

Principal Component Analysis for Enhancement of Infrared Spectra Monitoring

by

Ricky Lance Haney

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2011

Copyright 2011 by Ricky Lance Haney

Approved by

Jeffrey Fergus, Chair, Professor of Materials Engineering
Ruel (Tony) Overfelt, Professor of Mechanical Engineering
Bart Prorok, Associate Professor of Materials Engineering
Curtis Shannon, Professor of Chemistry and Biochemistry

Abstract

The issue of air quality within the aircraft cabin is receiving increasing attention from both pilot and flight attendant unions. This is due to exposure events caused by poor air quality that in some cases may have contained toxic oil components due to bleed air that flows from outside the aircraft and then through the engines into the aircraft cabin. Significant short and long-term medical issues for aircraft crew have been attributed to exposure. The need for air quality monitoring is especially evident in the fact that currently within an aircraft there are no sensors to monitor the air quality and potentially harmful gas levels (detect-to-warn sensors), much less systems to monitor and purify the air (detect-to-treat sensors) within the aircraft cabin.

The specific purpose of this research is to utilize a mathematical technique called principal component analysis (PCA) in conjunction with principal component regression (PCR) and proportionality constant calculations (PCC) to simplify complex, multi-component infrared (IR) spectra data sets into a reduced data set used for determination of the concentrations of the individual components. Use of PCA can significantly simplify data analysis as well as improve the ability to determine concentrations of individual target species in gas mixtures where significant band overlap occurs in the IR spectrum region. Application of this analytical numerical technique to IR spectrum analysis is important in improving performance of commercial sensors that airlines and aircraft manufacturers could potentially use in an aircraft cabin environment for multi-gas component monitoring.

The approach of this research is two-fold, consisting of a PCA application to compare simulation and experimental results with the corresponding PCR and PCC to determine quantitatively the component concentrations within a mixture. The experimental data sets consist of both two and three component systems that could potentially be present as air contaminants in an aircraft cabin. In addition, experimental data sets are analyzed for a hydrogen peroxide (H_2O_2) aqueous solution mixture to determine H_2O_2 concentrations at various levels that could be produced during use of a vapor phase hydrogen peroxide (VPHP) decontamination system. After the PCA application to two and three component systems, the analysis technique is further expanded to include the monitoring of potential bleed air contaminants from engine oil combustion. Simulation data sets created from database spectra were utilized to predict gas components and concentrations in unknown engine oil samples at high temperatures as well as time-evolved gases from the heating of engine oils.

Acknowledgments

I would like to express great appreciation towards Dr. Jeffrey Fergus, who was instrumental in my decision to attend Auburn University and pursue a Ph.D. in materials engineering. Dr. Fergus has also provided invaluable insight towards this research effort as well as significant funding support, along with Dr. Ruel (Tony) Overfelt, throughout my time at Auburn. I would like to express sincere appreciation to Dr. Overfelt for allowing me to work within the Air Transportation Center of Excellence (ACER), which has provided access to all the necessary resources for this research effort. I would like to thank Dr. Curtis Shannon, who guided the research effort towards the use of principal component analysis at the proposal stage. Dr. Shannon has also been vital in furthering my understanding of infrared spectroscopy. Additionally, I would like to thank Dr. Bart Prorok for serving on my Ph.D. committee and providing insight to further this research effort during weekly group meetings.

I would like to acknowledge fellow students that have helped tremendously with this research effort. Mobbassar Hassan Sk was instrumental in the data collection and analysis of the hydrogen peroxide in aqueous solution samples. John Andress helped significantly with using the FTIR instrument as well as the collection and analysis of numerous gas samples for this research effort as well as for commercial sensor testing. Amanda Neer's work with the data collection of the engine oil samples was critical in completion of this dissertation.

Most of all, I must thank my wonderful wife, Lacey, without whom the completion of this dissertation would not have been possible. She was always there to help me and always believed in my abilities to finish this research effort.

This project was funded by the U.S. Federal Aviation Administration (FAA) Office of Aerospace Medicine through the National Air Transportation Center of Excellence for Research in the Intermodal Transport Environment (RITE), Cooperative Agreement 07-C-RITE-AU.

Although the FAA has sponsored this project, it neither endorses nor rejects the findings of this research.

Table of Contents

Abstract.....	ii
Acknowledgments.....	iv
List of Tables	ix
List of Figures.....	xii
List of Abbreviations and Mathematical Symbols.....	xxiv
Chapter 1: Introduction.....	1
Research Purpose	3
Research Approach	4
Organization of Dissertation	5
Chapter 2: Commercial Aircraft Systems Overview	7
Commercial Aircraft Background	7
Vapor Phase Hydrogen Peroxide (VPHP) Aircraft Cabin Decontamination	11
Potential Environmental Air Contaminants within the Aircraft Cabin	12
Potential Bleed Air Contaminants within the Aircraft Cabin	13
Laboratory Test Environment	13
Chapter 3: Fourier Transform Infrared (FTIR) Spectroscopy Overview.....	23
FTIR Theoretical Background	23
FTIR Measurement Technique	24
IR Characteristics of CO/CO ₂ /H ₂ O	28

Carbon Monoxide (CO)	28
Carbon Dioxide (CO ₂)	29
Water (H ₂ O)	31
Chapter 4: Principal Component Analysis (PCA)	33
PCA Theoretical Background	33
Principal Component Regression (PCR)	36
Proportionality Constant Calculation (PCC)	38
Application to FTIR Spectroscopy Data	39
Chapter 5: PCA Application to FTIR Spectroscopy Data of Vapor Phase Hydrogen Peroxide (VPHP) Aircraft Cabin Decontamination Events	53
Discussion of Analyzed Data Sets	53
PCA Results and Discussion	55
Data Set 1	55
Data Set 2	67
Chapter 6: PCA Application to FTIR Spectroscopy Data of Potential Environment Air Contaminants within the Aircraft Cabin	71
Discussion of Analyzed Data Sets	71
PCA Results and Discussion: 2-Component Systems	72
CH ₂ O/C ₃ H ₄ O Simulation Data Set	72
CO/CO ₂ Simulation Data Set	80
CO/CO ₂ Experimental Data Set	87
PCA Results and Discussion: 3-Component Systems	98
CH ₂ O/C ₃ H ₄ O/H ₂ O Simulation Data Set	98
CO/CO ₂ /H ₂ O Experimental Data Set	109

Chapter 7: PCA Application to FTIR Spectroscopy Data of Potential Bleed Air Contaminants within the Aircraft Cabin	119
Discussion of Analyzed Data Sets	119
PCA Results and Discussion	121
Data Set 1 – Engine Oil Samples at Temperatures of Greatest Mass Loss	121
Data Set 2 – Simulated CH ₄ O/CH ₂ O/CO ₂ Gas Mixtures & Engine Oil Samples at Temperatures of Greatest Mass Loss	124
Data Set 3 – Simulated CH ₄ O/CH ₂ O/CO ₂ /CO/H ₂ O Gas Mixtures & BP Turbo Oil 2380 Engine Oil Time-Evolved Samples	143
Chapter 8: Conclusions	155
Chapter 9: Future Work	156
References	158
Appendix A: Principal Component Analysis MATLAB® Source Code	163

List of Tables

Table 2-1: Relationship between Altitude and Pressure with the Typical Aircraft Cabin Pressure Highlighted	8
Table 2-2: Limits on Contaminants that Could Potentially Be Found in Aircraft Cabin Air	13
Table 3-1: Relationship between I/I_0 , $\%T$, and A	26
Table 4-1: CH ₂ O/C ₃ H ₄ O Simplified Pure Spectra for Illustration of PCA Application to FTIR Spectroscopy Data	40
Table 4-2: CH ₂ O/C ₃ H ₄ O Spectra Compositions for Calibration Data Set, [X _C] _(10 x 7) , Ordered from Lowest to Highest Amount of CH ₂ O in the Gas Mixture	41
Table 4-3: CH ₂ O/C ₃ H ₄ O Spectra Compositions for Prediction Data Set, [X _P] _(5 x 7) , Ordered from Lowest to Highest Amount of CH ₂ O in the Gas Mixture	42
Table 4-4: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques	52
Table 5-1: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; H ₂ O ₂ in Aqueous Solution Data Set 1	67
Table 5-2: H ₂ O ₂ in Aqueous Solution Spectra Compositions for Prediction Data Set 2, [X _P] _(5 x 601)	67
Table 5-3: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; H ₂ O ₂ in Aqueous Solution Data Set 2	70
Table 6-1: CH ₂ O/C ₃ H ₄ O Spectra Compositions for Calibration Data Set, [X _C] _(8 x 2655) , Ordered from Low to High Concentration of CH ₂ O	73
Table 6-2: CH ₂ O/C ₃ H ₄ O Spectra Compositions for Prediction Data Set, [X _C] _(8 x 2655) , Ordered from Low to High Concentration of CH ₂ O	73

Table 6-3: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; CH ₂ O/C ₃ H ₄ O Gas Mixtures	80
Table 6-4: CO/CO ₂ Spectra Compositions for Calibration Data Set, [X _C] _(10 x 2636)	80
Table 6-5: CO/CO ₂ Spectra Compositions for Prediction Data Set, [X _P] _(5 x 2636)	81
Table 6-6: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; CO/CO ₂ Gas Mixtures	87
Table 6-7: CO/CO ₂ Spectra Compositions for Calibration Data Set, [X _C] _(8 x 5001)	87
Table 6-8: CO/CO ₂ Spectra Compositions for Prediction Data Set, [X _P] _(4 x 5001)	88
Table 6-9: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; CO/CO ₂ Experimental Gas Mixtures	98
Table 6-10: CH ₂ O/C ₃ H ₄ O/H ₂ O Spectra Compositions for Calibration Data Set, [X _C] _(8 x 2655) , Ordered from Low to High Concentration of CH ₂ O	99
Table 6-11: CH ₂ O/C ₃ H ₄ O/H ₂ O Spectra Compositions for Prediction Data Set, [X _P] _(3 x 2655) , Ordered from Low to High Concentration of CH ₂ O	100
Table 6-12: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; CH ₂ O/C ₃ H ₄ O Gas Mixtures	109
Table 6-13: CO/CO ₂ /H ₂ O Spectra Compositions for Calibration Data Set, [X _C] _(8 x 1001)	109
Table 6-14: CO/CO ₂ /H ₂ O Spectra Compositions for Prediction Data Set, [X _P] _(8 x 1001)	110
Table 6-15: Errors Associated with the Principal Component Regression (PCR) Analysis Technique; CO/CO ₂ /H ₂ O Experimental Gas Mixtures	116
Table 7-1: Calibration Data Set 2 Composed of Simulated Spectra of Various Concentrations of Methanol (CH ₄ O), Formaldehyde (CH ₂ O), and Carbon Dioxide (CO ₂)	125
Table 7-2: PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560	128
Table 7-3: RMSE for Predicted Spectra based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560	131

Table 7-4: PCR Calculated Concentrations of CH ₄ O (Modified), CH ₂ O, and CO ₂ for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560	135
Table 7-5: Comparison of RMSE using the Standard CH ₄ O Spectra and the Modified CH ₄ O Spectra for Predicted Spectra based on PCR Calculated Concentrations	138
Table 7-6: Truncated (3200 – 1600 cm ⁻¹) PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560	139
Table 7-7: Comparison of RMSE using the Standard CH ₄ O Spectra, the Modified CH ₄ O Spectra, and the Truncated Spectra for Predicted Spectra based on PCR Calculated Concentrations	141
Table 7-8: PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ for BP Turbo Oil 2380.....	141
Table 7-9: PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ for BP Turbo Oil 274	142
Table 7-10: PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ for Mobile Jet Oil II.....	142
Table 7-11: PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ for Aeroshell Turbine Oil 560	142
Table 7-12: Calibration Data Set 2 Composed of Simulated Spectra of Various Concentrations of Methanol (CH ₄ O), Formaldehyde (CH ₂ O), Carbon Dioxide (CO ₂), Carbon Monoxide (CO), and Water (H ₂ O)	144
Table 7-13: Comparison of RMSE using the Full and the Truncated Spectra, for the PCR Calculated Concentrations	153

List of Figures

Figure 1-1: Commercial Aircraft Bleed Air System	2
Figure 2-1: Characteristics of Ideal CO ₂ Partial Pressure Sensor; Measures $p_{\text{CO}_2} = 0.39$ at $P_T = 77$ kPa (7000 ft Altitude) for $p_{\text{CO}_2} = 0.51$ at $P_T = 101$ kPa (Sea Level)	9
Figure 2-2: Recirculation System Outside Air Changes per Hour for Aircraft Compared with Other Environments	10
Figure 2-3: Cabin Ventilation System Illustrating Typical Aircraft Recirculated Airflow	11
Figure 2-4: Laboratory Test Environment, Control Module	14
Figure 2-5: Laboratory Test Environment, Commercial Sensor Analysis Module	15
Figure 2-6: Laboratory Test Environment, FTIR Gas Analysis Module with Spectrum GX FTIR (Perkin Elmer, Shelton, CT, USA) with M-5-22-V Variable Pathlength Long Path Gas Cell (Infrared Analysis, Inc., Anaheim, CA, USA)	16
Figure 2-7: Internal Gold Plated Mirrors of the Variable Pathlength Long Path Gas Cell Highlighting the Path of the IR Beam within the Cell	17
Figure 2-8: System Block Diagram Illustrating Gas Flows through the Commercial Sensor Analysis Module to the FTIR Gas Analysis Module then Out through the Vacuum Pump	18
Figure 2-9: Theoretical Gas Flow Model Calculations Compared to Experimental Results for 800 ppm CO Test Gas in N ₂ with Total Inflow of 500 sccm with Vacuum Pressure (P_{vac}) = Atmospheric Pressure (Pa)	21
Figure 2-10: Theoretical Gas Flow Model Calculations for FTIR Gas Analysis Module in Series with the Commercial Gas Sensor Analysis Module	22
Figure 3-1: Energy Diagram Highlighting Transitions from Electronic Ground State to Electronic Excited State with Rotational and Vibrational Transitions	23
Figure 3-2: Beam Path from IR Source Highlighting the use of a Fixed Mirror and Moving Mirror to Construct an Interferogram Time Domain Signal Containing the IR Source Information	25

