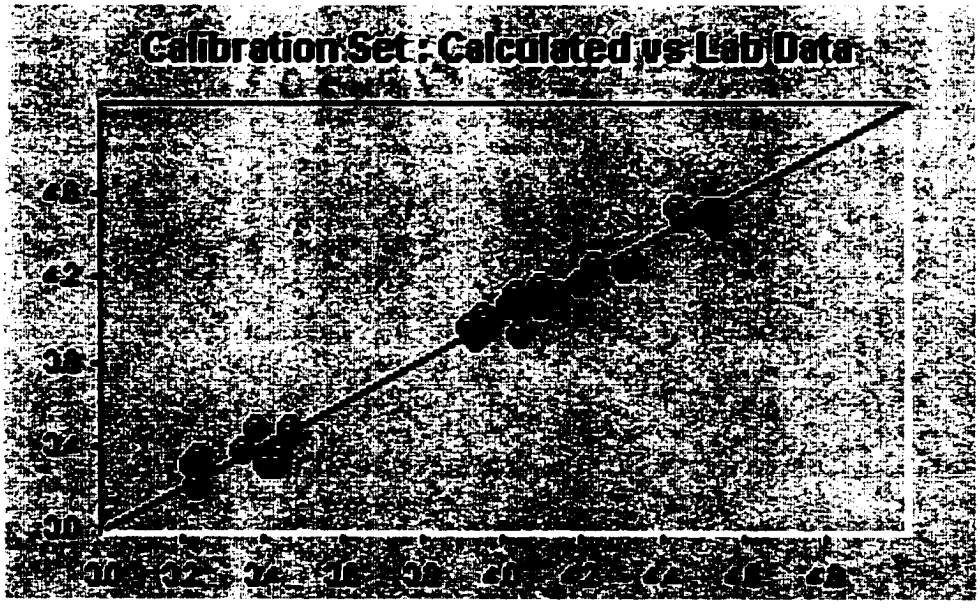

Figure 67. NIR vs. HPLC Plot, PLSR Calibration, s=10

9.2.5.2 Segment Size = 20

Figure 68 is the loading plot; in Table 32 are the summary statistical results of PLSR calibration; and in Figure 69 is the NIR vs. HPLC plot of calibration, all with s=20.

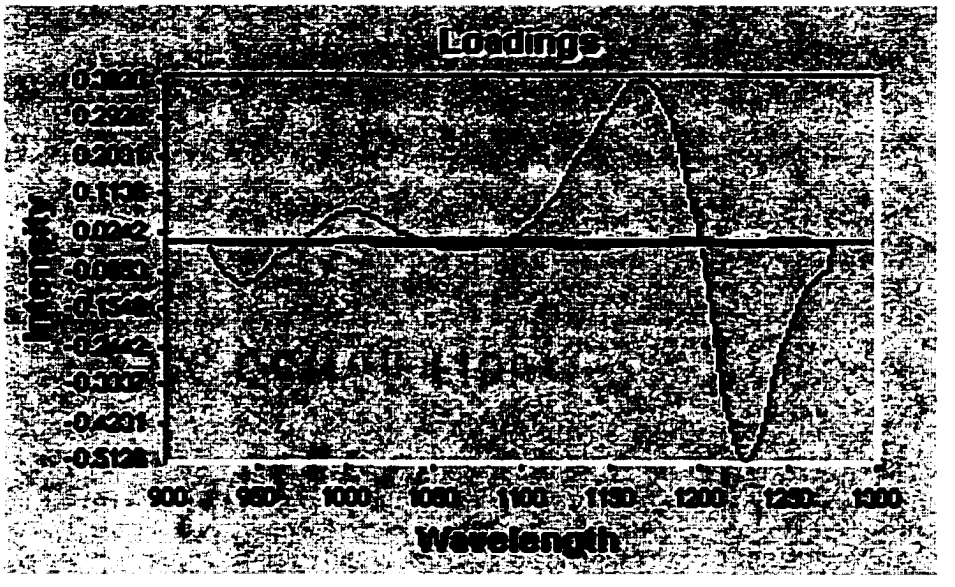

Figure 68. Loading Plot, PLSR Calibration, s = 20

Table 32. Statistical Results, PLSR Calibration, s=20

Factor	R^2	SEC	PRESS	F-value
1	0.8351	0.1831	1.2389	172.1
2	0.8849	0.1552	0.9533	126.9
3	0.9481	0.1059	0.6551	194.8
4	0.9600	0.0945	0.5353	185.8
5	0.9684	0.0853	0.3572	184.0
6	0.9771	0.0739	0.2342	206.2
7*	0.9822	0.0662	0.1994	220.9
8	0.9841	0.0638	0.2712	208.7
9	0.9857	0.0617	0.3194	198.8
10	0.9883	0.0568	0.2752	211.8
11	0.9903	0.0527	0.3330	223.9
12	0.9908	0.0525	0.4318	207.3
13	0.9913	0.0523	0.4525	192.8
14	0.9923	0.0504	0.5619	193.1
15	0.9928	0.0498	0.5636	184.2

**the number of factors used in calibration.*

Figure 69. NIR vs. HPLC Plot, PLSR Calibration, s=20

9.2.5.3 Segment Size = 30

Figure 70 is the loading plot, Table 33 are the summary statistical results and Figure 71 is the NIR vs. HPLC plot of calibration, all with s = 30.

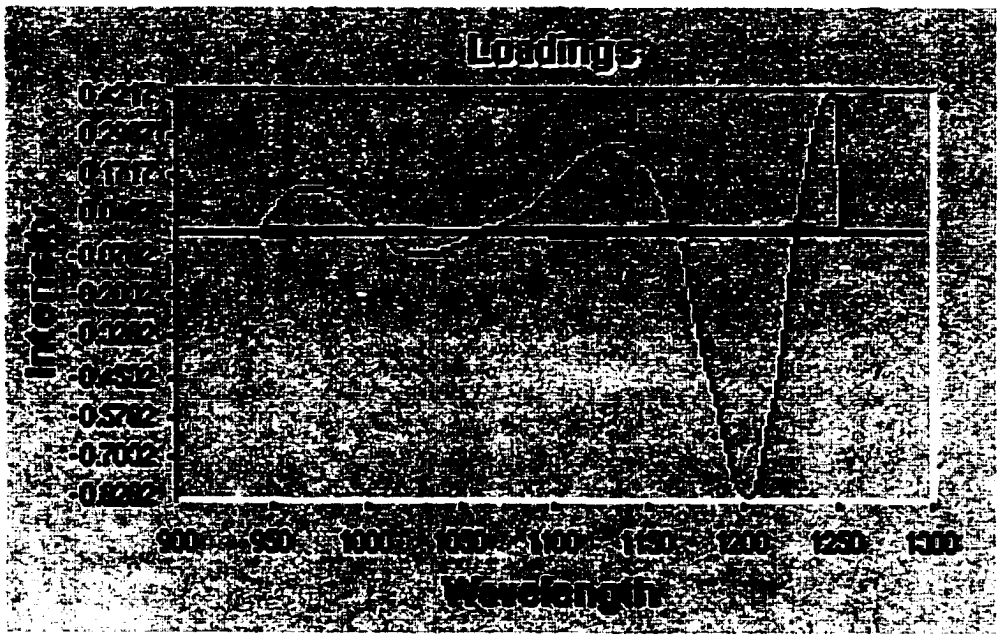

Figure 70. Loading Plot, PLSR Calibration, s=30

Table 33. Statistical Results, PLSR Calibration, s=30

Factor	R^2	SEC	PRESS	F-value
1	0.8371	0.1819	1.1945	174.7
2	0.8929	0.1498	0.8601	137.5
3	0.9380	0.1157	0.6521	161.2
4	0.9705	0.0811	0.3867	254.9
5*	0.9757	0.0747	0.2893	241.4
6	0.9791	0.0706	0.2889	226.5
7	0.9816	0.0674	0.2501	213.0
8	0.9841	0.0638	0.2770	209.1
9	0.9862	0.0605	0.2811	206.7
10	0.9869	0.0602	0.2927	188.4
11	0.9872	0.0607	0.3480	168.4
12	0.9882	0.0596	0.4037	160.2
13	0.9901	0.0557	0.4280	169.6
14	0.9923	0.0505	0.3903	192.3
15	0.9930	0.0493	0.3532	188.1

**the number of factors used in calibration.*

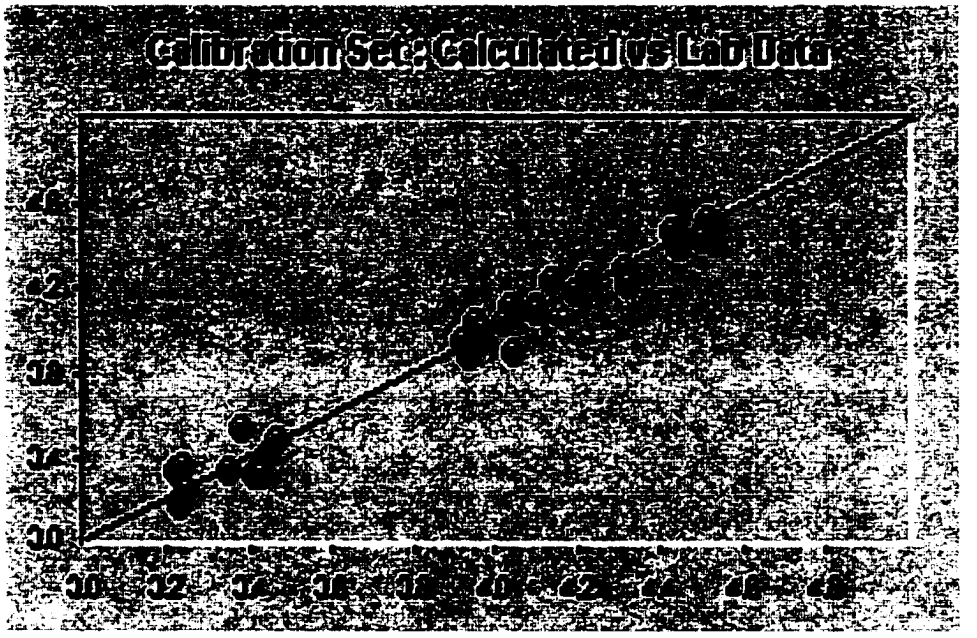

Figure 71. NIR vs. HPLC Plot, PLSR Calibration, s=30

9.2.5.4 Discussion

The derivative segment sizes of 10, 20, and 30 did not affect the calibration significantly. However, they affect the predictability of the calibration as it is discussed in section 9.2.7. In general, segment size of 20 is most often used for 2nd derivative concerning the averaging effect and information retention.

9.2.6 Regression Validation

The calibration equation of PLSR using the 2nd derivative pretreatment, segment size=20, in the range 900 nm – 1300 nm, was validated using the validation set selected by the Mahalanobis distance method. Table 34 gives the summary data of NIR vs. HPLC from the validation set and Figures 72 and 73 are the NIR vs. HPLC plot and NIR

residuals plot, respectively. The results and plots clearly demonstrate the validity of the calibration equation with comparable mean and SD to those of obtained from the calibration set and HPLC procedure. The residual plot shows the random distribution of the residuals, which indicates the low bias of the method.

Since the cross validation was performed during the calibration equation development and the samples in the validation set are from the same batches used in the calibration set, the residuals, mean, and standard deviation of the validation set were comparable to those of the calibration set. Therefore, only the validation results of PLSR with the 2nd derivative pretreatment, s=20, 7 factors, in the range 900 nm – 1300 nm, are listed here for demonstration purposes.

Table 34. NIR vs. HPLC Data, PLSR Validation, s=20, Factor=7, 2nd Derivative

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	2-03	3.93	3.85	0.08
2	2-04	4.01	4.02	-0.01
3	2-08	4.06	4.04	0.02
4	2-10	4.06	3.97	0.09
5	2-12	4.05	3.92	0.13
6	3-01	4.11	4.07	0.04
7	3-02	4.44	4.44	0.00
8	4-06	4.47	4.54	-0.07
9	4-07	4.44	4.38	0.06
10	4-18	4.39	4.51	-0.12
11	4-11	4.47	4.44	0.03
12	4-12	4.54	4.51	0.03
Mean	N/A	4.25	4.22	SAR: 0.68
SD	N/A	0.23	0.27	

Figure 72. NIR vs. HPLC Plot, PLSR Validation, 2nd Derivative

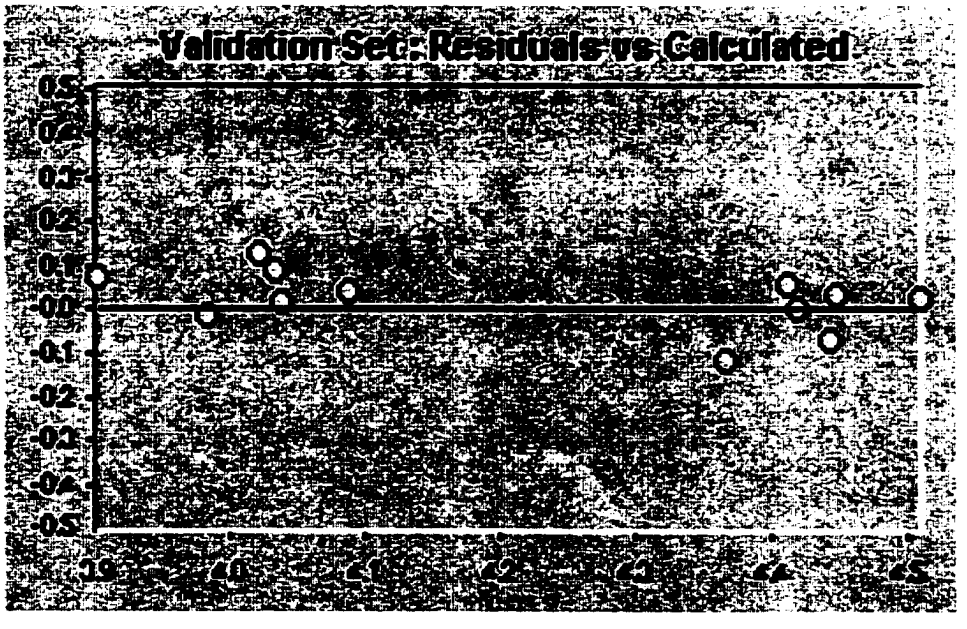

Figure 73. NIR Residual Plot, PLSR Validation, 2nd derivative

9.2.7 Prediction of Unknown Samples

To further confirm the suitability of validated calibration, 12 tablets from a different batch (batch 5) were analyzed as unknowns. After the unknown gum samples scanned on NIR, they were analyzed by HPLC method for the true potencies.

Table 35 summarizes the NIR vs. HPLC data of the prediction set using PLSR and Mahalanobis distance selection method.

Table 35. NIR Predicted Results of Unknown Samples by Different Methods

Method	Spl	LC	NIR Predicted Results (mg/tablet)											
			1	2	3	4	5	6	7	8	9	10	11	12
1	U-01	3.80	3.82	3.71	3.73	3.76	4.06	3.87	3.97	3.78	3.49	3.71	3.73	3.80
2	U-02	3.91	3.90	3.98	3.84	3.91	4.06	4.00	4.10	3.84	3.80	3.88	3.97	3.92
3	U-03	3.86	3.67	4.13	3.54	4.00	3.95	3.95	4.10	4.02	3.80	3.16	4.08	3.66
4	U-04	3.86	3.94	4.00	3.84	3.93	4.13	4.03	4.13	3.87	3.88	3.85	3.99	3.95
5	U-05	3.85	3.87	3.79	3.80	3.88	4.06	3.95	4.04	3.84	3.55	3.77	3.79	3.88
6	U-06	3.89	3.87	3.95	3.82	3.97	4.13	3.98	4.11	3.95	3.77	3.67	3.93	3.89
7	U-07	3.97	3.85	3.90	3.72	3.92	3.96	3.99	4.07	3.88	3.59	3.60	3.88	3.84
8	U-08	3.93	3.97	4.10	3.90	3.97	4.10	4.05	4.16	3.91	3.98	3.85	4.08	3.98
9	U-09	3.94	3.92	4.25	3.82	3.99	4.03	4.03	4.15	3.96	4.09	3.69	4.21	3.91
10	U-10	4.00	3.89	3.95	3.83	3.88	4.04	3.96	4.03	3.83	3.74	3.82	3.93	3.90
11	U-11	3.91	3.83	3.87	3.79	3.87	4.08	3.92	4.04	3.82	3.64	3.74	3.86	3.85
12	U-12	3.97	3.90	4.16	3.83	3.92	4.05	4.00	4.10	3.84	4.01	3.87	4.13	3.91
Mean		3.91	3.87	3.98	3.79	3.92	4.05	3.98	4.08	3.88	3.78	3.72	3.97	3.87
Min		3.80	3.67	3.71	3.54	3.76	3.95	3.87	3.97	3.78	3.49	3.16	3.73	3.66
Max		4.00	3.97	4.25	3.90	4.00	4.13	4.05	4.16	4.02	4.09	3.88	4.21	3.98
Max - Min		0.20	0.30	0.54	0.36	0.24	0.18	0.18	0.19	0.24	0.60	0.72	0.48	0.32
SD		0.058	0.076	0.157	0.092	0.066	0.056	0.051	0.055	0.069	0.189	0.197	0.142	0.083
t-Test (critical=2.07)			1.38	1.55	[redacted]	0.36	[redacted]	[redacted]	[redacted]	1.11	[redacted]	[redacted]	1.30	1.14
F-Test (critical=2.82)			1.70	[redacted]	2.48	1.27	1.08	1.31	1.12	1.40	[redacted]	[redacted]	[redacted]	2.01

Shaded area indicates significant difference. 1: 900 nm – 1300 nm, 2nd derivative, s=20, factors=7; 2: 2nd 850 nm – 1650 nm, derivative, s=20, factors=7; 3: 900 nm – 1300 nm, Raw Data, factors=7; 4: 900 nm – 1300 nm, 1st derivative, s=20, factors=7; 5:900 nm – 1300 nm, SNV, factors=7; 6: 900 nm – 1300 nm, Detrend, order=2, factors=7; 7: 900 nm – 1300 nm, S-G, points=7, factors=7; 8: 900 nm – 1300 nm, MSC, factors=7; 9:900 nm – 1300 nm, 2nd derivative, s=20, factors=2; 10: 900 nm – 1300 nm, 2nd derivative, s=20, factors=12; 11: 900 nm – 1300 nm, 2nd derivative, s=10, factors=7; 12:900 nm – 1300 nm, 2nd derivative, s=30, factors=7.

Tables 36 and 37 give examples of the t-test and F-test results.

*Table 36. t-Test of Method 1 vs. HPLC Data for Comparison of Means**

Statisticals	*HPLC*	*NIR*
Mean	3.91	3.87
Variance	0.00342	0.00583
Observations	12	12
Pooled Variance	N/A	0.00462
Degrees of Freedom	N/A	22
t Statistical	N/A	1.38
t Critical two-tail	N/A	2.07

*Two-Sample Assuming Equal Variances

Table 37. F-test of Method 1 vs. HPLC Data for Comparison of SD

Statisticals	*HPLC*	*NIR*
Mean	3.91	3.87
Variance	0.00342	0.00583
Observations	12	12
Degree of Freedom	11	11
F Statistical	N/A	1.70
F Critical one-tail	N/A	2.82

9.2.8 Discussion

The t-test and F-test results in Table 35 indicate that Method 1 (900nm – 1300 nm, 2nd derivative, s=20, f=7), Method 4 (900 nm – 1300 nm, 1st derivative, s=20, f=7), Method 8 (900 nm – 1300 nm, MSC, f=7), and Method 12 (900 nm – 1300 nm, 2nd derivative, s=30, f=7) passed both t-test and F-test and produced comparable results to those obtained from HPLC method; Method 2 (850 nm - 1650 nm, 2nd derivative, s=20, f=7) and Method 11 (900 nm – 1300 nm, 2nd derivative, s=10, f=7) failed to pass the F-test and produced significantly different variances but with comparable mean results; Method 3 (raw data), Method 5 (SNV), Method 6 (Detrend), and Method 7 (S-G) failed to pass the t-test and produced significantly different mean results but with comparable variances;

Method 9 (900 nm – 1300 nm, 2nd derivative s=20, f=2) and Method 10 (900 nm – 1300 nm, 2nd derivative, s=20, f=12) failed to pass both t-test and F-test and produced significantly different mean results and variances. The results clearly demonstrated the importance of selecting suitable wavelengths, appropriate mathematical pretreatment method, correct factors, and relative segment size (when the derivative method is used). In general, 2nd derivative and MSC are the two pretreatments used most often and segment size of 20 is often used in derivative method. Once again, different applications may require different methods and it may be necessary to evaluate all the parameters to obtain the most suitable and robust method.

9.3 Simple Linear Regression

Simple linear regression (SLR) is rarely used in NIR quantitation and is discussed here only for comparison purpose. Mathematical pretreatment of 2nd derivative with segment size of 20 was used in this section.

9.3.1 Regression Development

Figure 74 is the correlation and sensitivity plot of SLR which indicates high correlation and reasonable sensitivity at wavelength of 1204 nm. The squared line is the correlation and the two curved lines are the 2nd derivative of the samples containing the lowest and highest amount of constituents.

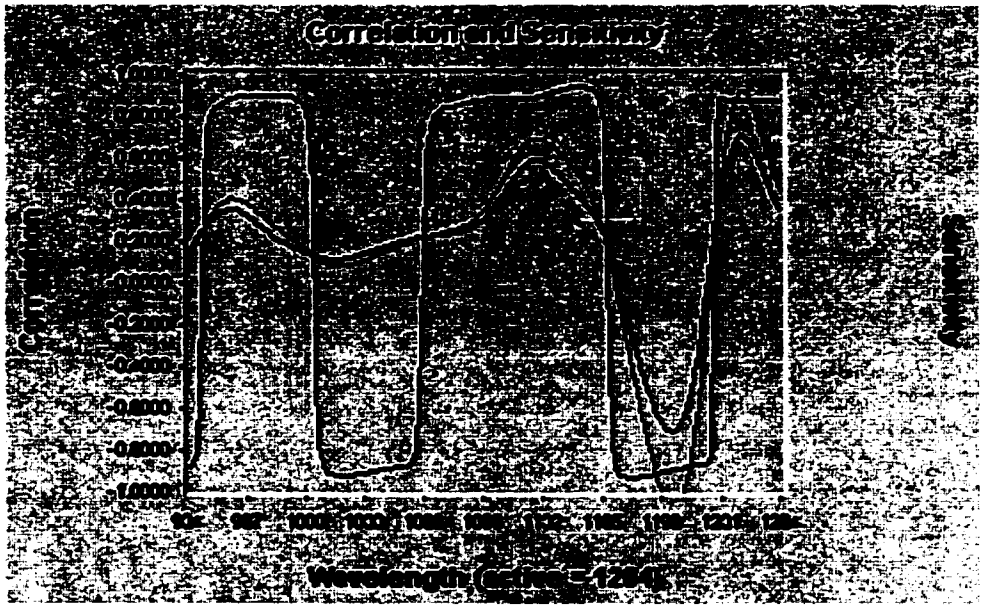

Figure 74. Correlation Coefficient and Sensitivity Plot, SLR Calibration

In Table 38 are the NIR vs. HPLC data of the calibration set at 1204 nm with k(0)=1.3716, k(1)=4.1466, R^2=0.7859, SEC=0.2059, and F-value=124.8. Figures 75 and 76 are the plots of NIR vs. HPLC and NIR residuals.

Table 38. NIR vs. HPLC Data, SLR Calibration

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-01	3.50	3.47	0.03
2	1-02	3.44	3.44	0.00
3	1-03	3.41	3.47	-0.06
4	1-04	3.34	3.22	0.12
5	1-05	3.43	3.38	0.05
6	1-06	3.40	3.41	-0.01
7	1-07	3.35	3.35	0.00
8	1-08	3.35	3.25	0.10
9	1-09	3.25	3.23	0.02
10	1-10	3.34	3.24	0.10
11	1-11	3.41	3.41	0.00
12	1-12	3.35	3.43	0.08
13	2-01	4.16	4.05	0.11
14	2-02	4.20	3.85	0.35
15	2-05	4.17	4.02	0.15
16	2-06	4.31	3.95	0.36
17	2-07	4.25	3.96	0.29
18	2-09	4.15	4.04	0.11
19	2-11	4.19	3.97	0.22
20	3-03	4.15	3.92	0.23
21	3-04	4.07	4.07	0.00
22	3-05	4.33	4.44	-0.11
23	3-06	4.24	4.20	0.04
24	3-07	4.29	4.32	-0.03
25	3-08	4.20	4.02	0.18
26	3-09	4.00	4.10	-0.10
27	3-10	4.27	4.10	0.17
28	3-11	4.05	4.04	0.01
29	3-12	4.17	4.14	0.03
30	4-01	4.13	4.53	-0.40
31	4-02	4.17	4.45	-0.28
32	4-03	4.26	4.43	-0.17
33	4-04	4.16	4.54	-0.38
34	4-05	4.00	4.38	-0.38
35	4-09	4.04	4.51	-0.47
36	4-10	4.25	4.44	-0.19
Mean	N/A	3.91	3.91	SAR: 5.33
SD	N/A	0.39	0.44	

SAR: Sum of Absolute Residuals.