Figure 3-3: Typical Atmospheric Background Spectra Highlighting CO ₂ and H ₂ O Interference, 16 Co-added Scans with Resolution 0.5 cm ⁻¹	27
Figure 3-4: CO Fundamental Vibrational Mode, k ₁ = 2143 cm ⁻¹	28
Figure 3-5: IR Absorbance Spectra for CO from QASoft® Database	28
Figure 3-6: CO ₂ First Fundamental Vibrational Mode, k ₁ = 1340 cm ⁻¹	29
Figure 3-7: CO ₂ Second Fundamental Vibrational Frequency, k ₂ = 667 cm ⁻¹	29
Figure 3-8: CO ₂ Third Fundamental Vibrational Frequency, k ₃ = 2350 cm ⁻¹	30
Figure 3-9: IR Absorbance Spectra for CO ₂ from QASoft® Database	31
Figure 3-10: H ₂ O Second Fundamental Vibrational Frequency, k ₂ = 1595 cm ⁻¹	32
Figure 3-11: IR Absorbance Spectra for H ₂ O from QASoft® Database	32
Figure 4-1: Variance-Covariance Matrix, [Z] _(p x p) Calculated from Mean Centered Data Matrix, [X _M] _(n x p)	34
Figure 4-2: Formaldehyde (CH ₂ O) and Acrolein (C ₃ H ₄ O) Complete Pure Spectra	40
Figure 4-3: CH ₂ O/C ₃ H ₄ O Simplified Pure Spectra for Illustration of PCA Application to FTIR Spectroscopy Data	40
Figure 4-4: CH ₂ O/C ₃ H ₄ O Calibration Data Set, [X _C] _(10 x 7) ; Note – Only 5 of 10 Calibration Spectra Shown	42
Figure 4-5: CH ₂ O/C ₃ H ₄ O Prediction Data Set, [X _P] _(5 x 7)	43
Figure 4-6: CH ₂ O/C ₃ H ₄ O Calibration Data Set, Mean Centered, [X _M] _(10 x 7) ; Note – Only 5 of 10 Calibration Spectra Shown	43
Figure 4-7: CH ₂ O/C ₃ H ₄ O Variance-Covariance Matrix, [Z] _(7 x 7)	44
Figure 4-8: CH ₂ O/C ₃ H ₄ O Eigenvalues from Solving Equation 4.2	44
Figure 4-9: SCREE Plot Indicating Eigenvalues for Each Principal Component; Principal Components 1 and 2 Explain 77.1% and 22.9% of the Total Calibration Data Set Variance	45
Figure 4-10: CH ₂ O/C ₃ H ₄ O Loadings Matrix, [V] _(7 x 7)	45
Figure 4-11: CH ₂ O/C ₃ H ₄ O Scores Matrix, [S] _(10 x 10)	46

Figure 4-12: CH ₂ O/C ₃ H ₄ O Estimates of Regression Coefficients from Calibration Data Set, [b] _(7x1)	46
Figure 4-13: Principal Component Regression (PCR) for CH ₂ O Concentrations in CH ₂ O/C ₃ H ₄ O Gas Mixtures; RMSE Calibration = 1.1x10 ⁻⁴ ppm, RMSE Prediction = 1.1x10 ⁻⁴ ppm	47
Figure 4-14: PCA Separated CH ₂ O Spectra in CH ₂ O/C ₃ H ₄ O Gas Mixtures	48
Figure 4-15: PCA Separated C ₃ H ₄ O Spectra in CH ₂ O/C ₃ H ₄ O Gas Mixtures	48
Figure 4-16: PCC C ₃ H ₄ O Calibration Data Set; Note – Only 5 of 10 Calibration Spectra Shown	49
Figure 4-17: PCC C ₃ H ₄ O Prediction Data Set	49
Figure 4-18: PCC CH ₂ O Calibration Data Set; Note – Only 5 of 10 Calibration Spectra Shown	50
Figure 4-19: PCC CH ₂ O Prediction Data Set	50
Figure 4-20: PCC CH ₂ O Calibration Data Set with No Baseline Correction	51
Figure 4-21: PCC Calibration Data Set with Baseline Correction	51
Figure 4-22: PCC CH ₂ O Prediction Data Set; RMSE Calibration = 2.6x10 ⁻² ppm, RMSE Prediction = 2.6x10 ⁻² ppm.....	52
Figure 5-1: H ₂ O ₂ in Aqueous Solution Calibration Data Set 1, [X _C] _(8 x 801) ; Note – Only 4 of 8 Calibration Spectra Shown, 0%, 10%, 20%, 30% H ₂ O ₂	56
Figure 5-2: H ₂ O ₂ in Aqueous Solution Prediction Data Set 1, [X _P] _(1 x 801) ; 63.7% H ₂ O ₂	56
Figure 5-3: SCREE Plot Indicating Eigenvalues for Each Principal Component for H ₂ O ₂ in Aqueous Solution Data Set 1; Principal Components 1 and 2 Explain 67.5% and 28.5% of the Total Calibration Data Set Variance	57
Figure 5-4: H ₂ O ₂ in Aqueous Solution Mixtures – Reduced Principal Component Loadings, [V*]; V-1 represents the variable in the original data set contributing the most variance within the spectra, the H ₂ O component, V-2 represents the variable in the original data set contributing the second most variance, the H ₂ O ₂ component	58
Figure 5-5: H ₂ O ₂ in Aqueous Solution Mixtures – Estimates of Regression Coefficients for wt. % of H ₂ O ₂ from Calibration Data Set 1 as Function of Wavenumber, [b _{H2O2}] _(801 x 1)	59

Figure 5-6: Principal Component Regression (PCR) – H ₂ O ₂ Concentrations for H ₂ O ₂ in Aqueous Solution Mixtures; RMSE Calibration = 2.1 wt.%, RMSE Prediction = 12.0 wt.%	59
Figure 5-7: H ₂ O ₂ RMSE Calibration and RMSE Prediction as a Function of the Number of Principal Components Used to Represent the Original Data Set for H ₂ O ₂ in Aqueous Solution Mixtures Data Set 1	60
Figure 5-8: PCC H ₂ O ₂ Calibration Data Set 1; Note – Only 3 of 8 Calibration Spectra Shown, 0%, 20%, 30% H ₂ O ₂	61
Figure 5-9: PCC H ₂ O ₂ Prediction Data Set 1; 63.7% H ₂ O ₂	62
Figure 5-10: PCC H ₂ O Calibration Data Set 1; Note – Only 2 of 8 Calibration Spectra Shown, 70%, 100% H ₂ O	62
Figure 5-11: PCC H ₂ O Prediction Data Set 1; 36.3% H ₂ O	63
Figure 5-12: PCC H ₂ O ₂ Calibration Data Set 1 with No Baseline Correction	64
Figure 5-13: PCC H ₂ O Calibration Data Set 1 with No Baseline Correction	64
Figure 5-14: PCC H ₂ O ₂ Calibration Data Set 1 with Baseline Correction	65
Figure 5-15: PCC H ₂ O Calibration Data Set 1 with Baseline Correction	66
Figure 5-16: PCC Model for wt.% of H ₂ O ₂ in Aqueous Solution with Variable Total Amount in Data Set 1; k _{H₂O₂} = 0.19, k _{H₂O} = 0.67; RMSE Calibration = 4.1 wt.%, RMSE Prediction = 5.2 wt.%	66
Figure 5-17: SCREE Plot Indicating Eigenvalues for Each Principal Component for H ₂ O ₂ in Aqueous Solution Data Set 2; Principal Components 1 and 2 Explain 61% and 34% of the Total Calibration Data Set Variance	68
Figure 5-18: Principal Component Regression (PCR) – H ₂ O ₂ Concentrations for H ₂ O ₂ in Aqueous Solution with Variable Total Amount in Data Set 2; RMSE Calibration = 10.7 wt.%, RMSE Prediction = 3.9 wt.%	68
Figure 5-19: PCC Model for wt.% of H ₂ O ₂ in Aqueous Solution with Variable Total Amount in Data Set 2; k _{H₂O₂} = 0.65, k _{H₂O} = 0.40; RMSE Calibration = 5.9 wt.%, RMSE Prediction = 6.0 wt.%	69
Figure 6-1: CH ₂ O/C ₃ H ₄ O Gas Mixtures Calibration Data Set, [X _C] _(8 x 2655) ; Note – Only 3 of 8 Calibration Spectra Shown	73
Figure 6-2: CH ₂ O/C ₃ H ₄ O Gas Mixtures Prediction Data Set, [X _P] _(3 x 2655)	74

Figure 6-3: SCREE Plot Indicating Eigenvalues for Each Principal Component for CH ₂ O/C ₃ H ₄ O Gas Mixtures Data Set; Principal Components 1 and 2 Explain 98.4% and 1.6% of the Total Calibration Data Set Variance	74
Figure 6-4: Principal Component Regression (PCR) – CH ₂ O Concentrations in CH ₂ O/C ₃ H ₄ O Gas Mixtures; RMSE Calibration = 0.0 ppm, RMSE Prediction = 0 ppm	75
Figure 6-5: Principal Component Regression (PCR) – C ₃ H ₄ O Concentrations in CH ₂ O/C ₃ H ₄ O Gas Mixtures; RMSE Calibration = 0.0 ppm, RMSE Prediction = 0 ppm	76
Figure 6-6: PCA Separated CH ₂ O Spectra in CH ₂ O/C ₃ H ₄ O Gas Mixtures	77
Figure 6-7: PCA Separated C ₃ H ₄ O Spectra in CH ₂ O/C ₃ H ₄ O Gas Mixtures	77
Figure 6-8: PCC CH ₂ O Calibration and Prediction Data Sets in CH ₂ O/C ₃ H ₄ O Gas Mixtures; k _{CH₂O} = 0.08; RMSE Calibration = 1 ppm, RMSE Prediction = 2 ppm	78
Figure 6-9: PCC C ₃ H ₄ O Calibration and Prediction Data Sets in CH ₂ O/C ₃ H ₄ O Gas Mixtures; k _{C₃H₄O} = 0.32; RMSE Calibration = 0 ppm, RMSE Prediction = 2 ppm	79
Figure 6-10: CO/CO ₂ Gas Mixtures Calibration Data Set, [X _C] _(10 x 2636) ; Note – Only 4 of 10 Calibration Spectra Shown	81
Figure 6-11: CO/CO ₂ Gas Mixtures Prediction Data Set, [X _P] _(5 x 2636)	82
Figure 6-12: SCREE Plot Indicating Eigenvalues for Each Principal Component for CO/CO ₂ Gas Mixtures Data Set; Principal Components 1 and 2 Explain 99.1% and 0.9% of the Total Calibration Data Set Variance	82
Figure 6-13: Principal Component Regression (PCR) – CO Concentrations in CO/CO ₂ Gas Mixtures; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm	83
Figure 6-14: Principal Component Regression (PCR) – CO ₂ Concentrations in CO/CO ₂ Gas Mixtures; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm	84
Figure 6-15: PCA Separated CO Spectra in CO/CO ₂ Gas Mixtures	85
Figure 6-16: PCA Separated CO ₂ Spectra in CO/CO ₂ Gas Mixtures	85
Figure 6-17: PCC CO Calibration and Prediction Data Sets in CO/CO ₂ Gas Mixtures; k _{CO} = 0.75; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm	86
Figure 6-18: PCC CO ₂ Calibration and Prediction Data Sets in CO/CO ₂ Gas Mixtures; k _{CO₂} = 0.06; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm	86

Figure 6-19: CO/CO ₂ Gas Mixtures Calibration Data Set, [X _C] _(8 x 5001) ; Note – Only 4 of 8 Calibration Spectra Shown	88
Figure 6-20: CO/CO ₂ Gas Mixtures Prediction Data Set, [X _P] _(4 x 5001)	88
Figure 6-21: SCREE Plot Indicating Eigenvalues for Each Principal Component for CO/CO ₂ Gas Mixtures Data Set; Principal Components 1 and 2 Explain 97.9% and 0.9% of the Total Calibration Data Set Variance	89
Figure 6-22: Principal Component Regression (PCR) – CO Concentrations in CO/CO ₂ Gas Mixtures; RMSE Calibration = 92 ppm, RMSE Prediction = 49 ppm	90
Figure 6-23: Principal Component Regression (PCR) – CO ₂ Concentrations in CO/CO ₂ Gas Mixtures; RMSE Calibration = 0.7 ppm, RMSE Prediction = 0.8 ppm	90
Figure 6-24: CO RMSE Calibration and RMSE Prediction as a Function of the Number of Principal Components Used to Represent the Original Data Set in CO/CO ₂ Gas Mixtures	91
Figure 6-25: Principal Component Regression (PCR) with 3 Principal Components – CO Concentrations in CO/CO ₂ Gas Mixtures; RMSE Calibration = 16 ppm, RMSE Prediction = 14 ppm	92
Figure 6-26: Principal Component Regression (PCR) with 4 Principal Components – CO Concentrations in CO/CO ₂ Gas Mixtures; RMSE Calibration = 8 ppm, RMSE Prediction = 14 ppm	92
Figure 6-27: CO/CO ₂ Gas Mixtures – Estimates of Regression Coefficients for Concentrations of CO from Calibration Data Set as Function of Wavenumber, [b _{CO}] _(5000 x 1) ; 2 Principal Components Used	93
Figure 6-28: CO/CO ₂ Gas Mixtures – Estimates of Regression Coefficients for Concentrations of CO from Calibration Data Set as Function of Wavenumber, [b _{CO}] _(5000 x 1) ; 3 Principal Components Used	94
Figure 6-29: CO/CO ₂ Gas Mixtures – Estimates of Regression Coefficients for Concentrations of CO from Calibration Data Set as Function of Wavenumber, [b _{CO}] _(5000 x 1) ; 4 Principal Components Used	94
Figure 6-30: PCA Separated CO Spectra in CO/CO ₂ Gas Mixtures Represented by V-2	95
Figure 6-31: PCA Separated CO ₂ Spectra in CO/CO ₂ Gas Mixtures Represented by V-1	96
Figure 6-32: PCC CO Calibration and Prediction Data Sets in CO/CO ₂ Gas Mixtures; k _{CO} = 20; RMSE Calibration = 204 ppm, RMSE Prediction = 194 ppm	97

Figure 6-33: PCC CO ₂ Calibration and Prediction Data Sets in CO/CO ₂ Gas Mixtures; k _{CO₂} = 0.25; RMSE Calibration = 1 ppm, RMSE Prediction = 1 ppm	97
Figure 6-34: Formaldehyde (CH ₂ O), Acrolein (C ₃ H ₄ O), and Water (H ₂ O) Pure Spectra for Simulated Data Sets Illustrating Spectral Overlap Between All Three Components	99
Figure 6-35: CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures Calibration Data Set, [X _C] _(8 x 2655) ; Note – Only 3 of 8 Calibration Spectra Shown	100
Figure 6-36: CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures Prediction Data Set, [X _P] _(3 x 2655)	101
Figure 6-37: SCREE Plot Indicating Eigenvalues for Each Principal Component for CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures Data Set; Principal Components 1, 2, and 3 Explain 71.9%, 26.9%, and 1.2%, respectively, of the Total Calibration Data Set Variance	102
Figure 6-38: Principal Component Regression (PCR) – CH ₂ O Concentrations in CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm	103
Figure 6-39: Principal Component Regression (PCR) – H ₂ O Concentrations in CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm	103
Figure 6-40: Principal Component Regression (PCR) – C ₃ H ₄ O Concentrations in CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm.....	104
Figure 6-41: PCA Separated CH ₂ O Spectra in CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures	105
Figure 6-42: PCA Separated H ₂ O Spectra in CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures	105
Figure 6-43: PCA Separated C ₃ H ₄ O Spectra in CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures	106
Figure 6-44: PCC CH ₂ O Calibration and Prediction Data Sets in CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures; k _{CH₂O} = 0.08; RMSE Calibration = 1 ppm, RMSE Prediction = 2 ppm	107
Figure 6-45: PCC H ₂ O Calibration and Prediction Data Sets in CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures; k _{H₂O} = 1.25; RMSE Calibration = 9 ppm, RMSE Prediction = 6 ppm	107
Figure 6-46: PCC C ₃ H ₄ O Calibration and Prediction Data Sets in CH ₂ O/C ₃ H ₄ O/H ₂ O Gas Mixtures; k _{C₃H₄O} = 0.39; RMSE Calibration = 1 ppm, RMSE Prediction = 1 ppm	108
Figure 6-47: CO/CO ₂ /H ₂ O Gas Mixtures Calibration Data Set, [X _C] _(8 x 1001) ; Note – Only 2 of 8 Calibration Spectra Shown	110

Figure 6-48: CO/CO ₂ /H ₂ O Gas Mixtures Prediction Data Set, [X _P] _(4 x 1001) ; Note – Only 2 of 4 Prediction Spectra Shown	111
Figure 6-49: SCREE Plot Indicating Eigenvalues for Each Principal Component for CO/CO ₂ /H ₂ O Gas Mixtures Data Set; Principal Components 1, 2, and 3 Explain 69.7%, 12.1%, and 7.7% of the Total Calibration Data Set Variance	112
Figure 6-50: Principal Component Regression (PCR) – CO Concentrations in CO/CO ₂ /H ₂ O Gas Mixtures; RMSE Calibration = 14 ppm, RMSE Prediction = 18 ppm	113
Figure 6-51: Principal Component Regression (PCR) – CO ₂ Concentrations in CO/CO ₂ /H ₂ O Gas Mixtures; RMSE Calibration = 2 ppm, RMSE Prediction = 4 ppm	113
Figure 6-52: Principal Component Regression (PCR) – H ₂ O Concentrations in CO/CO ₂ /H ₂ O Gas Mixtures; RMSE Calibration = 157 ppm, RMSE Prediction = 519 ppm	114
Figure 6-53: CO RMSE Calibration and RMSE Prediction as a Function of the Number of Principal Components Used to Represent the Original Data Set in CO/CO ₂ /H ₂ O Gas Mixtures	115
Figure 6-54: SCREE Plot Indicating Eigenvalues for Each Principal Component for CO/CO ₂ /H ₂ O Gas Mixtures Data Set; Principal Components 1, 2, and 3 Explain 73.4%, 12.6%, and 6.2% of the Total Calibration Data Set Variance	117
Figure 6-55: CO RMSE Calibration and RMSE Prediction as a Function of the Number of Principal Components Used to Represent the Original Data Set in CO/CO ₂ /H ₂ O Gas Mixtures with the IR Spectral Data Truncated to Include Only Contributions from CO and CO ₂ (2500 cm ⁻¹ to 2075 cm ⁻¹) for PCA	118
Figure 7-1: FTIR Scans of Engine Oil Samples at Temperatures of Greatest Mass Loss used for PCA; BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 at 306 °C (583 °F), 301 °C (574 °F), 306 °C, and 326 °C (619 °F), respectively	121
Figure 7-2: SCREE Plot Indicating Eigenvalues for Each Principal Component for Engine Oil Data Set 1; Principal Components 1, 2, and 3 Explain 76.1%, 21.6%, and 2.3% of the Total Engine Oil Data Set 1 Variance	122
Figure 7-3: Mass Spectrometry (MS) Data of Engine Oil Samples at Temperatures of Greatest Mass Loss; BP Turbo Oil 274, and Mobile Jet Oil II at 301 °C (574 °F), and 306 °C (578 °F), respectively	123

Figure 7-4: Mass Spectrometry (MS) Data Database Files for Formaldehyde (CH ₂ O), Methanol (CH ₄ O), and Carbon Dioxide (CO ₂)	123
Figure 7-5: Formaldehyde (CH ₂ O), Methanol (CH ₄ O), and Carbon Dioxide (CO ₂) Pure Spectra for Simulated Data Set Illustrating Spectral Overlap Between CH ₂ O and CH ₄ O	124
Figure 7-6: CH ₄ O/CH ₂ O/CO ₂ Gas Mixtures Calibration Data Set, [X _C] _(10 x 1763) ; Note – Only 3 of 10 Calibration Spectra Shown	125
Figure 7-7: SCREE Plot Indicating Eigenvalues for Each Principal Component for CH ₄ O/CH ₂ O/CO ₂ Gas Mixtures Data Set; Principal Components 1, 2, and 3 Explain 86.8%, 95.4%, and 4.6%, respectively, of the Total Calibration Data Set Variance	126
Figure 7-8: Principal Component Regression (PCR) – CH ₄ O Concentrations in CH ₄ O/CH ₂ O/CO ₂ Gas Mixtures; RMSE Calibration = 0.0 ppm; Predicted concentrations for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 206, 221, 214, and 292 ppm, respectively	127
Figure 7-9: Principal Component Regression (PCR) – CH ₂ O Concentrations in CH ₄ O/CH ₂ O/CO ₂ Gas Mixtures; RMSE Calibration = 0.0 ppm; Predicted concentrations for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 145, 190, 142, and 263 ppm, respectively	127
Figure 7-10: Principal Component Regression (PCR) – CO ₂ Concentrations in CH ₄ O/CH ₂ O/CO ₂ Gas Mixtures; RMSE Calibration = 0.0 ppm; Predicted concentrations for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 30, 49, 20, and 63 ppm, respectively	128
Figure 7-11: Predicted Spectra for BP Turbo Oil 2380 based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 2.4%	129
Figure 7-12: Predicted Spectra for BP Turbo Oil 274 based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 3.1%	129
Figure 7-13: Predicted Spectra for Mobile Jet Oil II based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 2.7%	130
Figure 7-14: Predicted Spectra for Aeroshell Turbine Oil 560 based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 3.9%	130

Figure 7-15: Illustration of Mutual Interactions within the Hydroxyl Group via Hydrogen Bonded Methanol (CH ₄ O) Clusters that Leads to O-H Bond Lengthening and C-O Bond Shortening Explaining the Observed Red and Blue Wavenumber Shifts, Respectively, within the CH ₄ O IR Spectra Component of the Engine Oil at High Temperature	132
Figure 7-16: Comparison of Actual Methanol (CH ₄ O) Spectra to Modified CH ₄ O Spectra with the O-H Stretching Bands Red Shifted and the C-O Stretching Bands Blue Shifted Bands Due to High Temperature Disturbance of Hydrogen Bonds	134
Figure 7-17: CH ₄ O (Modified)/CH ₂ O/CO ₂ Gas Mixtures Calibration Data Set, [X _C] _(10 x 1763) ; Note – Only 3 of 10 Calibration Spectra Shown	135
Figure 7-18: Modified Predicted Spectra for BP Turbo Oil 2380 based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 2.1%	136
Figure 7-19: Modified Predicted Spectra for BP Turbo Oil 274 based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 2.7%	136
Figure 7-20: Modified Predicted Spectra for Mobile Jet Oil II based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 2.2%	137
Figure 7-21: Modified Predicted Spectra for Aeroshell Turbine Oil 560 based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 3.4%	137
Figure 7-22: Truncated (3200 – 1600 cm ⁻¹) CH ₄ O/CH ₂ O/CO ₂ Gas Mixtures Calibration Data Set, [X _C] _(10 x 830) ; Note – Only 3 of 10 Calibration Spectra Shown %	138
Figure 7-23: Modified Predicted Spectra for BP Turbo Oil 2380 based on Truncated (3200 – 1600 cm ⁻¹) PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 2.1%	139
Figure 7-24: Modified Predicted Spectra for BP Turbo Oil 274 based on Truncated (3200 – 1600 cm ⁻¹) PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 2.7%	140
Figure 7-25: Modified Predicted Spectra for Mobile Jet Oil II based on Truncated (3200 – 1600 cm ⁻¹) PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 2.3%	140
Figure 7-26: Modified Predicted Spectra for Aeroshell Turbine Oil 560 based on Truncated (3200 – 1600 cm ⁻¹) PCR Calculated Concentrations of CH ₄ O, CH ₂ O, and CO ₂ Gas Mixtures; RMSE = 3.4%	141

Figure 7-27: Methanol (CH ₄ O), Formaldehyde (CH ₂ O), Carbon Dioxide (CO ₂), Carbon Monoxide (CO), and Water (H ₂ O) Pure Spectra for Simulated Data Set Illustrating Spectral Overlap Between CH ₄ O, CH ₂ O, and H ₂ O	143
Figure 7-28: BP Turbo Oil 2390 Time Evolved Spectra Used for Prediction Data Set, [X _P] _(20 x 1763) ; Note – Only 2 of 20 Prediction Spectra Shown, Time 5 minutes and 90 minutes	144
Figure 7-29: Principal Component Regression (PCR) Calculated Gas Concentrations for CH ₄ O, CH ₂ O, and CO of BP Turbo Oil 2380 Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off	145
Figure 7-30: Principal Component Regression (PCR) Calculated Gas Concentrations for CO ₂ of BP Turbo Oil 2380 Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off	146
Figure 7-31: Principal Component Regression (PCR) Calculated Gas Concentrations for H ₂ O of BP Turbo Oil 2380 Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off	146
Figure 7-32: Predicted Spectra for BP Turbo Oil 2380 at Time = 10 min. based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, CO ₂ , CO, and H ₂ O Gas Mixtures; RMSE = 2.7%	147
Figure 7-33: Predicted Spectra for BP Turbo Oil 2380 at Time = 30 min. based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, CO ₂ , CO, and H ₂ O Gas Mixtures; RMSE = 8.4%	148
Figure 7-34: Predicted Spectra for BP Turbo Oil 2380 at Time = 60 min. based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, CO ₂ , CO, and H ₂ O Gas Mixtures; RMSE = 9.2%	148
Figure 7-35: Predicted Spectra for BP Turbo Oil 2380 at Time = 90 min. based on PCR Calculated Concentrations of CH ₄ O, CH ₂ O, CO ₂ , CO, and H ₂ O Gas Mixtures; RMSE = 9.0%	149
Figure 7-36: Principal Component Regression (PCR) Calculated Gas Concentrations for CH ₄ O, CH ₂ O, and CO of BP Turbo Oil 2380 Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off	150
Figure 7-37: Principal Component Regression (PCR) Calculated Gas Concentrations for H ₂ O of BP Turbo Oil 2380 Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off	150