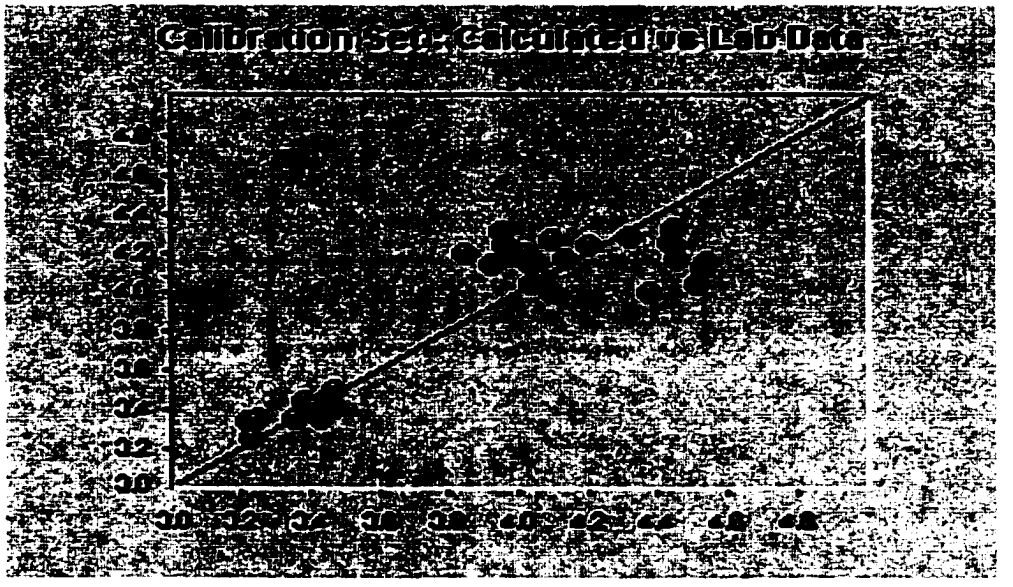

Figure 75. NIR vs. HPLC Plot, SLR Calibration

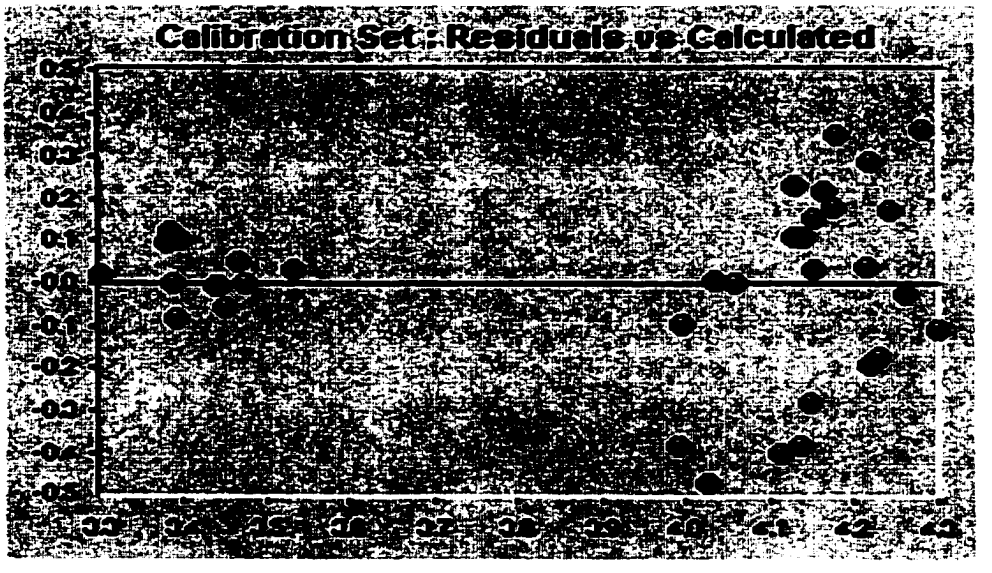

Figure 76. NIR Residual Plot, SLR Calibration

9.3.2 Regression Validation

Table 39 gives the NIR vs. HPLC results of the validation set with SLR at 1204 nm and Figures 77 and 78 are the plots of the NIR vs. HPLC data and the NIR residuals.

Table 39. NIR vs. HPLC Data, SLR Validation

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	2-03	4.02	3.94	0.08
2	2-04	4.30	3.97	0.33
3	2-08	4.21	3.92	0.29
4	2-10	4.33	3.97	0.36
5	2-12	4.36	4.23	0.13
6	3-01	3.98	4.05	-0.07
7	3-02	4.16	4.31	-0.15
8	4-06	4.33	4.52	-0.19
9	4-07	4.16	4.50	-0.34
10	4-18	4.21	4.53	-0.32
11	4-11	4.30	4.53	-0.23
12	4-12	4.20	4.51	-0.31
Mean	N/A	4.21	4.25	SAR: 2.80
SD	N/A	0.12	0.26	

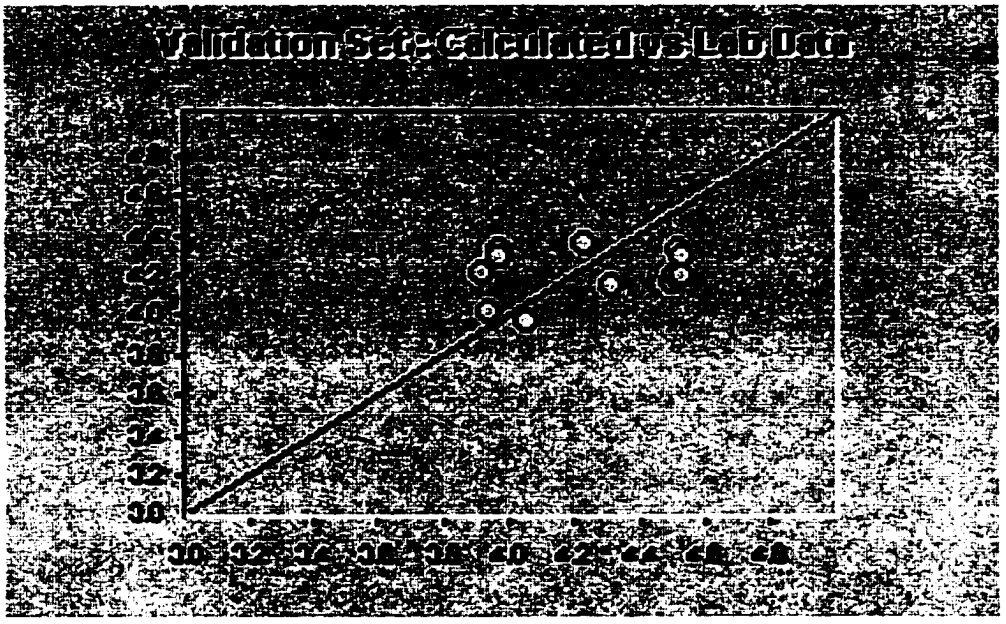

Figure 77. NIR vs. HPLC Plot, SLR Validation

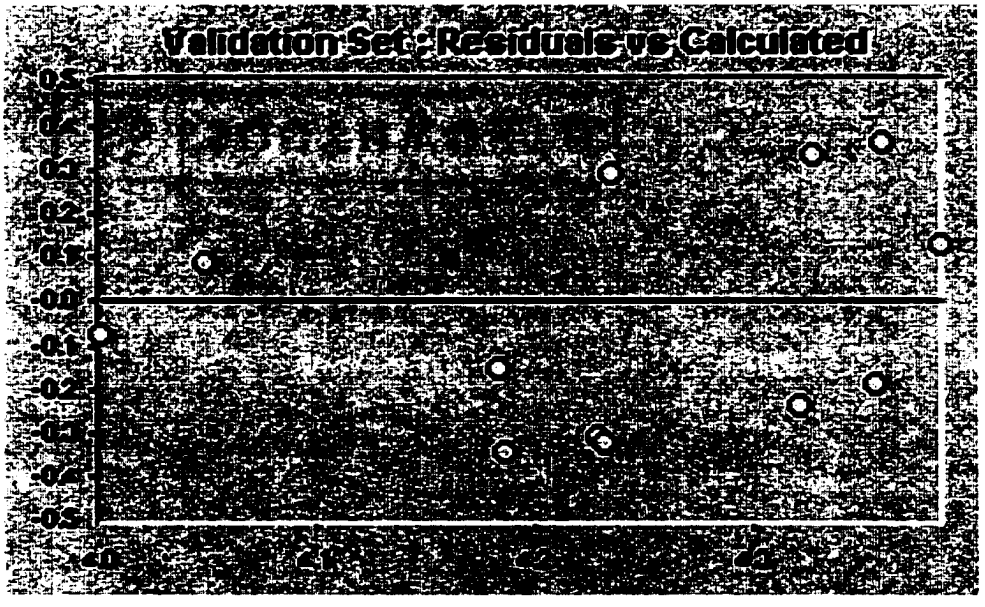

Figure 78. NIR Residual Plot, SLR Validation

9.3.3 Prediction of Unknown Samples

In Table 40 are the NIR vs. HPLC data for the prediction of the unknown samples with SLR at 1204 nm.

Table 40. NIR vs. HPLC Data, SLR Prediction

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	U-01	3.60	3.80	-0.20
2	U-02	3.69	3.91	-0.22
3	U-03	3.03	3.86	-0.83
4	U-04	3.70	3.86	-0.16
5	U-05	3.67	3.85	-0.18
6	U-06	3.61	3.89	-0.28
7	U-07	3.48	3.97	-0.49
8	U-08	3.73	3.93	-0.20
9	U-09	3.60	3.94	-0.34
10	U-10	3.72	4.00	-0.28
11	U-11	3.62	3.91	-0.29
12	U-12	3.66	3.97	-0.31
Mean	N/A	3.59	3.91	SAR: 3.78
SD	N/A	0.19	0.058	

9.3.4 Discussion

The calibration and validation results show high variations with SLR method. The predicted results of unknown samples show clear large biases, which indicate that the SLR method is not robust and not suitable for use in NIR quantitation here. The results of t-test and the F-test are discussed in Section 9.5.

9.4 Multiple Linear Regression

Multiple Linear Regression (MLR) is a least squares method that uses spectral information at several wavelengths. Three wavelengths of 1038 nm, 1156 nm, and 1204 nm were used here.

Since the multiple correlation coefficient is a number which doesn't necessarily tell all about a situation, especially about the reliability of predicting future samples, the standard error of calibration (SEC) and the statistic F for regression are also used to select suitable wavelengths (which wavelengths and how many) for quantitation.

9.4.1 Regression Development

Figures 79-81 show the correlation and sensitivity plots at wavelength 1204 nm, 1156 nm, and 1038 nm, respectively. Table 41 gives the NIR vs. HPLC data of MLR calibration with k(0)=3.339, k(1)=15.5138, k(2)=16.1608, R^2=0.9554, SEC=0.0982, and F-value=228.3.

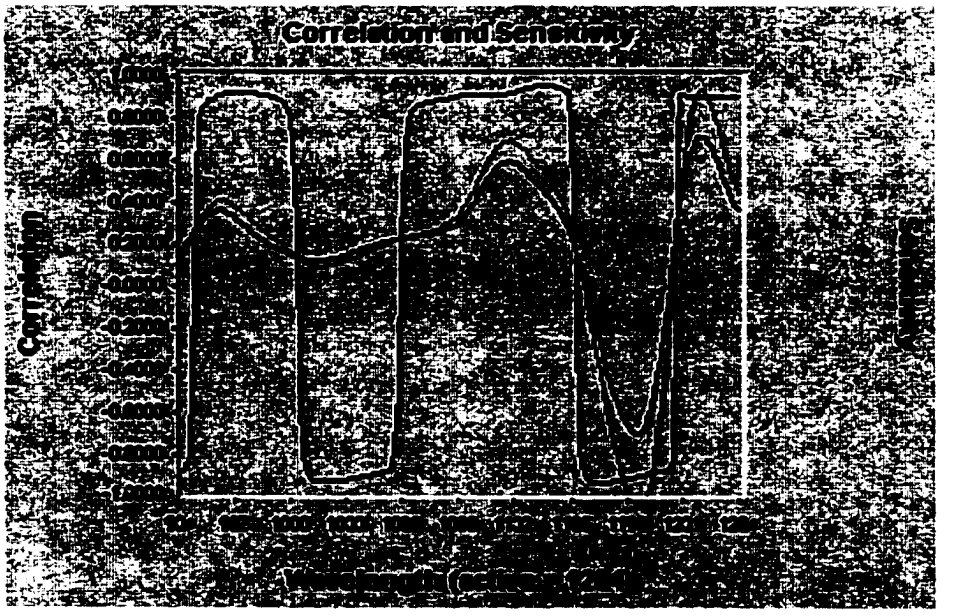

Figure 79. Correlation & Sensitivity Plot, MLR Calibration, 1204 nm

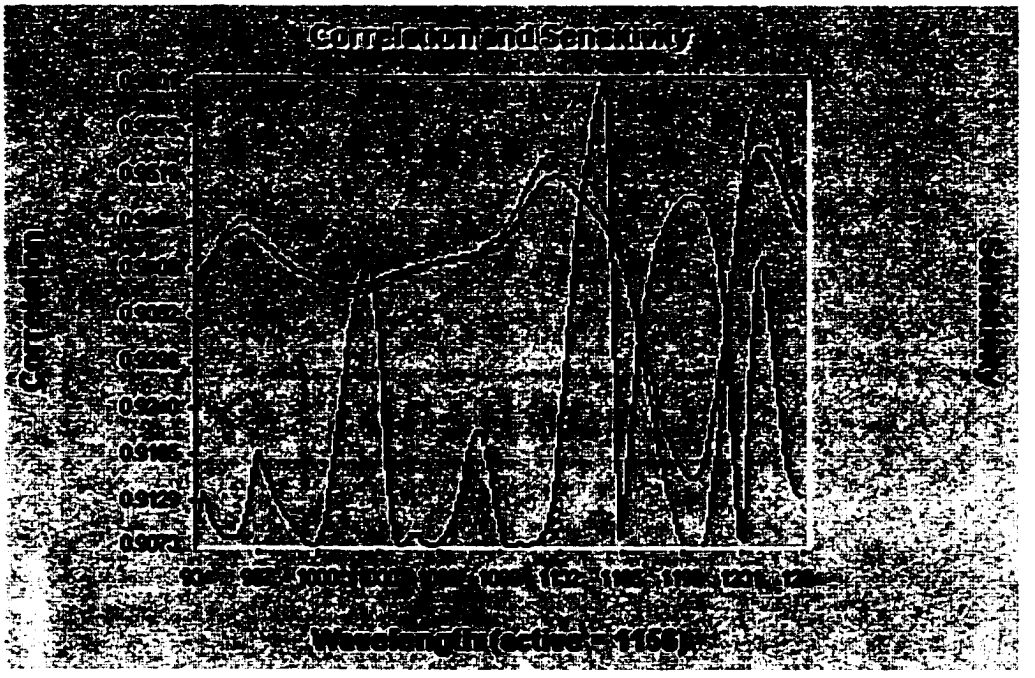

Figure 80. Correlation and Sensitivity Plot, MLR Calibration, 1156 nm

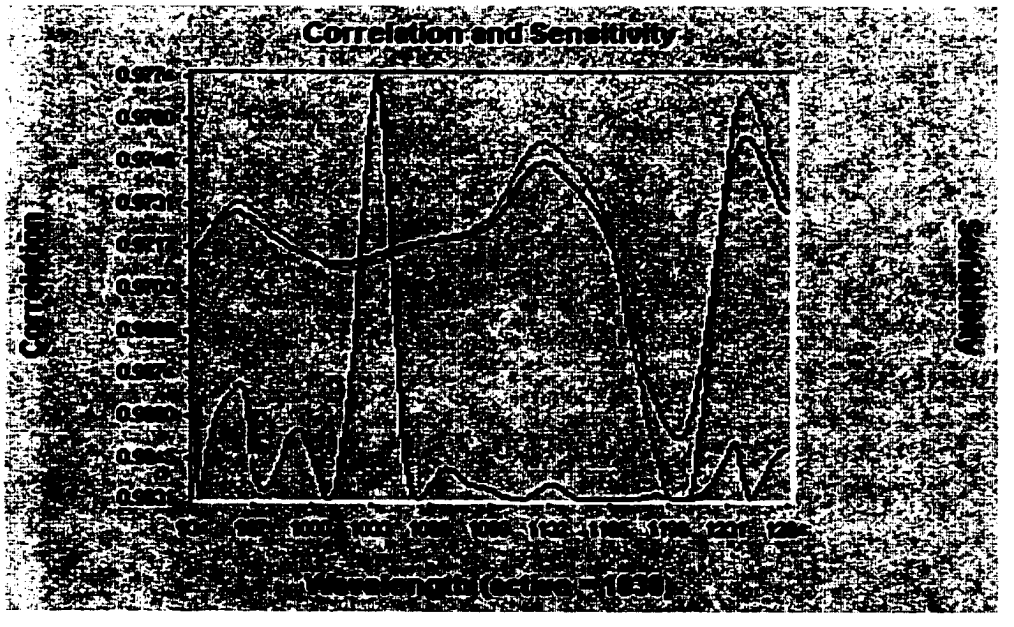

Figure 81. Correlation and Sensitivity Plot, MLR Calibration, 1038 nm

Table 41. NIR vs. HPLC Data, MLR Calibration

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-01	3.36	3.47	-0.11
2	1-02	3.33	3.44	-0.11
3	1-03	3.45	3.47	-0.02
4	1-04	3.28	3.22	0.06
5	1-05	3.39	3.38	0.01
6	1-06	3.45	3.41	0.04
7	1-07	3.43	3.35	0.08
8	1-08	3.40	3.25	0.15
9	1-09	3.24	3.23	0.01
10	1-10	3.37	3.24	0.13
11	1-11	3.27	3.41	-0.14
12	1-12	3.38	3.43	-0.05
13	2-01	3.87	4.05	-0.18
14	2-02	3.90	3.94	-0.04
15	2-05	4.04	3.95	0.09
16	2-06	3.96	3.97	-0.01
17	2-07	3.95	3.96	-0.01
18	2-09	3.98	3.92	0.06
19	2-11	3.93	3.97	-0.04
20	3-03	4.24	4.20	0.04
21	3-04	4.35	4.32	0.03
22	3-05	4.20	4.02	0.18
23	3-06	4.07	4.10	-0.03
24	3-07	4.29	4.10	0.19
25	3-08	4.10	4.04	0.06
26	3-09	4.33	4.23	0.10
27	3-10	4.06	4.05	0.01
28	3-11	4.19	4.14	0.05
29	3-12	4.26	4.31	-0.05
30	4-01	4.49	4.52	-0.03
31	4-02	4.35	4.53	-0.18
32	4-03	4.38	4.45	-0.07
33	4-04	4.51	4.43	0.08
34	4-05	4.41	4.50	-0.09
35	4-09	4.40	4.53	-0.13
36	4-10	4.49	4.53	-0.04
Mean	N/A	3.91	3.92	SAR: 2.70
SD	N/A	0.44	0.43	

SAR: Sum of Absolute Residuals.

Figure 82 is the NIR vs. HPLC plot with MLR and Figure 83 is the NIR residual plot of the calibration.

Figure 82. NIR vs. HPLC Plot, MLR Calibration

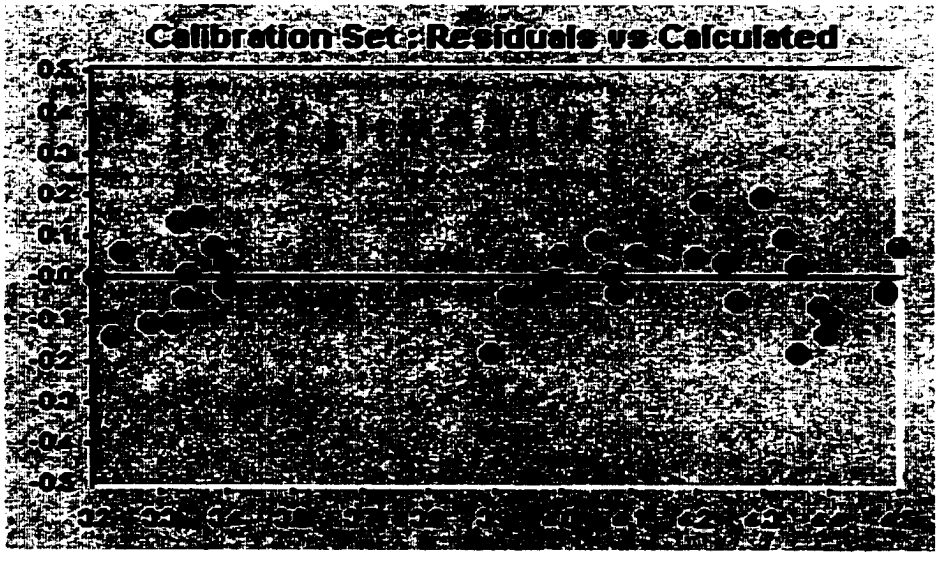

Figure 83. NIR Residual Plot, MLR Calibration

9.4.2 Regression Validation

Table 42 gives the NIR vs. HPLC of the validation set and Figures 84 and 85 are the plots of NIR vs. HPLC and NIR residuals, respectively, with MLR.

Table 42. NIR vs. HPLC Data, MLR Validation

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	2-03	3.95	3.85	0.10
2	2-04	4.05	4.02	0.03
3	2-08	4.02	4.04	-0.02
4	2-10	3.99	3.97	0.02
5	2-12	4.01	3.92	0.09
6	3-01	4.17	4.07	0.10
7	3-02	4.46	4.44	0.02
8	4-06	4.42	4.54	-0.12
9	4-07	4.37	4.38	-0.01
10	4-18	4.32	4.51	-0.19
11	4-11	4.48	4.44	0.04
12	4-12	4.47	4.51	-0.04
Mean	N/A	4.23	4.22	SAR: 0.78
SD	N/A	0.21	0.27	

Figure 84. NIR vs. HPLC Plot, MLR Validation

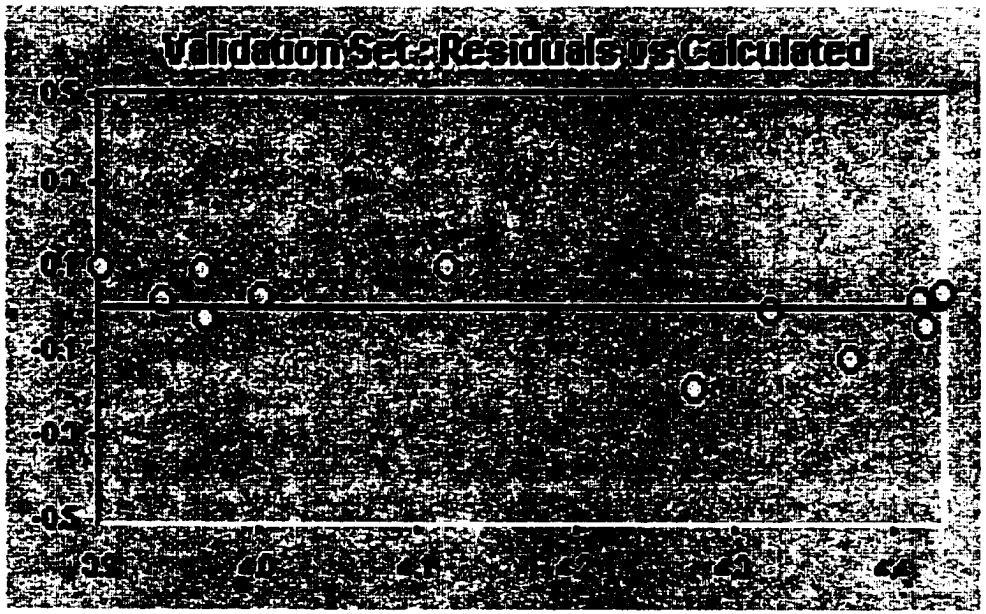

Figure 85. NIR Residuals Plot, MLR Validation

9.4.3 Prediction of Unknown Samples

In Table 43 are the NIR predicted results of unknown gum samples with MLR at 1038 nm, 1156 nm, and 1204 nm.