Figure 7-38: Predicted Spectra for BP Turbo Oil 2380 at Time = 10 min. based on Truncated (4000-2000 cm^{-1}) PCR Calculated Concentrations of CH_4O , CH_2O , CO_2 , CO , and H_2O Gas Mixtures; RMSE = 2.2% 151

Figure 7-39: Predicted Spectra for BP Turbo Oil 2380 at Time = 30 min. based on Truncated (4000-2000 cm^{-1}) PCR Calculated Concentrations of CH_4O , CH_2O , CO_2 , CO , and H_2O Gas Mixtures; RMSE = 7.2% 152

Figure 7-40: Predicted Spectra for BP Turbo Oil 2380 at Time = 60 min. based on Truncated (4000-2000 cm^{-1}) PCR Calculated Concentrations of CH_4O , CH_2O , CO_2 , CO , and H_2O Gas Mixtures; RMSE = 7.8% 152

Figure 7-41: Predicted Spectra for BP Turbo Oil 2380 at Time = 90 min. based on Truncated (4000-2000 cm^{-1}) PCR Calculated Concentrations of CH_4O , CH_2O , CO_2 , CO , and H_2O Gas Mixtures; RMSE = 7.5% 153

List of Abbreviations and Mathematical Symbols

ACER	Airliner Cabin Environment Research Center
AMU	Atomic Mass Unit
BTI	Business Travel Index
CFD	Computational Fluid Dynamics
CH ₂ O	Formaldehyde
CH ₄ O	Methanol
C ₃ H ₄ O	Acrolein
CO	Carbon Monoxide
CO ₂	Carbon Dioxide
FAA	Federal Aviation Administration
FTIR	Fourier Transform Infrared
H ₂ O	Water
H ₂ O ₂	Hydrogen Peroxide
HEPA	High Efficiency Particulate Air
IR	Infrared
MS	Mass Spectrometry
NBTA	National Business Travel Association
O ₃	Ozone
OHRCA	Occupational Health Research Consortium in Aviation

OSHA	Occupational Safety and Health Administration
Pa	Pascal
PCA	Principal Component Analysis
PCC	Proportionality Constant Calculation
PCR	Principal Component Regression
PPM	Parts per Million
PSI	Pound per Square Inch
RMSE	Root Mean Square Error
RMSEC	Root Mean Square Error Calibration
RMSEP	Root Mean Square Error Prediction
SARS	Severe Acute Respiratory Syndrome
SCCM	Standard Cubic Centimeter
S/N	Signal-to-Noise
STP	Standard Temperature and Pressure
SVD	Singular Value Decomposition
TCP	Tricresyl Phosphate
TGA	Thermogravimetric Analysis
TWA	Time Weighted Average
VPHP	Vapor Phase Hydrogen Peroxide
XPS	X-ray Photoelectron Spectroscopy

Chapter 1: Introduction

In today's turbulent airline industry, a company's survival is highly dependent on maintaining strict operational cost controls. With the current global economic downturn not expected to recover for some time, the operational cost driver is becoming even more important in order for a company to maintain expected overall profit margins and in some cases to reduce massive operational net losses with declining revenues. The Business Travel Index (BTI) indicated that 2010 would only have slight increases, 3.8% compared to 2009, in overall business travel growth, which is a major revenue source for the airline industry [1]. Michael McCormick, National Business Travel Association (NBTA) Executive Director and COO, recently said in an interview "we're looking forward to the end of 2012 – when the industry should see a return to peak levels" [1]. In the long-term, the Federal Aviation Administration (FAA) has projected that one billion passengers will be flying per year in 2023 and that the number of passengers will continue to grow [2].

Even though the industry has a strong focus on operational cost controls to maintain profitability, this should not come at the expense of safety. Terrorist attacks utilizing aircraft, such as those against the World Trade Center on September 11, 2001, are a global safety concern that not only the airline industry has worked hard to prevent, but that also has required significant contribution from many national governments in the form of national defense intelligence gathering and information sharing [3, 4]. In addition, both domestic and international transmission of diseases, such as severe acute respiratory syndrome (SARS) and the

recent outbreak of H1N1 flu in Latin America, on aircraft is a major concern as well [5-7]. These two focus points regarding safety currently receive the majority of media attention and rightfully so, but there is also a safety risk that many passengers as well as a number of airline employees are just beginning to realize.

The issue of air quality within the aircraft cabin is receiving increasing attention from both pilot and flight attendant unions [7]. This is due to exposure events caused by poor air quality that in some cases may have contained toxic oil components due to bleed air that flows from outside the aircraft through the engines into the aircraft cabin [8-10]. The system that supplies bleed air to the aircraft is shown schematically in Figure 1-1.

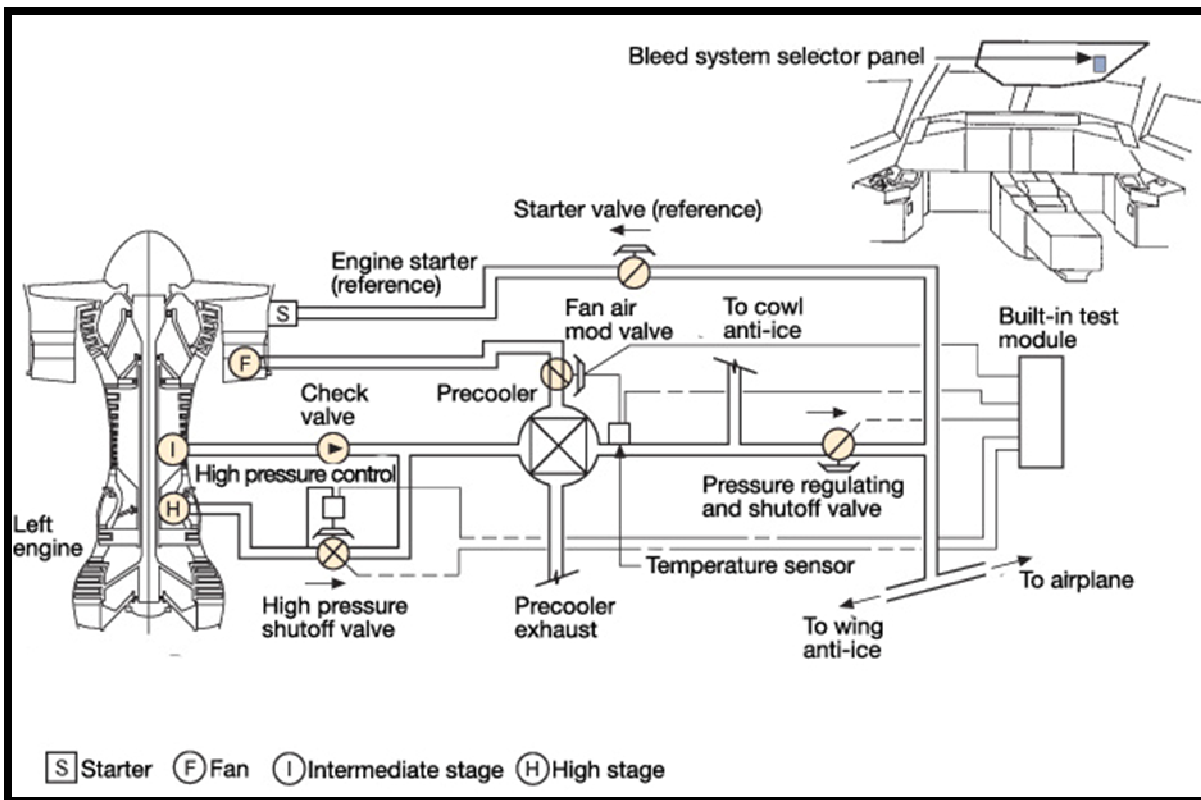


Figure 1-1: Commercial Aircraft Bleed Air System [11]

Exposure events to contaminated air have been suggested as the primary cause of significant short and long-term medical issues for aircraft crew in which it was thought that a

leak in the bleed air system occurred during flight. In 2009, the FAA Office of Aviation Medicine collaborated with the Occupational Health Research Consortium in Aviation (OHRCA) as well as the Airliner Cabin Environment Research Center (ACER) to create a guide for health care providers to deal with this specific issue [12]. The need for air quality monitoring is especially evident in the fact that currently within an aircraft there are no sensors to monitor the air quality and potentially harmful gas levels (detect-to-warn sensors), much less systems to monitor and purify the air (detect-to-treat sensors) within the aircraft cabin [13]. In 2009, the FAA estimated the total number of aircraft in the U.S. commercial fleet was 7,132, with 3,666 mainline passenger aircraft and 2,612 regional carrier aircraft [7]. With these numbers of aircraft currently in service, it is not feasible to replace the current aircraft fleet with new model aircraft that do not utilize a bleed air system, such as the Boeing 787 Dreamliner [14].

Research Purpose

The specific purpose of this research is to utilize a mathematical technique called principal component analysis (PCA) in conjunction with principal component regression (PCR) and proportionality constant calculations (PCC) to simplify complex, multi-component infrared (IR) spectra data sets into a reduced data set used for determination of the concentrations of the individual components [15]. This can significantly simplify data analysis as well as improve the ability to determine concentrations of individual target species in gas mixtures where significant band overlap in the IR spectrum region occurs. PCR is a mathematical technique that determines component concentrations of a prediction data set based on multivariate regression of a calibration data set. For PCC, the total integrated intensity of an IR absorbance band is assumed to increase linearly with the amount of the component concentrations in a mixture. Based on this

assumption, one can determine the proportionality constant for each of the individual components in a calibration data set. With the proportionality constants known for each component, a relationship to calculate the prediction data set component concentrations can be derived. In some cases where the overall volume is not constant, the PCC technique is expanded to perform the analysis on volume and integrated area fractions instead of components. Application of this analytical numerical technique to IR spectrum analysis is important in improving performance of commercial sensors that airlines and aircraft manufacturers could potentially use in an aircraft cabin environment for multi-gas component monitoring.

Research Approach

The approach of this research is to utilize PCA with both simulation and experimental results in conjunction with PCR and PCC to determine quantitatively the component concentrations within a mixture. In the simulation data sets, pure spectra from the QASoft® database (Infrared Analysis, Inc., Anaheim, CA, USA) are used. To form a simulated mixture, pure spectra are added together and different multiplication factors are applied to achieve a range of component concentrations. The simulated data sets consist of various spectra of two and three targeted component systems.

The experimental data sets consist of both two and three targeted component systems that could potentially be present as air contaminants in an aircraft cabin. In addition, experimental data sets are analyzed for a hydrogen peroxide (H_2O_2) aqueous solution mixture to determine H_2O_2 concentrations at various levels that could be produced during use of a vapor phase hydrogen peroxide (VPHP) decontamination system. After the PCA application to two and three component systems, the analysis technique is further expanded to include the monitoring of

potential bleed air contaminants from engine oil combustion, in which a simulation data set is utilized to predict gas components and concentrations in unknown engine oil samples. For the analysis of combusted aircraft engine oils, a simulated data set is used for PCA to determine regression coefficients for PCR to apply to the experimentally obtained data.

Organization of Dissertation

This dissertation contains a systematic utilization of PCA in conjunction with Fourier Transform infrared (FTIR) spectroscopy data for components that are applicable to the airline industry. Chapter 2 provides a detailed background of commercial aircraft systems that are applicable to this research. The background information includes a discussion on environmental air contaminants and bleed air contaminants that could potentially be present in aircraft cabins due to circulation of outside air of the combusted engine oils. Also within Chapter 2, the experimental aircraft cabin simulation environment is detailed. Next, in Chapter 3 a detailed overview of FTIR spectroscopy is presented. This discussion provides theoretical background on the FTIR spectroscopy technology as well as details on the experimental procedure and materials used in the collection of FTIR spectroscopy data. In addition, Chapter 3 also provides in-depth background on the characteristic IR modes of vibration on the key molecules of interest to this research, CO, CO₂, and H₂O. Chapter 4 then provides a detailed and mathematical background of PCA as well as the associated PCR and PCC techniques. Chapter 5 highlights the application of PCA to FTIR spectroscopy data from solutions to calculate H₂O₂ concentrations in an aqueous solution that could potentially be present during a VPHP decontamination event. The PCC technique for a variable volume of solution is utilized with this analysis in Chapter 5. Chapter 6 then uses PCA to obtain results that identify the individual component spectra within a

multi-component system with both simulated and experimental FTIR spectroscopy data sets. Chapter 6 includes the application of PCA to the monitoring of concentrations of environmental air contaminants. Concluding the PCA on FTIR spectroscopy data in Chapter 7, is an application of the technique to the monitoring of potential bleed air contaminants within the aircraft cabin, which requires the use of simulated data sets to predict the components and concentrations of gas species. The PCR technique is used to determine the quantitative concentrations of the individual components of the system and these concentrations are used to calculate predicted spectra for the oils and then compared to the original spectra to determine the experimental error. Chapter 8 summarizes the research findings and contains concluding remarks about the potential scientific impact of the results found within this dissertation. Chapter 9 contains a brief discussion of potential future work that could further both scientific and engineering understanding of PCA application to FTIR spectroscopy. The references cited are found at the end of this dissertation. Appendix A contains the MATLAB® source code used to perform the matrix manipulations for PCA.