Table 43. NIR vs. HPLC Data, MLR Prediction

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	U-01	4.00	3.80	0.20
2	U-02	4.10	3.91	0.19
3	U-03	4.13	3.86	0.27
4	U-04	4.14	3.86	0.28
5	U-05	4.04	3.85	0.19
6	U-06	4.11	3.89	0.22
7	U-07	4.12	3.97	0.15
8	U-08	4.19	3.93	0.26
9	U-09	4.18	3.94	0.24
10	U-10	4.06	4.00	0.06
11	U-11	4.03	3.91	0.12
12	U-12	4.12	3.97	0.15
Mean	N/A	4.10	3.91	SAR: 2.33
SD	N/A	0.059	0.058	

9.4.4 Discussion

The calibration and validation results of MLR demonstrate the good fit of between NIR and HPLC results. However, the NIR predicted results of unknown samples showed certain level of bias on the high side. The t-test and F-test results are discussed in Section 9.5.

9.5 Comparison of SLR, MLR, and PLSR

In Table 44 are the NIR predicted results of unknown gum samples with PLSR, MLR, and SLR methods. These show that the PLSR method is superior to both MLR and SLR, and MLR is better than SLR

although the calibration and validation results of PLSR and MLR are comparable. The t-test and F-test results indicate that compared to the HPLC results, PLSR produced comparable mean and variance, MLR produced comparable variance but significantly different mean, and SLR produced significantly different variance and mean.

Table 44. NIR Predicted Results of Unknowns by Different Regressions

No.	Sample Name	HPLC mg/tablet	PLSR mg/tablet	MLR mg/tablet	SLR mg/tablet
1	U-01	3.80	3.82	4.00	3.60
2	U-02	3.91	3.90	4.10	3.69
3	U-03	3.86	3.67	4.13	3.03
4	U-04	3.86	3.94	4.14	3.70
5	U-05	3.85	3.87	4.04	3.67
6	U-06	3.89	3.87	4.11	3.61
7	U-07	3.97	3.85	4.12	3.48
8	U-08	3.93	3.97	4.19	3.73
9	U-09	3.94	3.92	4.18	3.60
10	U-10	4.00	3.89	4.06	3.72
11	U-11	3.91	3.83	4.03	3.62
12	U-12	3.97	3.90	4.12	3.66
Mean		3.91	3.87	4.10	3.59
Min		3.80	3.67	4.00	3.03
Max		4.00	3.97	4.19	3.73
Max – Min		0.20	0.30	0.19	0.70
SD		0.058	0.076	0.059	0.190
t-Test (critical =2.07)			1.38		
F-test (critical=2.82)			1.70	1.01	

Shaded area indicates the significant difference.

10. DATA ANALYSIS – REFLECTANCE NIR

An additional twelve tablets were selected from the same five batches used in transmittance NIR. Samples from batch 1-4 were used for calibration and validation. Batch 5 was treated as unknown samples and the potencies were predicted by NIR and then compared to the results obtained from the HPLC procedure.

10.1 Sample Presentation

All the gum samples were scanned with one presentation directly on the reflectance NIR spectrophotometer with a cover on to protect the samples from stray light.

10.2 Regression Development

Based on the results from the studies performed on transmittance NIR, PLSR with 2nd derivative and segment size=20 were used. The effect of wavelength range on reflectance NIR was investigated since reflectance NIR acquired a wider wavelength range than transmittance and used both silicon and lead sulfide detectors with switching at 1100 nm. A total of 36 samples were used in the calibration set and 12 samples were used in the validation set.

10.2.1 Range 750 nm – 2250 nm

The NIR spectra were acquired from 400 nm – 2500 nm and the range 750 nm – 2250 nm was used here to avoid the peak before 700 nm, which might be associated with Queniline Yellow, and the noisy baseline above 2250 nm (refer to Figure 36 for the details). Figures 86-88 are the loading, correlation coefficient, and PRESS plots, respectively.

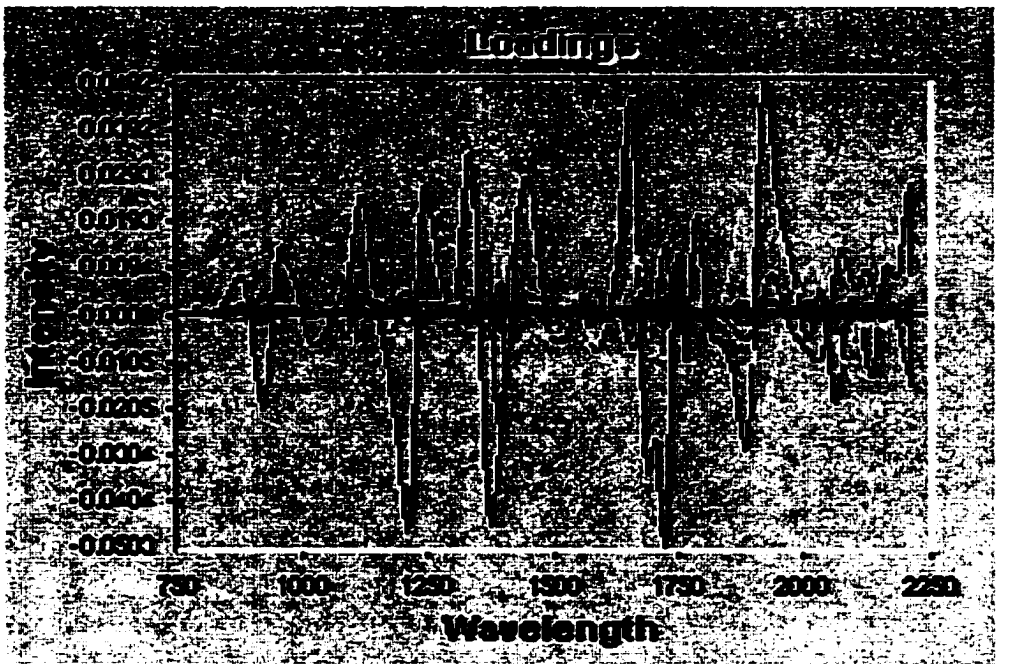

Figure 86. Loading Plot, R-PLSR Calibration, 750 nm – 2250 nm

Figure 87. Coefficient Plot R-PLSR Calibration, 750 nm – 2250 nm

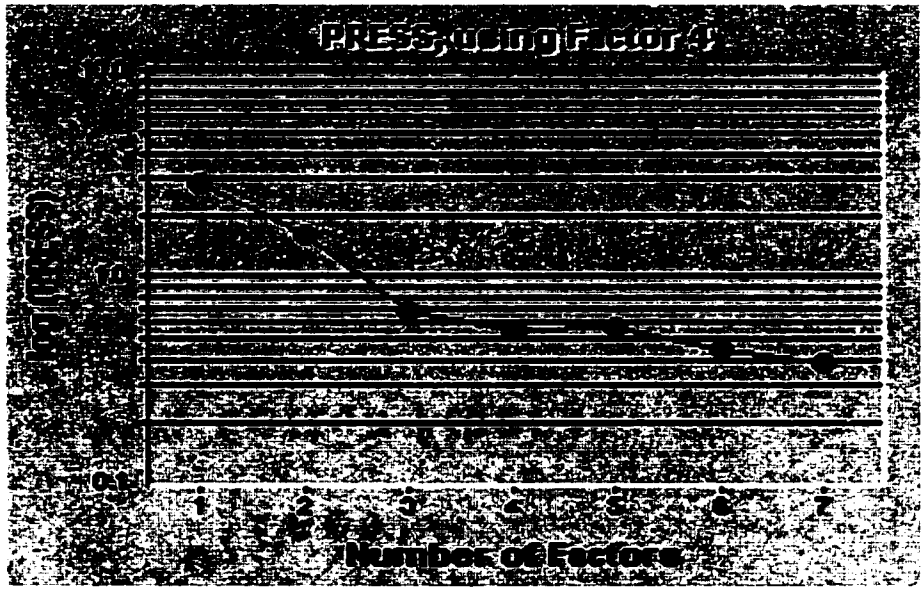

Figure 88. PRESS Plot R-PLSR Calibration, 750 nm – 2250 nm

Table 45 gives the statistical results and Table 46 the NIR vs. HPLC data of the calibration set. Figures 89 and 90 show the plots of NIR vs. HPLC Data and NIR residuals.

Table 45. Statistical Results, R-PLSR Calibration, 750 nm – 2250 nm

Factor	R^2	SEC	PRESS	F-value
1	0.6547	0.2802	2.8014	64.5
2	0.8175	0.2068	1.6007	73.9
3	0.9309	0.1292	0.6945	143.8
4*	0.9453	0.1168	0.5590	133.9
5	0.9649	0.0951	0.5669	164.9
6	0.9770	0.0783	0.4459	205.2
7	0.9789	0.0763	0.3855	185.7
8	0.9821	0.0715	0.5848	185.6
9	0.9858	0.0651	0.8822	200.0
10	0.9883	0.0602	1.0047	210.5
11	0.9903	0.0560	0.9926	222.1
12	0.9937	0.0461	0.8897	301.5
13	0.9966	0.0346	0.9311	493.8
14	0.9971	0.0328	0.8950	510.9
15	0.9983	0.0253	0.9393	806.7

**the number of factors used in calibration.*

Table 46. NIR vs. HPLC, R-PLSR Calibration, 750 nm – 2250 nm

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-01	2.88	2.88	0.00
2	1-03	3.49	3.47	0.02
3	1-04	3.44	3.49	-0.05
4	1-06	3.61	3.54	0.07
5	1-07	3.40	3.41	-0.01
6	1-09	3.37	3.39	-0.02
7	1-10	3.50	3.38	0.12
8	1-11	3.33	3.32	0.01
9	1-12	3.32	3.37	-0.05
10	2-01	3.90	3.90	0.00
11	2-02	4.12	3.79	0.33
12	2-04	3.87	3.92	-0.05
13	2-05	3.79	3.86	-0.07
14	2-07	3.83	4.01	-0.18
15	2-08	3.85	3.80	0.05
16	2-10	3.98	3.97	0.01
17	2-11	4.12	3.94	0.18
18	2-12	3.89	4.03	-0.14
19	3-01	4.38	4.20	0.18
20	3-02	4.18	4.20	-0.02
21	3-03	4.31	4.18	0.13
22	3-05	4.39	4.25	0.14
23	3-06	4.28	4.31	-0.03
24	3-07	4.36	4.21	0.15
25	3-09	4.33	4.33	0.00
26	3-11	4.15	4.28	-0.13
27	3-12	4.15	4.21	-0.06
28	4-01	4.57	4.53	0.04
29	4-02	4.59	4.64	-0.05
30	4-03	4.61	4.60	0.01
31	4-05	4.52	4.61	-0.09
32	4-06	4.58	4.63	-0.05
33	4-07	4.50	4.53	-0.03
34	4-09	4.47	4.60	-0.13
35	4-10	4.37	4.58	-0.21
36	4-12	4.48	4.56	-0.08
Mean	N/A	4.03	4.03	SAR: 2.89
SD	N/A	0.46	0.47	

SAR: Sum of Absolute Residuals.

Figure 89. NIR vs. HPLC Plot, R-PLSR Calibration, 750nm – 2250 nm

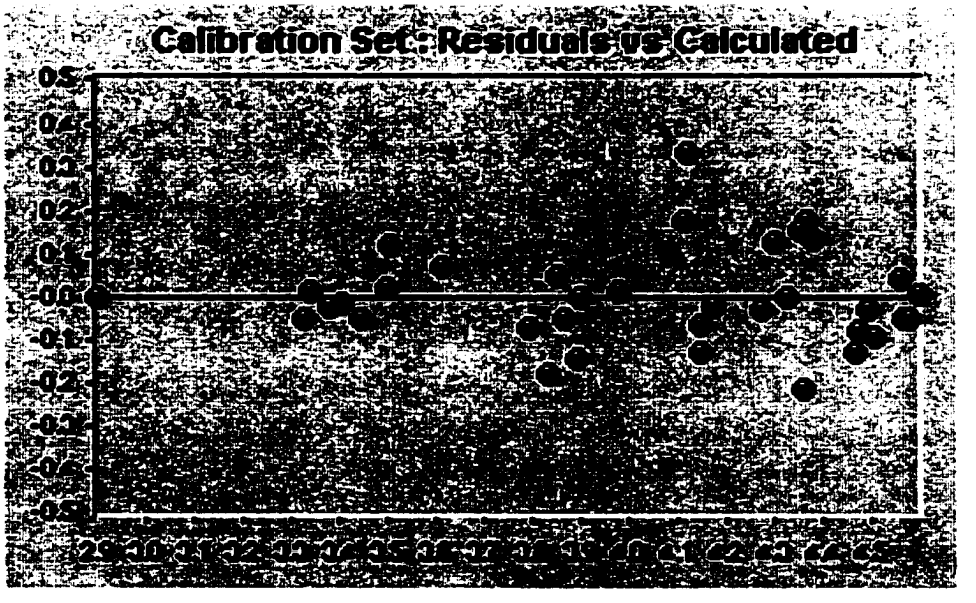

Figure 90. NIR Residual Plot, R-PLSR Calibration, 750nm – 2250 nm

10.2.2 Range 750 nm – 1050 nm Plus 1150 nm – 2250 nm

The calibration equation using two wavelength ranges, 750 nm – 1050 nm and 1150 nm – 2250 nm was evaluated to remove the detector switching effect at 1100 nm. Figure 91-93 are the PRESS plot, NIR vs. HPLC data plot, and NIR residuals plot, respectively.

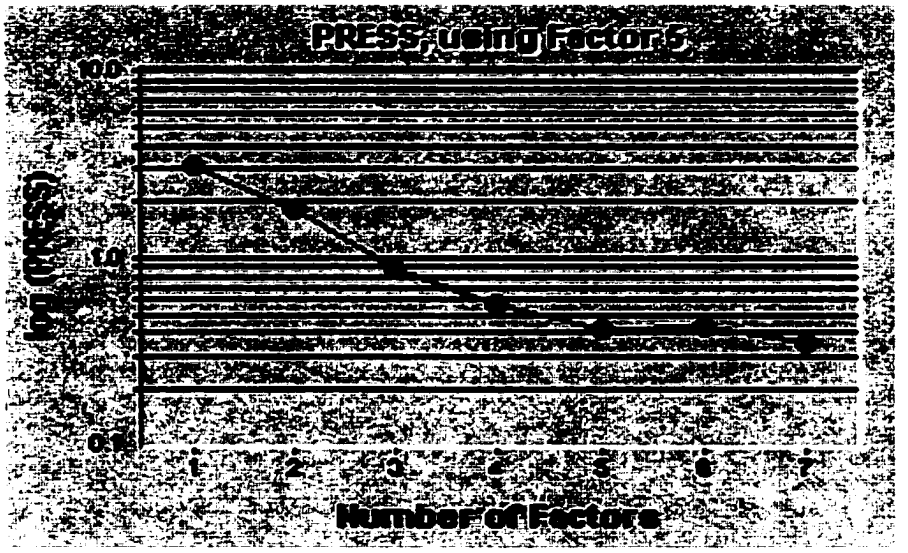

Figure 91. PRESS Plot, 750 nm – 1050 nm & 1150 nm – 2250 nm

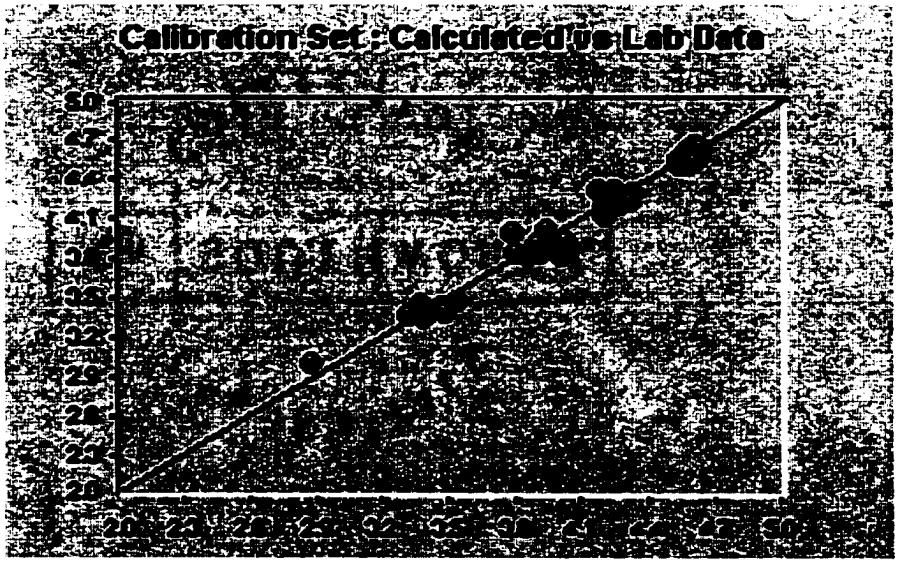

Figure 92. NIR vs. HPLC Plot, 750 nm – 1050 nm & 1150 nm – 2250 nm

Figure 93. NIR Residual Plot, 750 nm – 1050 nm & 1150 nm – 2250 nm

Table 47 gives the statistical results and Table 48 the NIR vs. HPLC results for the calibration set.

Table 47. Statistical Results, 750 nm – 1050 nm & 1150 nm – 2250 nm

Factor	R^2	SEC	PRESS	F-value
1	0.6410	0.2857	3.0919	60.7
2	0.7821	0.2259	1.8055	59.2
3	0.9196	0.1393	0.8826	122.1
4	0.9464	0.1156	0.5587	136.9
5*	0.9697	0.0884	0.4125	191.8
6	0.9763	0.0795	0.4277	199.2
7	0.9817	0.0711	0.3489	214.6
8	0.9843	0.0670	0.6037	211.9
9	0.9877	0.0604	0.7278	232.3
10	0.9902	0.0551	0.7854	252.5
11	0.9925	0.0493	0.8143	287.4
12	0.9955	0.0390	0.7344	420.7
13	0.9963	0.0363	0.6487	450.0
14	0.9976	0.0295	0.6575	633.4
15	0.9989	0.0209	0.6233	1179.6

**the number of factors used in calibration.*

Table 48. NIR vs. HPLC Data, 750 nm – 1050 nm & 1150 nm – 2250 nm

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-01	3.00	2.88	0.12
2	1-03	3.44	3.47	-0.03
3	1-04	3.39	3.49	-0.10
4	1-06	3.47	3.54	-0.07
5	1-07	3.39	3.41	-0.02
6	1-09	3.36	3.39	-0.03
7	1-10	3.46	3.38	0.08
8	1-11	3.38	3.32	0.06
9	1-12	3.42	3.37	0.05
10	2-01	3.92	3.90	0.02
11	2-02	4.00	3.79	0.21
12	2-04	3.84	3.92	-0.08
13	2-05	3.82	3.86	-0.04
14	2-07	3.81	4.01	-0.20
15	2-08	3.84	3.80	0.04
16	2-10	3.94	3.97	-0.03
17	2-11	4.01	3.94	0.07
18	2-12	3.94	4.03	-0.09
19	3-01	4.29	4.20	0.09
20	3-02	4.13	4.20	-0.07
21	3-03	4.31	4.18	0.13
22	3-05	4.33	4.25	0.08
23	3-06	4.23	4.31	-0.08
24	3-07	4.31	4.21	0.10
25	3-09	4.31	4.33	-0.02
26	3-11	4.21	4.28	-0.07
27	3-12	4.20	4.21	-0.01
28	4-01	4.52	4.53	-0.01
29	4-02	4.58	4.64	-0.06
30	4-03	4.67	4.60	0.07
31	4-05	4.54	4.61	-0.07
32	4-06	4.64	4.63	0.01
33	4-07	4.56	4.53	0.03
34	4-09	4.54	4.60	-0.06
35	4-10	4.51	4.58	-0.07
36	4-12	4.61	4.56	0.05
Mean	N/A	4.03	4.03	SAR: 2.42
SD	N/A	0.46	0.47	

SAR: Sum of Absolute Residuals.

10.2.3 Range 900 nm – 1300 nm

The calibration equation was evaluated with the same wavelength range as the transmittance NIR so that the results could be compared directly.

Table 49 gives the summary statistical results and Table 50 the NIR vs. HPLC data of reflectance NIR calibration in the range of 900 nm – 1300 nm.

Table 49. Statistical Results, R-PLSR Calibration, 900 nm – 1300 nm

Factor	R^2	SEC	PRESS	F-value
1	0.7216	0.2516	2.2796	88.1
2	0.8686	0.1755	1.1776	109.1
3	0.9136	0.1445	0.8853	112.8
4*	0.9269	0.1350	0.8037	98.2
5	0.9339	0.1305	0.8012	84.7
6	0.9518	0.1134	1.2712	95.4
7	0.9735	0.0855	1.5371	147.0
8	0.9862	0.0628	0.9273	241.4
9	0.9890	0.0572	0.6076	259.3
10	0.9894	0.0573	0.4124	233.1
11	0.9906	0.0550	0.5582	229.7
12	0.9919	0.0521	1.5618	235.6
13	0.9929	0.0499	1.6753	237.2
14	0.9952	0.0420	1.9506	311.7
15	0.9961	0.0387	1.1768	342.8

**the number of factors used in calibration.*

Table 50. NIR vs. HPLC Data, R-PLSR Calibration, 900 nm – 1300 nm

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-01	3.14	2.88	0.26
2	1-03	3.45	3.47	-0.02
3	1-04	3.42	3.49	-0.07
4	1-06	3.56	3.54	0.02
5	1-07	3.32	3.41	-0.09
6	1-09	3.27	3.39	-0.12
7	1-10	3.42	3.38	0.04
8	1-11	3.40	3.32	0.08
9	1-12	3.39	3.37	0.02
10	2-01	3.90	3.90	0.00
11	2-02	4.11	3.79	0.32
12	2-04	3.80	3.92	-0.12
13	2-05	3.80	3.86	-0.06
14	2-07	3.79	4.01	-0.22
15	2-08	3.81	3.80	0.01
16	2-10	3.99	3.97	0.02
17	2-11	4.10	3.94	0.16
18	2-12	3.93	4.03	-0.10
19	3-01	4.36	4.20	0.16
20	3-02	4.07	4.20	-0.13
21	3-03	4.32	4.18	0.14
22	3-05	4.40	4.25	0.15
23	3-06	4.27	4.31	-0.04
24	3-07	4.38	4.21	0.17
25	3-09	4.34	4.33	0.01
26	3-11	4.18	4.28	-0.10
27	3-12	4.17	4.21	-0.04
28	4-01	4.48	4.53	-0.05
29	4-02	4.49	4.64	-0.15
30	4-03	4.65	4.60	0.05
31	4-05	4.52	4.61	-0.09
32	4-06	4.63	4.63	0.00
33	4-07	4.53	4.53	0.00
34	4-09	4.49	4.60	-0.11
35	4-10	4.43	4.58	-0.15
36	4-12	4.60	4.56	0.04
Mean	N/A	4.03	4.03	SAR: 3.31
SD	N/A	0.45	0.47	

SAR: Sum of Absolute Residuals.

Figures 94 – 96 are the PRESS plot, NIR vs. HPLC plot, and NIR residual plot of reflectance NIR calibration in range of 900 nm – 1300 nm, respectively.