Chapter 2: Commercial Aircraft Systems Overview

Commercial Aircraft Background

Within the typical aircraft, numerous systems are responsible for the stability and operation of a desirable environmental control. This project focuses on the main subsystems of the aircraft environmental control system (ECS) relating to cabin air quality, which are the bleed air controls, air conditioning pack, mix manifold, recirculation devices, and the cabin vents [11, 16].

The ECS air controls takes in outside air while the aircraft is in operation using bleed air controls. This outside air is compressed to 220 kPa (32 psi) and rises to a temperature of 160 °C (320 °F). This system, shown in Figure 1-1, has a number of valves and heat exchangers that conditions the air to a desirable temperature and pressure for the other flight systems. During flight, air entering the bleed air system could potentially have high concentrations of ozone (O₃) due to elevated atmospheric concentrations at the flight altitude. Typical levels of O₃ in the outside air range from 0.5 to 1.0 parts per million (ppm) [11, 17, 18]. Most of this O₃ partially dissociates when it goes through compression stages of the engine and the catalytic converter but significant and harmful amounts have been measured in simulated aircraft cabin environments [17, 18].

After air leaves the bleed air system, it then enters the air conditioning (AC) packs where it is cooled to a temperature of about 15 °C (59 °F) and decompressed to a pressure of 78-82 kPa

before continuing to the mix manifold. This pressure range corresponds to the typical aircraft cabin altitude setting that ranges from 6,000-8,000 feet above sea level. Table 2-1 summarizes the changes in pressure in relation to altitude. One item to note is the air that enters the mix manifold is not monitored for harmful gas concentrations. In addition to O₃ that may still be present, the CO₂ and CO proportional concentrations in the air are unchanged from outside levels.

Table 2-1: Relationship between Altitude and Pressure with the Typical Aircraft Cabin Pressure Highlighted

Altitude		Pressure		
[feet]	[m]	[kPa]	[atm]	[in Hg]
0	0.0	101.3	1.00	29.9
500	152.4	99.5	0.98	29.4
1000	304.8	97.7	0.96	28.9
1500	457.2	96.0	0.95	28.3
2000	609.6	94.2	0.93	27.8
2500	762.0	92.5	0.91	27.3
3000	914.4	90.8	0.90	26.8
3500	1066.8	89.1	0.88	26.3
4000	1219.2	87.5	0.86	25.8
4500	1371.6	85.9	0.85	25.4
5000	1524.0	84.3	0.83	24.9
6000	1828.8	81.2	0.80	24.0
7000	2133.6	78.2	0.77	23.1
8000	2438.4	75.3	0.74	22.2
9000	2743.2	72.4	0.71	21.4
10000	3048.0	69.7	0.69	20.6

The change in total pressure has an effect that is described by Dalton’s Law of Partial Pressure shown in Equation 2.1, in which the partial pressure of a gas, p_i , is a product of the mole fraction, X_i , of the gas and the total pressure of the gas mixture, P_T .

$$p_i = X_i P_T \qquad \text{(Dalton’s Law of Partial Pressures)} \qquad (2.1)$$

With total pressure changes due to changes in altitude, the measured concentration of a gas of interest, if the sensor measurement principal is based on partial pressure, will be affected. This effect will thus change readings of a gas ideal partial pressure sensor. As shown in Figure 2-1, an ideal partial pressure sensor used to measure CO₂ levels at approximately 7,000 ft altitude, where the total pressure is approximately 77% that of sea level, would read $p_{CO_2} = 0.39$ kPa for a true concentration of $p_{CO_2} = 0.51$ kPa measured at standard temperature and pressure (STP).

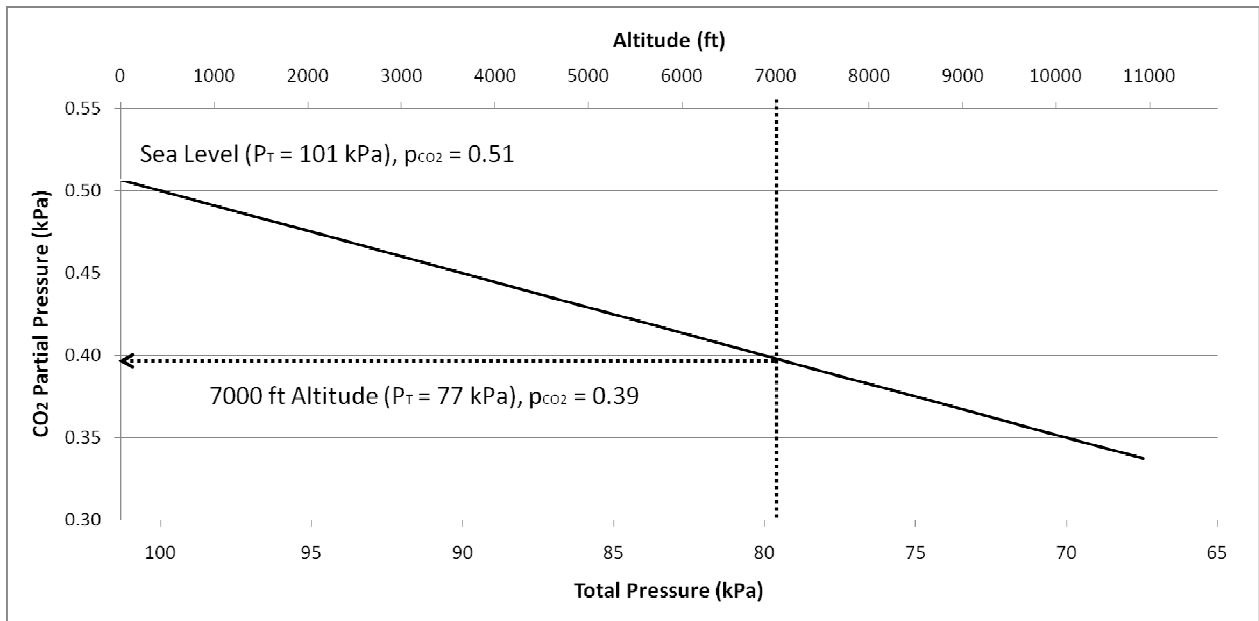


Figure 2-1: Characteristics of Ideal CO₂ Partial Pressure Sensor; Measures $p_{CO_2} = 0.39$ at $P_T = 77$ kPa (7000 ft Altitude) for $p_{CO_2} = 0.51$ at $P_T = 101$ kPa (Sea Level)

The HEPA filters, similar to those used in critical wards of a hospital, are present in the recirculation system, and when in a relatively new condition remove 99.97% of bacteria and viruses at a particle size of 0.3 μm produced or brought on board the aircraft by passengers [11, 16]. These filters, however, do not filter harmful gases such as CO or CO₂ that may be present. The system attempts to control the levels of gases that may be present due to internal contamination of air within the aircraft through dilution with high quantities of outside air as

highlighted by the 10-15 outside air changes per hour shown in Figure 2-2 in comparison to hospital delivery and operating rooms as well as the typical building [16].

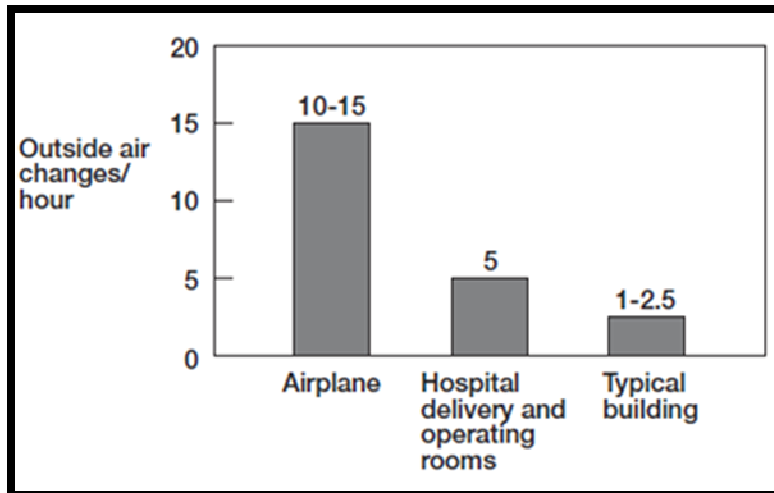


Figure 2-2: Recirculation System Outside Air Changes per Hour for Aircraft Compared with Other Environments [16]

The final subsystem the air passes through as it reaches the passengers is the cabin ventilation system shown in Figure 2-3. The airflow is directed from overhead air supply nozzles and extracted through return air grilles where the sidewall meets the floor along the length of the cabin. The air here has a typical temperature of 18-30 °C (64-86 °F) and a relative humidity of 10-20% [19]. Within the ventilation system, for nearly all commercial aircraft there are currently no sensors or monitoring for potentially harmful gases but recent computational fluid dynamics (CFD) simulation work has been conducted to determine the optimal position to place sensors when and if they are installed to ensure the earliest warning possible to both flight crew and passengers [20].

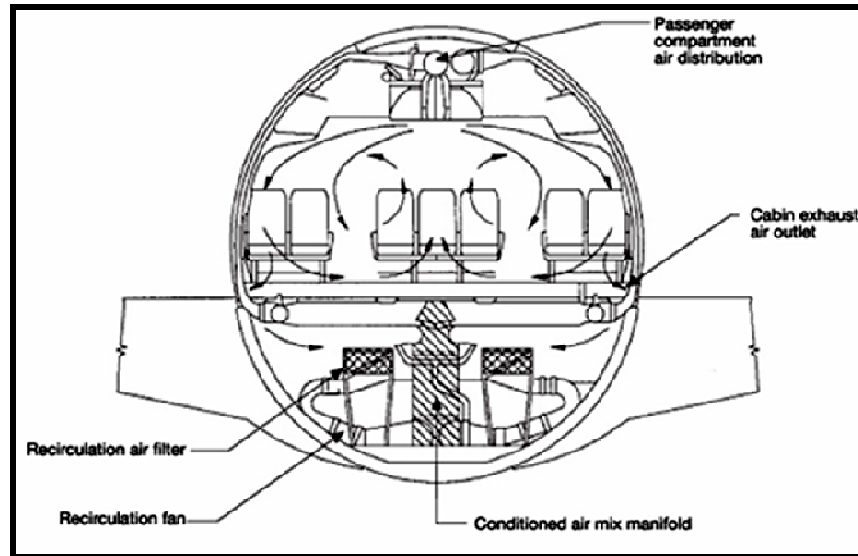


Figure 2-3: Cabin Ventilation System Illustrating Typical Aircraft Recirculated Airflow [19]

Vapor Phase Hydrogen Peroxide (VPH) Aircraft Cabin Decontamination

An increasing awareness towards aircraft cabin sterilization for biological and chemical contaminants has led to the development of full-scale methods using VPH [21, 22, 23]. VPH at concentrations greater than 80 ppm have been shown to have sporicidal effects, while the typical aircraft cabin sterilization utilizes VPH concentrations in the range of 150-600 ppm [21]. In addition, the typical VPH process contains an initial dehumidification process that reduces the relative humidity to less than 10% [21]. Concentrations of the initial liquid condensing from the H_2O_2 - H_2O vapor can be as high as 50-75 wt.% H_2O_2 even though the original flash vaporized liquid is only 35 wt.% and these high H_2O_2 concentration in the condensate have been shown to increase susceptibility to hydrogen embrittlement for 4340 high strength steel [24].

For VPH, three major processing parameters affect inactivation of microorganisms. These three factors are sterilant concentration, exposure time, and percent saturation. Although

commercial systems are available, monitoring the H_2O_2 and H_2O conditions during operation require specialized sensors, such as those from Analytical Technology, Inc (Collegeville, PA, USA). These sensors as well as others are primarily used to monitor H_2O_2 and H_2O in the gas phase. Although the occurrence of microcondensation can be detected with optical dew point sensors, accurate monitoring of the concentrations of condensates is not routinely practiced [25].

Potential Environmental Air Contaminants within the Aircraft Cabin

During flight operations, air quality of an aircraft cabin is critical to crew and passenger safety and comfort and as noted previously. However, there are currently no environmental monitoring sensors present in the aircraft cabin. By diluting the aircraft cabin air with high quantities of outside air toxic gases, such as CO , CO_2 , and O_3 are assumed to be below harmful levels. Even so, recent aircraft cabin studies have shown that formaldehyde (CH_2O) has been specifically detected as a reaction product of ozone-initiated chemistry due to ozone reactions with human skin oils, hair, and clothing as well as the fabric within the aircraft cabin [26]. In [17], both CH_2O and acrolein ($\text{C}_3\text{H}_4\text{O}$) resulting from O_3 interactions were detected at concentrations exceeding their OSHA recommended exposure limits.

Table 2-2 highlights the limits that the FAA and the Occupational Safety and Health Administration (OSHA) currently have on some contaminants of interests that could potentially be found in aircraft cabin air [27]. In Table 2-2, it should be noted that time weighted average (TWA) is the average concentration in a normal 8-hour workday and a 40-hour workweek. In addition, the ppm levels are sea level equivalent values.

Table 2-2: Limits on Contaminants that Could Potentially Be Found in Aircraft Cabin Air [27]

Contaminants	FAA Limit	OSHA Permissible Exposure Limit
Carbon Monoxide (CO)	50 ppm	50 ppm
Carbon Dioxide (CO ₂)	5000 ppm	5000 ppm
Ozone (O ₃)	0.1 ppm	0.1 ppm
Formaldehyde (CH ₂ O)	N/A	0.75 ppm (TWA)
Acrolein (C ₃ H ₄ O)	N/A	0.1 ppm

* TWA – Average concentration in a normal 8-hour workday and a 40-hour workweek

Potential Bleed Air Contaminants within Aircraft Cabin

A bleed-air system within an aircraft is very beneficial in that the compressed air that it produces can be used as a major power source for many environmental control systems from de-icing the wings of a plane to pressurizing the cabin. A drawback of this system though, is that it has the potential to allow contaminated air from the environment during taxiing operations, as well as noxious gases due to possible leaks of engine oil, hydraulic fluid, de-icing fluid, etc., into the aircraft cabin.

In addition to forming due to O₃ reactions, CH₂O and C₃H₄O have been shown to form when engine oil is burned [28]. The oils and hydraulics used in aircraft are also known to contain toxic chemicals, such as the irritant phenyl-alpha-naphthylamine and the neurotoxin tricresyl phosphate (TCP) [10, 29]. In 2000, measurement of various gases and volatile compounds from various engine oils showed that oil pyrolyzed at 525 °C (977 °F) generated significant amounts of CO₂ and CO in excess of 100 ppm [30]. In addition, TCP was found within the samples using a gas chromatography (GC) laboratory measurement technique [30].

Laboratory Test Environment

The experimental setup consists of three major modules. The first is the Control Module, shown in Figure 2-4, which is responsible for control of pressure and the flow of both inert

carrying gases as well as test gases of interest. The pressure setting within the system allows for testing of sensors at various altitudes that are encountered in the airplane cabin environment. The gas lines in the system are rated for vacuum pressures of 15 inches of mercury (380 mm Hg) or about 50% of atmospheric pressure (50.5 kPa), which corresponds to altitudes up to 12,000 feet (3,700 meters). The flow meters allow precise control of these gases and allow custom mixing ratios for sensor performance testing.

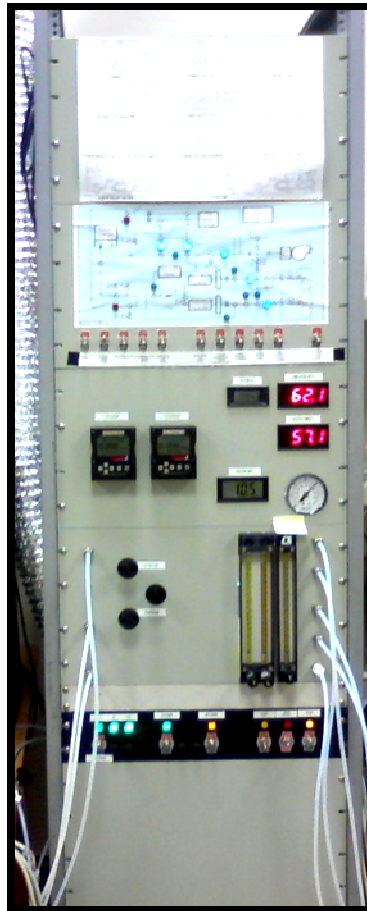


Figure 2-4: Laboratory Test Environment, Control Module

The second module, the Commercial Sensor Analysis Module, shown in Figure 2-5, is an enclosed, vacuum-sealed, Plexiglas (PMMA) chamber, which has a total volume of 42.4 liters.

With this module, an environment replicating the various airplane cabin conditions can be maintained to test commercial sensor performance in regards to detection of the gases of interest.

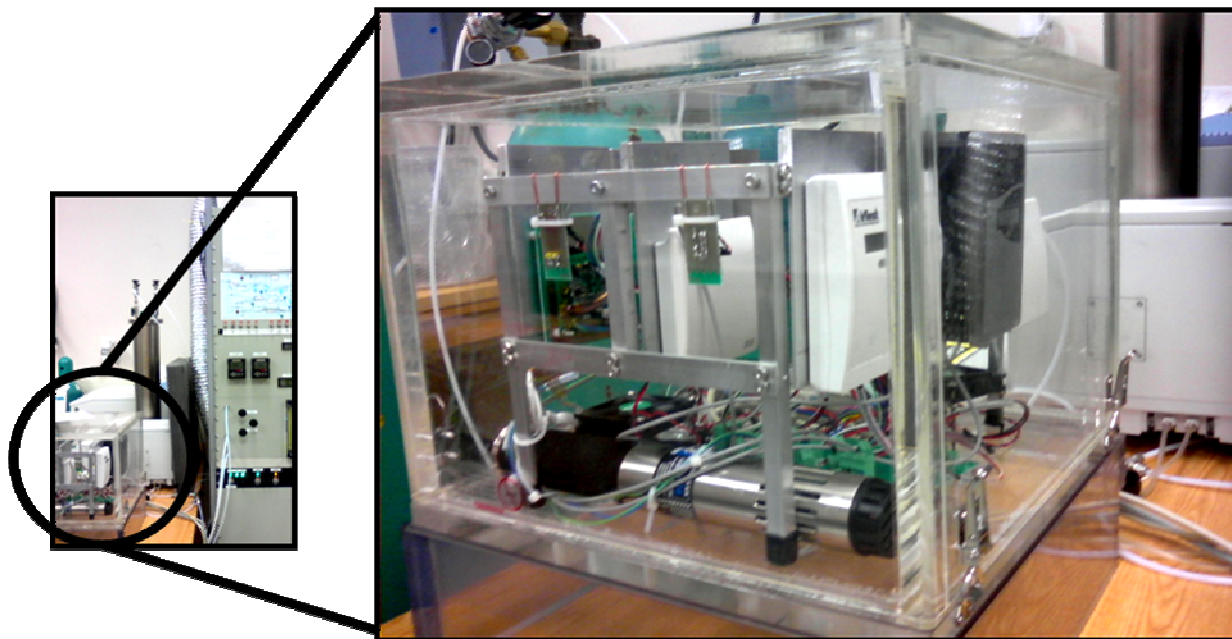


Figure 2-5: Laboratory Test Environment, Commercial Sensor Analysis Module

The final module shown in Figure 2-6, which is used as the standard in evaluating commercial sensor performance as well as for independent gas analysis studies, is the FTIR Gas Analysis Module. This module contains a Spectrum GX FTIR System (Perkin Elmer, Shelton, CT, USA), as well as an M-5-22-V variable pathlength long path gas cell (Infrared Analysis, Inc., Anaheim, CA, USA). The optical path is folded in a volume of 8.5 liters, while the cell path length is determined by the number of passes times the base path length (56 cm). The FTIR spectrometer can take scans over a possible wavenumber scan range from $10,000\text{ cm}^{-1}$ to 400 cm^{-1} with possible spectral resolutions of 64 cm^{-1} to 0.5 cm^{-1} . The IR source is produced by a temperature stabilized wire coil that operates at 1350 K. The windows in the variable pathlength long pass gas cell are made of potassium chloride (KCl) and are 4 mm thick. The detector for

the IR beam is a fast recovery deuterated triglycine sulfate (FR-DTGC) module, which is standard for the mid-IR region of interest.



Figure 2-6: Laboratory Test Environment, FTIR Gas Analysis Module with Spectrum GX FTIR (Perkin Elmer, Shelton, CT, USA) with M-5-22-V Variable Pathlength Long Path Gas Cell (Infrared Analysis, Inc., Anaheim, CA, USA)

Shown in Figure 2-7 are the internal gold plated mirrors of the variable pathlength long path gas cell. Adjustments to the mirrors allow for multiple passes of the IR beam within the gas cell. Using a viewing window located on the top of the long path gas cell, a laser can be used to see the number of passes that the IR beam will make. The number of passes that the IR beam traverses in the long path gas cell is 4 times the number of laser dots found on the bottom row on the gold plated mirror when looking through the viewing window. The base pathlength of the M-5-22-V variable pathlength long path gas cell is 0.56 m, with a minimum number of passes of

4 and a maximum number of passes of 64. These adjustments to the pathlength allows for variable pathlength within the long path gas cell ranging from 2.24 m to 35.84 m (64 passes).

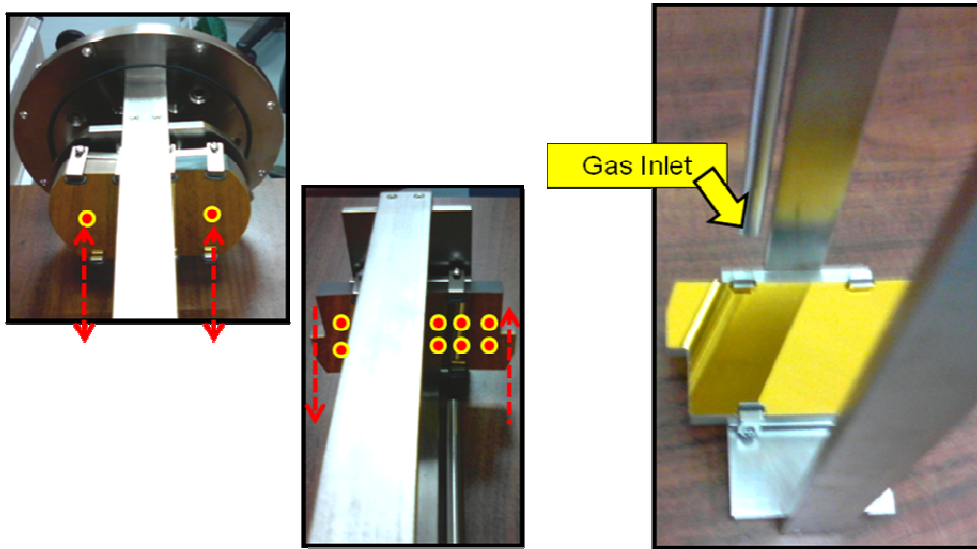


Figure 2-7: Internal Gold Plated Mirrors of the Variable Pathlength Long Path Gas Cell Highlighting the Path of the IR Beam within the Cell

With the flow controls on the Control Module and known volumes of the Commercial Sensor Analysis Module as well as the FTIR Gas Analysis Module, a differential equation based theoretical mixing model, based on the simplified system diagram shown in Figure 2-8, can be constructed. This model details the expected concentrations of gas at a given time during an experiment. The gases within the modules are assumed to be well mixed and this is accomplished using fans in the Commercial Sensor Analysis Module and the input of gas for the FTIR Gas Analysis Module being sufficiently far away from the output.

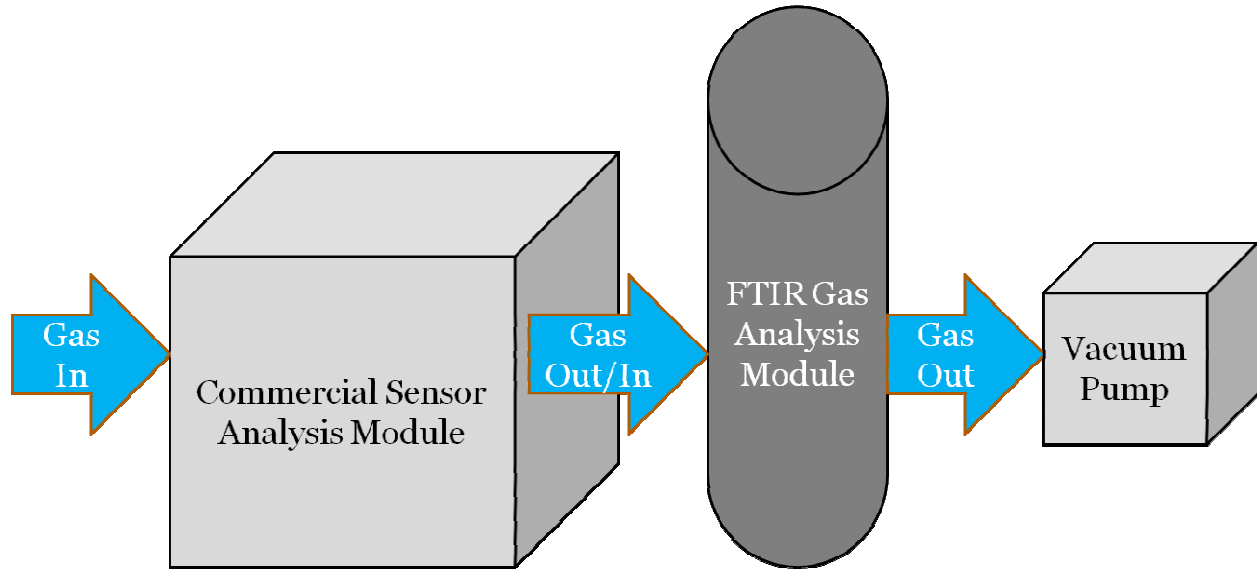


Figure 2-8: System Block Diagram Illustrating Gas Flows through the Commercial Sensor Analysis Module to the FTIR Gas Analysis Module then Out through the Vacuum Pump