Figure 94. PRESS Plot, R-PLSR Calibration, 900 nm – 1300 nm

Figure 95. NIR vs. HPLC Plot, R-PLSR Calibration, 900 nm – 1300 nm

Figure 96. NIR Residual Plot, R-PLSR Calibration, 900 nm – 1300 nm

10.3 Regression Validation

The validation results of the calibration sets are very similar to those of the calibration set since the validation sets were used as cross validation during the regression development process. Table 51 gives NIR vs. HPLC data and Figures 97 and 98 the plot of NIR vs. HPLC data and NIR residuals of the reflectance NIR validation set in range of 900 nm – 1300 nm, respectively

Table 51. NIR vs. HPLC Data, R-PLSR Validation, 900 nm – 1300 nm

No.	Sample Name	NIR Results (mg/tablet)	HPLC Results (mg/tablet)	Residual (mg/tablet)
1	1-02	3.41	3.42	-0.01
2	2-05	3.34	3.49	-0.15
3	2-08	3.40	3.51	-0.11
4	2-03	3.93	3.84	0.09
5	2-06	3.61	3.89	-0.28
6	2-09	4.11	3.98	0.13
7	3-04	4.20	4.23	-0.03
8	3-08	4.29	4.22	0.07
9	3-10	4.34	4.27	0.07
10	4-04	4.60	4.57	0.03
11	4-08	4.75	4.78	-0.03
12	4-11	4.60	4.53	0.07
Mean	N/A	4.05	4.06	SAR: 1.07
SD	N/A	0.51	0.45	

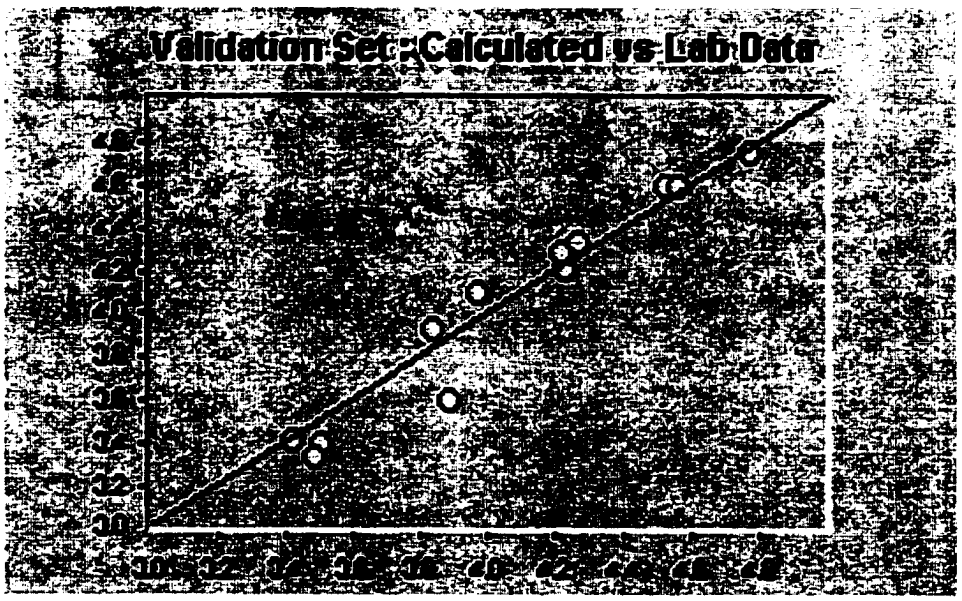

Figure 97. NIR vs. HPLC Plot, R-PLSR Validation, 900 nm – 1300 nm

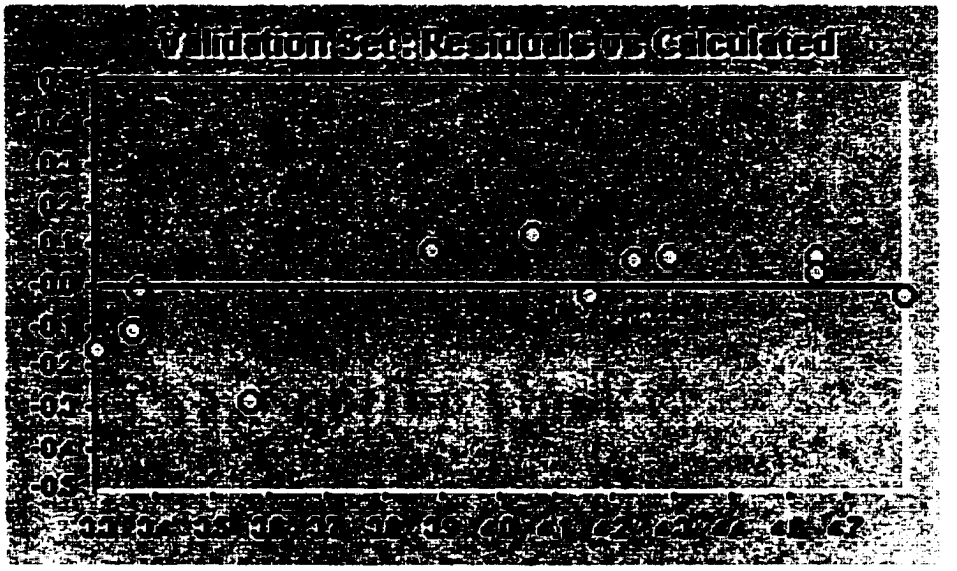

Figure 98. NIR Residuals Plot, R-PLSR Validation, 900 nm – 1300 nm

10.4 Prediction of Unknown Samples

Table 52 gives the reflectance NIR predicted results of unknown samples analyzed by the methods with different wavelength ranges.

Table 52. Reflectance NIR Predicted Results (mg/tablet)

No.	Sample Name	HPLC	750 nm - 2250 nm	750-1050 & 1150-2250	900 nm - 1300 nm
1	U-01	3.95	4.13	4.22	4.24
2	U-02	3.92	4.28	4.37	4.39
3	U-03	3.93	4.24	4.24	4.30
4	U-04	3.85	3.64	3.51	3.97
5	U-05	3.80	4.31	4.37	4.54
6	U-06	3.81	3.87	4.00	3.91
7	U-07	3.92	4.46	4.44	4.65
8	U-08	3.94	4.18	4.30	4.37
9	U-09	3.92	4.05	4.24	4.33
10	U-10	3.84	4.21	4.31	4.45
11	U-11	3.89	4.16	4.30	4.38
12	U-12	3.91	4.18	4.29	4.42
Mean		3.89	4.14	4.22	4.33
Min		3.80	3.64	3.51	3.91
Max		3.95	4.46	4.44	4.65
Max – Min		0.15	0.82	0.93	0.74
SD		0.0517	0.2133	0.2472	0.2114
t-Test (critical=2.07)					
F-test (critical=2.82)					

Shaded area indicates significant difference.

10.5 Discussion

It is very clear that all three procedures produced means and variances significantly different from those obtained from the HPLC procedure. The t-test and F-test results at all three wavelength ranges are higher than the critical values.

Although there are no significant differences between the wavelength ranges used, it is generally recommended to avoid the detector switching wavelength if more than one type of detectors used.

11. COMPARISON OF TRANSMITTANCE AND REFLECTANCE NIR

Table 53 are the predicted results of unknown samples using transmittance and reflectance NIR spectrophotometers with PLSR calibration, 2nd derivative (s=20), in the wavelength range of 900 nm – 1300 nm. The t-test and F-test results indicate that transmittance NIR produced results with comparable mean and variance to those obtained by HPLC procedure and reflectance NIR produced significant different mean and variance.

Table 53. Predicted Results of Transmittance and Reflectance NIR

No.	Sample Name	Transmittance mg/tablet		Reflectance, mg/tablet	
		HPLC	NIR	HPLC	NIR
1	U-01	3.80	3.82	3.95	4.24
2	U-02	3.91	3.90	3.92	4.39
3	U-03	3.86	3.67	3.93	4.30
4	U-04	3.86	3.94	3.85	3.97
5	U-05	3.85	3.87	3.80	4.54
6	U-06	3.89	3.87	3.81	3.91
7	U-07	3.97	3.85	3.92	4.65
8	U-08	3.93	3.97	3.94	4.37
9	U-09	3.94	3.92	3.92	4.33
10	U-10	4.00	3.89	3.84	4.45
11	U-11	3.91	3.83	3.89	4.38
12	U-12	3.97	3.90	3.91	4.42
Mean		3.91	3.87	3.89	4.33
Min		3.80	3.67	3.80	3.91
Max		4.00	3.97	3.95	4.65
Max – Min		0.20	0.30	0.15	0.74
SD		0.058	0.076	0.0517	0.2114
t-Test*			1.38	N/A	
F-test (critical=2.82)			1.70	N/A	

**the critical value is 2.07 for the test with equal variances (transmittance NIR) and 2.18 for the test with unequal variances (reflectance NIR).*

12. SUMMARY

Quantitative NIR is a powerful tool for on-line and off-line analysis with great advantages of no sample destruction, no sample preparation, speed, simplicity, and significant cost saving.

The first step in quantitative NIR is sample collection for calibration and validation. To obtain a robust quantitative method, the samples used in the study must cover all the possible variables that the real life samples may contain, such as variations in particle size, instrument, water content, temperature, and different lots of ingredients.

The second step is the sample selection in which the samples are divided into three sets: calibration set, validation set, and outliers, with different selection methods available. In general, a random selection method can be used if no samples should be treated as outliers (all samples will be randomly selected and placed into the calibration set and validation set according to the defined ratio or numbers). Maximum distance in wavelength space and Mahalanobis distance in principal component space methods can be used to detect and reject the outliers so that the calibration and validation sets will not contain samples with outlying spectra caused by error. Outliers should not be rejected without careful evaluation since some "outliers" may contain valid information. Different mathematical pretreatments of the spectra can be used (for maximum distance and Mahalanobis distance methods) for the samples to remove baseline noise, shifting, scattering, and background changes prior to the selection process.

The next step is the calibration method development. Different calibration methods are available for different applications with different

mathematical pretreatments. The calibrations used most often are partial least squares regression (PLSR) and multiple linear regression (MLR). When PLSR is used, the PRESS, R^2, SEC, F-value, correlation coefficient plot, and loading plot should be used to select the appropriate number of factors. When MLR is used, the correlation coefficient and sensitivity plot, R^2, SEC, and F-value should be used to select the appropriate wavelengths to avoid colinearity and calibration overfitting.

Method validation is the next step where the validation set will be used for checking the validity of the calibration equation using the samples from the same batches as the calibration set. If necessary, the calibration equation bias and/or slope can be adjusted during this stage. However, care must be taken when adjusting the bias and/or slope since fitting the calibration equation well during validation does not necessarily mean better results for the prediction of unknown samples.

This research studied three chemometric calibration methods: simple linear regression (SLR for single wavelength), multiple linear regression (MLR for 2-3 wavelengths), and partial least squares regression (PLSR for wavelength ranges) and compared their differences.

An efficient HPLC method was developed and validated for linearity, precision, accuracy, specificity, and robustness so that the true potencies of the nicotine gum samples could be determined, the calibration equation could be developed, and the NIR-predicted results for unknown samples could be compared. Four batches of nicotine tablets were used in regression development and validation, and twelve tablets were selected randomly from each batch. In addition, one different batch of nicotine tablets was treated as unknown sample to check the predictability of each regression.

Three sample selection methods, random selection, maximum distance selection, and Mahalanobis distance selection, were studied to reject outliers and to obtain suitable calibration and validation sets. The effects of different sample presentations, different mathematical pretreatments, different wavelength ranges, different segment sizes of the 2^{nd} derivative, and different numbers of factors were also studied.

The differences and effects of transmittance and reflectance NIR spectrophotometers were investigated and results were discussed. It was found that transmittance NIR produced more a robust calibration method than reflectance NIR. Transmittance NIR using PLSR calibration with 2^{nd} derivative or MSC produced results very comparable to those obtained using the HPLC method.

It is safe to say that transmittance NIR spectrophotometry is an excellent technique with precision and accuracy comparable to HPLC when appropriate sample collection, sample selection, and calibration parameters are used.

13. GLOSSARY

CL: Confidence Limit

DQ: Design Qualification

ILS: Inverted Least Squares

IQ: Installation Qualification

K-M Function: Kubelka and Munk Function for Diffuse Reflectance

MIR: Mid-Infrared

MLR: Multiple Linear Regression

MQ: Maintenance Qualification

NIR: Near-Infrared

OQ: Operational Qualification

PC: Principal Component

PCA: Principal Component Analysis

PLS: Partial Least Squares

PLSR: Partial Least Squares Regression

PQ: Performance Qualification

PRESS: Prediction Residual Error – Sum of Squares

R^2: Coefficient of Multiple Determination

SD: Standard Deviation

SEC: Standard Error of Calibration

SEP: Standard Error of Prediction

SLR: Simple Linear Regression

14. APPENDIX

Table 54. Critical Value of the Student t-test (two-tailed)

Confidence Interval	90%	95%	98%	99%
α value	0.10	0.05	0.02	0.01
Degrees of freedom				
1	6.314	12.71	31.82	63.66
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
∞	1.645	1.960	2.326	2.576

Note: The rows correspond to the numbers of degrees of freedom and the columns to values of α. The tables is two tailed, meaning that the probability for the variable to lie outside the interval equals α.[91]

Table 55. Critical Values of F for a One-tailed Test (P = 0.05)

	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.703	8.660
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.803
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.619	4.558
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.874
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511	3.445
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.218	3.150
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006	2.936
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.845	2.774
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.646
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.544
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.459
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.388
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.403	2.328
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.352	2.276
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381	2.308	2.230
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.269	2.191
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.234	2.155
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.203	2.124

Note: ν_1 = number of degrees of freedom of the numerator and ν_2 = number of degrees of freedom of the denominator.[92]

15. REFERENCES

1. K.I. Hildrum, T. Isaksson, T. Naces, and A. Tandberg. *Near Infrared Spectroscopy: Bridging the Gap Between Data Analysis and NIR Applications.* Ellis Horwood, Chichester, 1992.