The basis for this theoretical gas flow model is the assumption that the change in concentration of a test gas within the Commercial Sensor Analysis Module, dC/dt , equals the inflow rate, F_{in} , of the test gas minus the outflow rate of the test gas, F_{out} (Equation 2.2). The term F_{out} is multiplied by the concentration of the test gas at a given time divided by the total of volume of the module, V . This multiplication is necessary because the total outflow volume includes the carrier gases as well as the test gas. In addition, F_{out} , is a function of the ratio of atmospheric pressure, P (1 atm. = 101.325 kPa), versus the applied vacuum pressure to the system, P_{vac} . The inflow rate for the test and carrier gases are set with the flow controllers.

$$\frac{dC}{dt} = F_{in} + F_{out} \left(\frac{C}{V} \right) \left(\frac{P}{P_{vac}} \right) \quad (2.2)$$

The second step in the derivation of a theoretical model is to use algebra and separation of variables producing Equations 2.3 and 2.4.

$$\frac{dC}{dt} = -\left(\frac{F_{out}}{V}\right)\left(\frac{P}{P_{vac}}\right)\left[C - V\left(\frac{F_{in}}{F_{out}}\right)\left(\frac{P_{vac}}{P_a}\right)\right] \quad (2.3)$$

$$\frac{dC}{\left[C - V\left(\frac{F_{in}}{F_{out}}\right)\left(\frac{P_{vac}}{P_a}\right)\right]} = -\left(\frac{F_{out}}{V}\right)\left(\frac{P}{P_{vac}}\right)dt \quad (2.4)$$

Integrating Equation 2.4 gives Equation 2.5, with k being a constant of integration.

$$\ln\left[C - V\left(\frac{F_{in}}{F_{out}}\right)\left(\frac{P_{vac}}{P_a}\right)\right] = -\left(\frac{F_{out}}{V}\right)\left(\frac{P}{P_{vac}}\right)t + k \quad (2.5)$$

Next, taking the exponential of each side in Equation 2.5, the relationship shown in Equation 2.6 is produced and with further simplification, the equation for concentration as a function of time is given in Equation 2.7.

$$\left[C - V\left(\frac{F_{in}}{F_{out}}\right)\left(\frac{P_{vac}}{P_a}\right)\right] = k \exp\left[-\left(\frac{F_{out}}{V}\right)\left(\frac{P_a}{P_{vac}}\right)t\right] \quad (2.6)$$

$$C = V\left(\frac{F_{in}}{F_{out}}\right)\left(\frac{P_{vac}}{P_a}\right) + k \exp\left[-\left(\frac{F_{out}}{V}\right)\left(\frac{P_a}{P_{vac}}\right)t\right] \quad (2.7)$$

From Equation 2.7 and incorporation of the initial condition, the concentration of the test gas in the chamber, $C(0)$, at time, $t=0$, the value of the integration constant, k can be found as shown in Equation 2.8.

$$k = C(0) - V \left(\frac{F_{in}}{F_{out}} \right) \left(\frac{P_{vac}}{P_a} \right) \quad (2.8)$$

Now, the final relationship of test gas concentration within the chamber at a given time, t , can be found using Equation 2.9. An initial concentration of the test gas, $C(0)$, can be present before the gas flow begins and this value must be measured either with a sensor in the Commercial Sensor Analysis Module or with the FTIR in the FTIR Gas Analysis Module.

$$C = V \left(\frac{F_{in}}{F_{out}} \right) \left(\frac{P_{vac}}{P_a} \right) + \left[C(0) - V \left(\frac{F_{in}}{F_{out}} \right) \left(\frac{P_{vac}}{P_a} \right) \right] \exp \left[- \left(\frac{F_{out}}{V} \right) \left(\frac{P_a}{P_{vac}} \right) t \right] \quad (2.9)$$

This model was used to predict the expected CO concentration within the FTIR Gas Analysis Module for a given time. Shown in Figure 2-9 is a comparison between the calculations derived from the simplified model versus the FTIR measured CO concentrations for given test parameters. For this particular experiment, the inflow gas was a 800 ppm CO in N₂, which corresponds to a 0.40 standard cubic centimeter (sccm) flow rate with the flow controller set to 500 sccm.

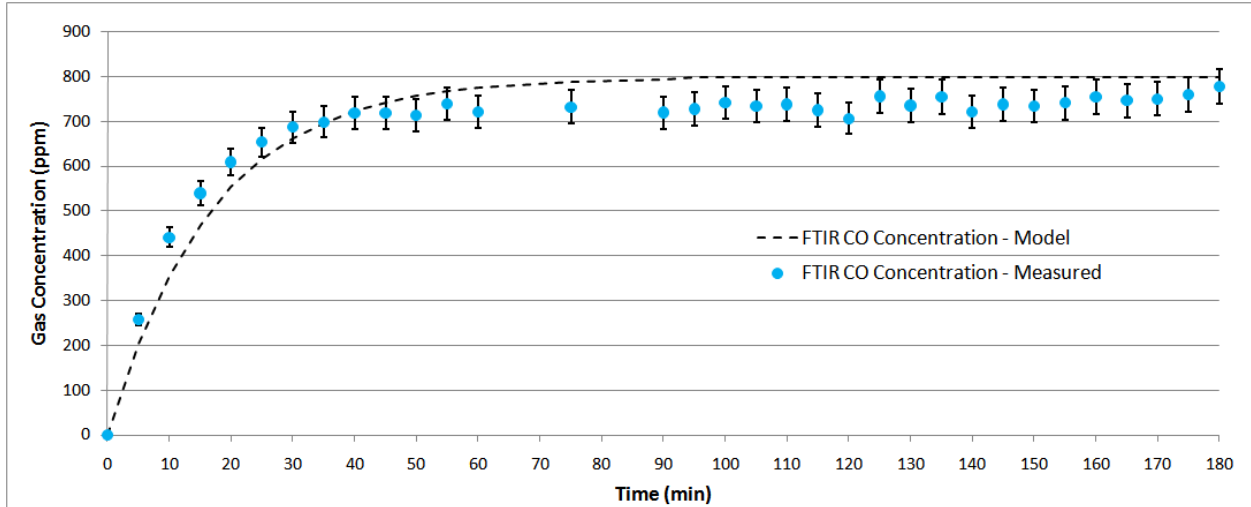


Figure 2-9: Theoretical Gas Flow Model Calculations Compared to Experimental Results for 800 ppm CO Test Gas in N₂ with Total Inflow of 500 sccm with Vacuum Pressure (P_{vac}) = Atmospheric Pressure (P_a)

For the case when the FTIR Gas Analysis Module is in series with the Commercial Gas Sensor Analysis Module, the inflow value to the model for FTIR Gas Analysis Module will be used to calculate outflow from the Commercial Sensor Analysis Module. This produces a lagging effect for both the theoretical and measured test gas concentrations in the FTIR Gas Analysis Module during the initial ramp up to the final steady-state test gas concentration. This effect is shown graphically for the theoretical models in Figure 2-10.

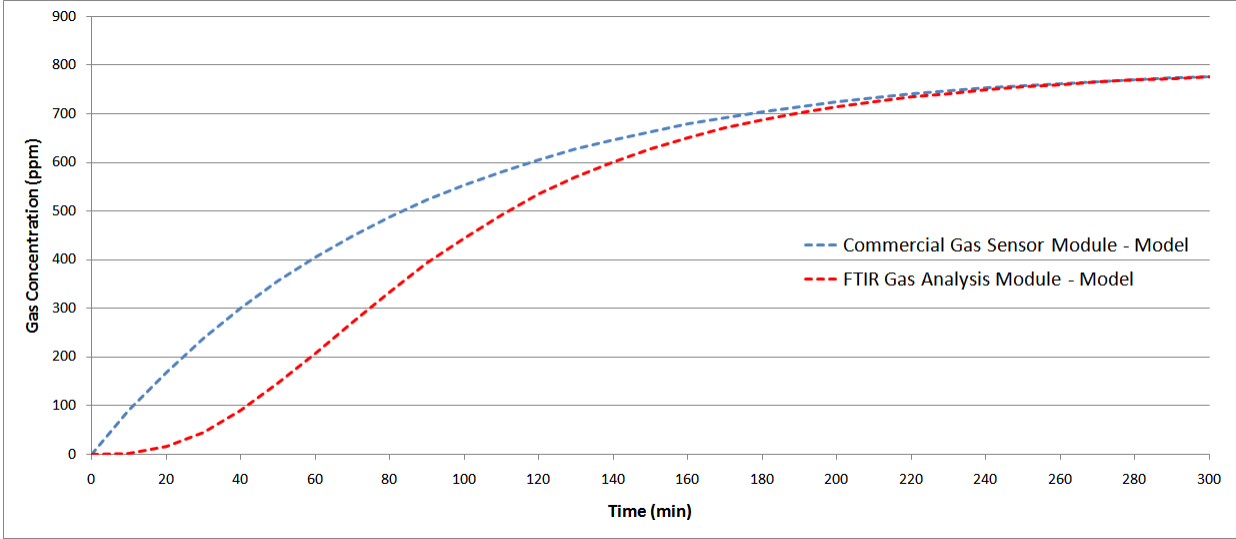


Figure 2-10: Theoretical Gas Flow Model Calculations for FTIR Gas Analysis Module in Series with the Commercial Gas Sensor Analysis Module

Chapter 3: Fourier Transform Infrared (FTIR) Spectroscopy Overview

FTIR Theoretical Background

FTIR analysis relies on the principle that all polyatomic molecules and hetero-nuclear diatomic molecules absorb IR radiation. When a molecule interacts with an IR source, it experiences a vibrational transition due to photon absorption, illustrated in Figure 3-1 in red, thus placing the molecule at a higher energy state [31-33]. In addition, when a molecules of gas absorbs IR energy, rotational transitions can occur in conjunction with the vibrational transitions. These rotational and vibrational transitions produce a number of relatively closely spaced absorption lines [34]. The total energy (E_{tot}) within a molecule is defined as three additive components (Equation 3.1), energy due to rotation of the molecule (E_{rot}), energy due to vibration of atoms (E_{vib}), and energy due motion of electrons (E_e).

$$E_{tot} = E_{rot} + E_{vib} + E_e \quad (3.1)$$

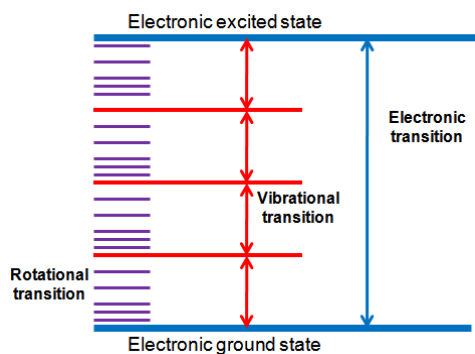


Figure 3-1: Energy Diagram Highlighting Transitions from Electronic Ground State to Electronic Excited State with Rotational and Vibrational Transitions

The physical properties of a molecule determine the pattern of absorption, defined by the IR wavelength at which the species absorbs. The major physical properties that help define a characteristic IR absorption spectrum for a molecule are the number of atoms, the bond angles, and the bond strengths [31]. Each IR absorption band, which is due to a particular vibrational energy change, is composed of a number of relatively closely spaced absorption lines and these components can be related to simultaneous rotational changes that accompany vibrational energy changes [31, 32]. For N atomic nuclei within a molecule, there are $3N-6$ degrees of freedom for a nonlinear molecule and $3N-5$ degrees of freedom for a linear molecule [33]. These degrees of freedom refer to the number of vibrational modes expected within a molecular structure. The absorption frequency depends on the molecular vibrational frequency, while the intensity of the molecular absorption depends on how efficiently IR energy is transferred to the molecule.

FTIR Measurement Technique

In regards to the FTIR measurement technique, the system is technically referred to as a Michelson Interferometer that produces a time domain measurement based on a path difference of two beams from a single an IR source. These beams are combined before interacting with an IR absorbing species using a specialized signal called an interferogram [35]. The interferometer utilizes a beam splitter that transmits about 50% and reflects about 50% of the incoming IR source thus dividing the signal it into two. One beam reflects off a flat mirror that is fixed and the other is reflected off a mirror that is allowed to move. When the two beams are recombined after reflecting off their respective mirrors, the resulting beams interfere with each other (Figure 3-2). Because the path of the beam on the moving mirror is constantly changing, the data points, which make up the interferogram signal, contain all the infrared source information from the IR

source within a very short time domain signal. Once the interferogram interacts with the sample, the resulting signal is then converted to the frequency domain through a mathematical technique called a Fourier Transform.