2. B.G. Osborne, T. Fearn, and P.H. Hindle. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis,* 2nd edition. Longman Scientific and Techniacal, Essex, 1993.

3. F.E. Fowle. *Astrophys. J.,* 1912: **35**; 149-162.

4. P. Kubelka and E. Munk. *Zeits. Tech. Physik,* 1931: **12**; 593.

5. H. Hotelling. *J. Ed. Psych.,* 1933: **24**; 417-441, 489-520.

6. I. Ben-Gera and K.H. Norris. *J. Feed Sci.,* 1968: **64**; 33.

7. *Smoking Control Science.* SmithKline Beecham Consumer Healthcare, Parsippany, 1998.

8. B. Holmstedt. Toxicity of nicotine and related compounds. In: *The pharmacology of nicotine.* M. J. Rand and K. Thurau (Eds.). McLean VA: IRL Press, ICSU Symposium Series 1988: **9**; 61-68

9. International Conference on Harmonization, Q1C (1997) and Q6A (2000).

10. Association of Official Agricultural Chemists. *Official and Provisional Methods of Analysis.* H.W. Wiley (Ed.). US Department of Agriculture, Washington DC, Government Printing Office, 1908.

11. R.H. Cundiff and P.C. Markunas. *Anal. Chem.,* 1955: **27**; 1650-1653.

12. C.O. Willits, M.L. Swaim, J.A. Connelly, and B.A. Brice. *Anal. Chem.,* 1950: **22**; 430-433.

13. L. Leiserson and T.B. Walker. *Anal. Chem.,* 1955: **27**; 1129-1130.

14. X. Xiao. *Fenxi Huaxue*, 1986: **14(8)**; 626.

15. L.D. Quin. *J. Org. Chem.*, 1959: **24**; 911.

16. A.H. Beckett and E.J. Triggs. *Nature*, 1966: **211**; 1415-1417.

17. P. Jacob III, M. Wilson, and N.L. Benowitz. *J. Chromatog.*, 1981: **222**; 434-436.

18. A.H. Beckett, J.W. Gorrod, and P. Jenner. *J. Pharm. Pharmac.*, 1971: **23**; 55S-61S.

19. M. Curvall, E. Kazemi-Vala, and C.R. Enzell. *J. Chromatog.*, 1982: **232**; 283-293.

20. N.L. Benowitz, P. Jacob III, C. Denaro, and R. Jenkins. *Clin. Pharmacol. Ther.*, 1991: **49**; 270-277.

21. P. Jacob III, L. Yu, G. Liang, A.T. Shulgin, and N.L. Benowitz. *J. Chromatogr., Biomed. Applic.*, 1993: **619**; 49-61.

22. A. Pilotti and C.R. Enzell. *Beitr. Tabakforsch.*, 1976: **8**; 339-349.

23. T. Sakaki, H. Sakuma, and S. Sugawara. *Agr. Biol. Chem.*, 1984: **48(11)**; 2719-2724.

24. I.D. Watson. *J. Chromatogr.*, 1977: **143**; 203-206.

25. J.A. Saunders and D.E. Blume. *J. Chromatogr.*, 1981: **205**; 147-154.

26. K.C. Cundy and P.A. Crooks. *J. Chromatogr. Biomed Appl.*, 1984: **306**; 291-301.

27. P. Zuccaro, S. Pichini, I. Altieri, M. Rosa, M. Pellegrini, and R. Pacifici. *Clin. Chem.*, 1997: **43(1)**; 180-181.

28. Nicotine Polacrilex. *The United States Pharmacopeia. USP 24.* Rockville MD: US Pharmacopeia, 2000; 1182.

29. B. Sellergren, A. Zander, T. Renner, and A. Swietlow. *J. Chromatogr. (A)*, 1998: **829**; 143-152.

30. K.T. McManus, J.D. deBethizy, D.A. Garteiz, G.A. Kyerematen, and E.S. Vesell. *J. Chrom. Sci.*, 1990: **28**; 510-516.

31. G.D. Byrd. *44th Annual conference on mass spectrometry and allied topics*, Portland, OR, May 12-16, 1996.

32. P. Crooks and G.D. Byrd. *Analytical determination of nicotine and related compounds and their metabolites*. J.W. Gorrod and P. Jacob III (Eds). Elsevier Science, The Netherlands, 1999; 225-264.

33. S.S. Yang and I. Smetena. *Chromatographia*, 1995: **40**; 375-378.

34. S. Ralapati. *J. Chromatogr.* (B), 1997: **695**; 117-129.

35. G.H. Lu and S. Ralapati. *Electrophoresis*, 1998: **19**; 19-26.

36. J.A. Zoltewicz, L.B. Bloom, and W.R. Kem. *J. Org. Chem.*, 1989: **54**; 4462-4468.

37. D.R. Downs, T.M. Long. *Int. Tobacco Sci. Cong.*, Manila, 1980.

38. W. McClure. *NIR 84 Proc. Int. Symp. Near Infrared Reflectance Spectrosc.*, 1984; 127-133.

39. M. Hana, W. McClure, T.B. Whitaker, M. White, and D.R. Bahler. J. *Near Infrared Spectrosc.*, 1995: **3**; 133-142.

40. Presentation. Foss NIRSystems, Silver Spring, MD.

41. R.S. Drago. *Physical Methods in Chemistry*. W.B. Saunders Co., Philadelphia, 1977.

42. E.U. Cpndon and H. Odishaw. *Handbook of Physics, 2nd edition.* McGraw-Hill Book Company, New York, 1967.

43. J.D. Ingle, Jr. and S.R. Crouch. *Spectrochemical Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

44. I.N. Levine. *Quantum Chemistry, 4th edition.* Prentice Hall, Englewood Cliffs, New Jersey, 1991.

45. J.S. Shenk, J.J. Workman, Jr., M.O. Westerhaus. Application of NIR Spectroscopy to Agricultural Products. In: *Handbook of Near-Infrared Analysis*. D.A. Burns and E.W. Ciurczak (Eds). Marcel Dekker, Inc., New York, 2001; 431-433.

46. *Near-Infrared Absorptions*. Foss NIRSystems, Silver Spring, MD.

47. J.J. Workman, Jr., and D.A. Burns. Commercial NIR Instrumentation. In: *Handbook of Near-Infrared Analysis*. D.A.

Burns and E.W. Ciurczak (Eds). Marcel Dekker, Inc., New York, 2001; 53-70.

48. A. Butler. *Cer. Foods World*, 1983: **28**; 238.

49. E.W. Ciurczak. *Spectroscopy*, 1991: **6(4)**; 12.

50. Instrumentation 1991 Review, *Chem. Eng. News*, March 18, 1991; 20-63.

51. Code of Federal Regulations of the Food and Drug Administration, 21 CFR Parts 11, 210, and 211.

52. V. Grisarti and E.J. Zachowski, *LC/GC North America* , 2002: **20(4)**; 356, 358, 360-362.

53. A. Schuster, *Astrophys. J.*, 1905: **21**; 1.

54. N.T. Melamed, *J. Appl. Phys.*, 1963: **34**; 560.

55. Z. Bodo, *Acta. Phys. Hung.*, 1951: **1**; 135.

56. J. Broser, *Ann. Physik*, 1950: **5**; 401.

57. E.O. Hulbert, *J. Opt. Soc. Am.*, 1943: **33**; 42.

58. H.E.I. Neugebauer, *Z. Tech Physik.*, 1937: **18**; 137.

59. J.H. Lambert. *Photometria sive de mensura et grodibus luminis colorum et umbrae.* Augustae Vindelicorum, 1760.

60. G. Mie, *Ann. Physik*, 1908: **25**; 377.

61. P. Kubelka and F. Munk, *Z. Tech. Physik*, 1931: **12**; 593.

62. J.M. Olinger, P.R. Griffiths, and T. Burger. Theory of Diffuse Reflection in the NIR Region. In: *Handbook of Near-Infrared Analysis*. D.A. Burns and E.W. Ciurczak (Eds). Marcel Dekker, Inc., New York, 2001; 19-51.

63. L. Bush, W.P. Hempfling, and H. Burton. *Analytical Determination of Nicotine and Related Compounds and Their Metabolites.* J.W. Gorrod and P. Jacob III (Eds), Elsevier, 1999; 13.

64. M. Pailer. Tobacco Alkaloids and related compounds, Von Euler (Ed.), Proc. IVth Wenner-Gren International Symposium, New York, 1965; 15-36.

65. P.A. Crooks. *Analytical Determination of Nicotine and Related Compounds and Their Metabolites.* J.W. Gorrod and P. Jacob III (Eds), Elsevier, 1999; 69-147.

66. D.W. Armstrong, X. Wang, J.T. Lee, and Y.S. Liu. *Chirality*, 1999: **11**; 82-84.

67. *The Merck Index*, 12th edition, Merck & Co., Inc., New Jercy, 1996; 1119.

68. *Intact® Analyzers*. Foss NIRSystems, Silver Spring, MD, 2000.

69. *Rapid-Content® Analyzer*. Foss NIRSystems, Silver Spring, MD, 2000.

70. H. Mark. *Handbook of Near-Infrared Analysis*. D.A. Burns and E.W. Ciurczak (Eds). Marcel Dekker, Inc., New York, 2001; 363-400.

71. An excellent discussion can be found on two websites: www.galactic.com and www.statsoft.com.

72. K.G. Joreskog and H. Wold. *Systems Under Indirect Observations. Causality/Structure/Prediction, Vols. I and II.* North-Holland, Amsterdam, 1983.

73. D. Metzler, C.M. Harris, R.L. Reeves, W.H. Lawton, and M.S. Maggio. *Anal. Chem.*, 1977: **49**; 864A.

74. H. Mark. *Anal Chem.*, 1986: **58**; 2814.

75. R. Marbach and H.M. Heise. *TRAC*, 1992: **11**; 270-275.

76. M. Stone and R.J. Brooks. *J. R. Stat. Soc. B.*, 1990: **52**; 337-369.

77. A. Lorber, L.E. Wangen, and B.R. Kowalski. *J. Chemom.*, 1987: **1**; 19-31.

78. S. Wold, P. Geladi, K. Esbensen, and J. Ohman, *J. Chemom.*, 1987: **1**; 41.

79. D. Haaland and E. Thomas, *Anal. Chem.*, 1988: **60**; 1193.

80. A. Lorber, L.E. Wangen, and B.R. Kowalski, *J. Chemom.*, 1987: **1**; 19.

81. H. Martens and T. Nas. *Multivariate Calibration.* John Wiley and Sons, New York, 1989.

82. K. Pearson. *Phil. Mag. (Ser. 6).* 1901: **2**; 559-572.

83. *Standard Definitions of Terms and Symbols Relating to Molecular Spectroscopy.* ASTM vol. 14.01 Standard E131-90.

84. K.R. Beebe, R.J. Pell, and M.B. Seasholtz. *Chememetrics.* John Wiley & Sons, Inc., New York, 1998.

85. *Reference and Theory Manual.* Foss NIRSystems, Silver Spring, MD.

86. D.M. Haaland and E.V. Thomas, *Anal. Chem.* 1988: **60;** 1193-1202.

87. N.R. Draper and A. Smith, *Applied Regression Analysis.* John Wiley and Sons, New York, 1981.

88. D.A. Burns and E.W. Ciurczak. *Handbook of Near-Infrared Analysis, 2nd edition.* Marcel Dekker, New York, 2001.

89. J.C. Miller and J.N. Miller. *Statistics for Analytical Chemistry, 2nd edition.* John Wiley & Sons, New York, 1988.

90. T. Isaksson and T. Naes. *Applied Spectroscopy,* 1988: **42**; 1273-1284.

91. [91] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y Michotte, and L. Kaufman. *Chemometrics: a textbook.* Elsevier, 1988; 470.

92. [92] J.C. Miller and J.N. Miller. Statistics for Analytical Chemistry, 3rd edition. Ellis Horwood Limited, 1993; 223.