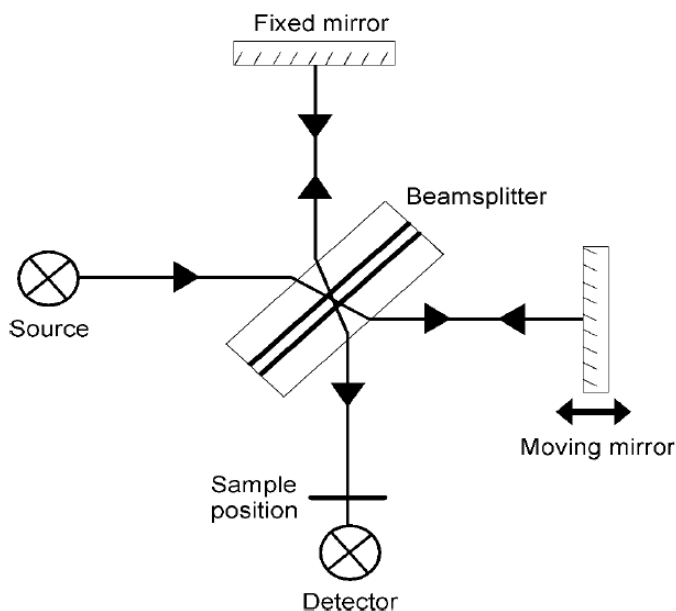


Figure 3-2: Beam Path from IR Source Highlighting the use of a Fixed Mirror and Moving Mirror to Construct an Interferogram Time Domain Signal Containing the IR Source Information [35]

The intensity of the transmitted power through the sample, I , is compared to the intensity of the IR radiation incident on the sample, I_0 and percent transmittance, $\%T$, value for each frequency is calculated (Equation 3.2). For quantitative analysis of a spectrum, this transmittance value is converted to a unit-less absorbance value, A , (Equation 3.3). Absorbance values are ideal for quantitative studies because as shown in Equation 3.4, the Beer-Lambert Law, absorbance is directly proportional to the concentration of light-absorbing species [31]. In Equation 3.4, ϵ is the molar absorptivity, b is the pathlength that the light source travels, and c is the concentration of the light-absorbing species. Table 3-1 highlights the relationship between I/I_0 , $\%T$, and A .

$$\%T = \frac{I}{I_0} \times 100\% \quad (3.2)$$

$$A = -\log_{10}(T) = \log_{10}\left(\frac{I_0}{I}\right) \quad (3.3)$$

$$A = \epsilon bc \quad (\text{Beer-Lambert Law}) \quad (3.4)$$

Table 3-1: Relationship between I/I_0 , $\%T$, and A

I/I_0	$\%T$	A
1	100	0
0.1	10	1
0.01	1	2
0.001	0.1	3
0.0001	0.01	4

It is necessary to collect a background spectrum sample before taking scans of the sample to remove instrument and atmospheric characteristics. It is important to note that the signal-to-noise (S/N) ratio is determined by both the sample spectrum and the background spectrum and typically it is necessary to have as many background co-added scans (scans averaged over a given spectral range) as samples co-added scans to obtain the best S/N. The S/N determines the weakest feature that can be confidently identified within a spectrum and is directly proportional to the number of scans taken for a sample or background. As an example, if it is necessary to double the S/N value of four co-added scans, it is necessary to collect 16 co-added scans.

When analyzing samples for strongly absorbing species such as CO_2 , it is not necessary to take as many scans, but to analyze species that do not absorb significantly in the IR region, it is necessary to have a high S/N value. Figure 3-3 shows a background scan of 16 co-added spectra, taken at a resolution of 0.5 cm^{-1} over a wavenumber range of $2500\text{-}1000 \text{ cm}^{-1}$. This

spectrum highlights the typical atmospheric interference of CO₂ in the wavenumber range of 2400-2250 cm⁻¹ and H₂O (water vapor) in the wavenumber range of 2100-1300 cm⁻¹.

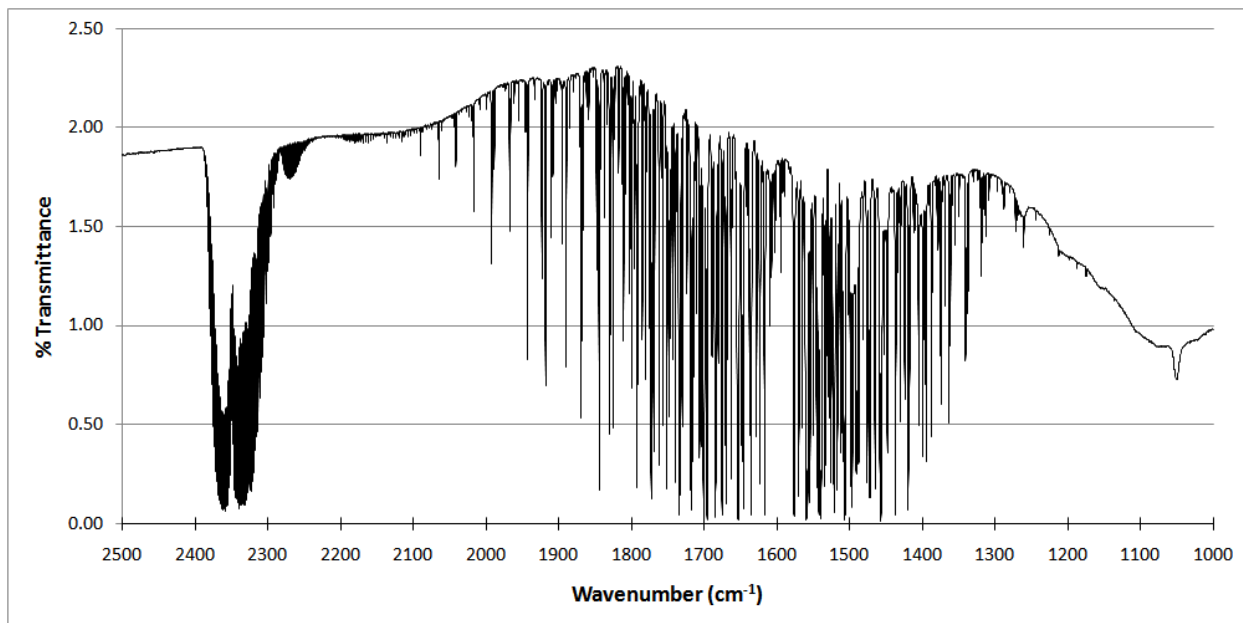


Figure 3-3: Typical Atmospheric Background Spectra Highlighting CO₂ and H₂O Interference, 16 Co-added Scans with Resolution 0.5 cm⁻¹

The most generally accepted resolution for gas analysis is 0.5 cm⁻¹ because this takes advantage of detailed fine structure in the bands of gaseous molecules and widens the range over which absorption is valid. In addition, the S/N is proportional to the resolution squared [35]. The resulting spectrum is then analyzed by comparison to known databases. The database used in this study is from QASoft® and contains molecular absorption spectra for 386 gases, with most of them being in an inert gas, such as N₂, to maintain total pressure of 1 atmosphere. The purpose of the background gas is to establish the total pressure of the system as close to atmosphere as possible while limiting background gas interference. Since N₂ is inactive in the IR region, it is a desirable gas for this purpose. This database covers the region of 3700 cm⁻¹ to 500 cm⁻¹, which is the fundamental IR region where rotation and vibrations of molecules give rise to IR absorption.

IR Characteristics of CO/CO₂/H₂O

Carbon Monoxide (CO)

CO is a homogeneous, linear, diatomic molecule and thus should have a single characteristic mode of vibration ($3N-5 = 3 \times 2 - 5 = 1$) [31, 33]. This vibrational mode, which is along the chemical bond, shown in Figure 3-4, has a characteristic vibrational frequency of 2143 cm^{-1} . The CO spectrum from the QASoft® database is shown in Figure 3-5.

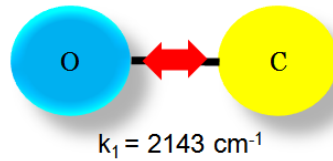


Figure 3-4: CO Fundamental Vibrational Mode, $k_1 = 2143 \text{ cm}^{-1}$

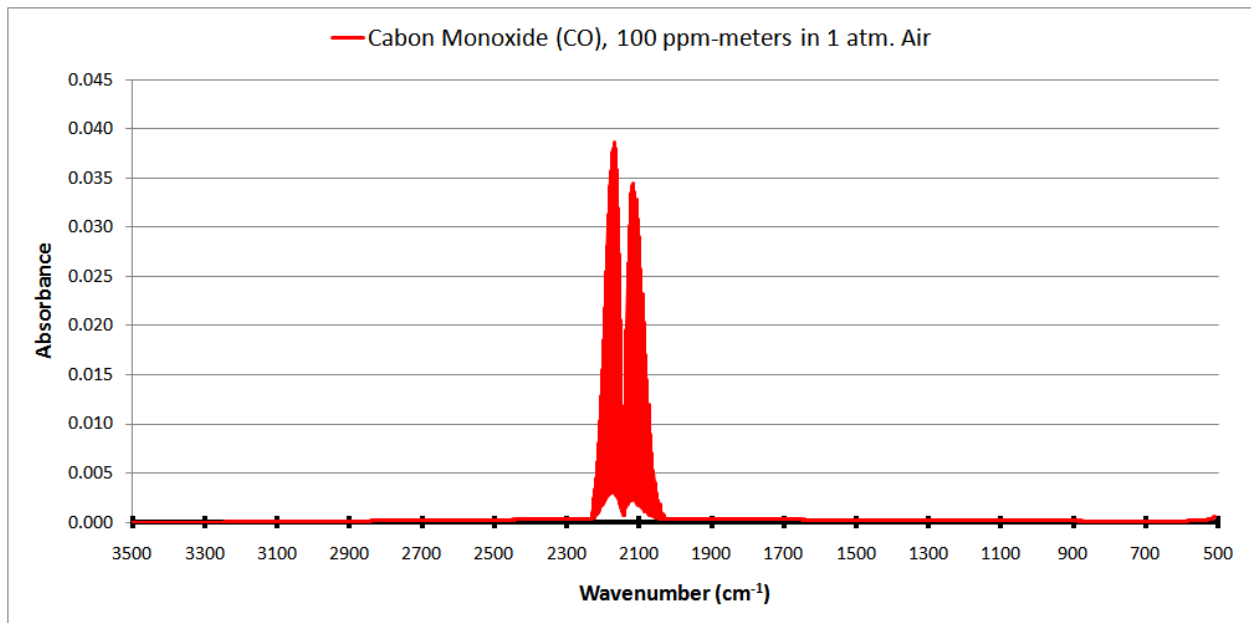


Figure 3-5: IR Absorbance Spectra for CO from QASoft® Database

Carbon Dioxide (CO₂)

CO₂ is a heterogeneous, linear, tri-atomic molecule that has four characteristic modes of vibration, based on the calculated 4 degrees of freedom ($3N-5 = 3 \times 3 - 5 = 4$) [31]. Figure 3-6 shows the CO₂ first fundamental vibrational mode, k_1 , which occurs with symmetrical motion of the oxygen atoms while the carbon atom is fixed. The characteristic vibrational frequency associated with this vibration occurs at 1340 cm^{-1} . For pure CO₂, this vibrational mode is inactive in the IR energy region because the molecular dipole moment is not changed with vibration.

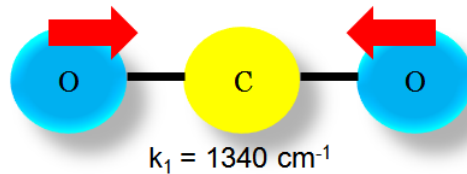


Figure 3-6: CO₂ First Fundamental Vibrational Mode, $k_1 = 1340 \text{ cm}^{-1}$

Figure 3-7 shows the CO₂ second fundamental vibrational mode, k_2 , which results when the carbon atom oscillates perpendicular to the oxygen atoms with the two vibrational modes arising from rotations by 90° [36]. Because the two vibrational modes are just rotations of the same molecular motion, they have the same fundamental vibrational frequency of 667 cm^{-1} .

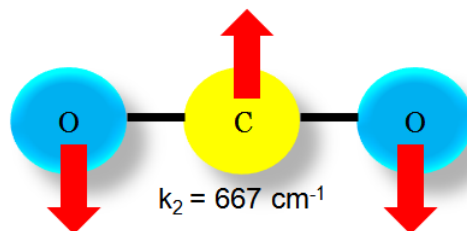


Figure 3-7: CO₂ Second Fundamental Vibrational Frequency, $k_2 = 667 \text{ cm}^{-1}$

The third fundamental vibrational mode for CO₂, k_3 , which is shown in Figure 3-8 results when the carbon atom moves relative to the center of mass of the oxygen atoms. The characteristic vibrational frequency associated with this motion is 2350 cm⁻¹.

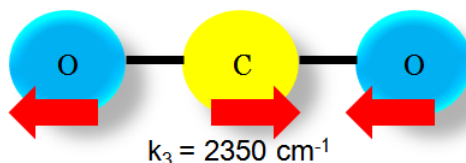


Figure 3-8: CO₂ Third Fundamental Vibrational Frequency, $k_3 = 2350 \text{ cm}^{-1}$

In regards to FTIR measurements, CO₂ has IR bands that will absorb so strongly that that it is possible to reach a concentration level where the energy transmitted to the detector will not produce a spectrum with features that are distinguishable from the noise level of the instrument. At this point, above 3 absorbance units (< 0.1% Transmission), the Beer-Lambert law is no longer applicable, and typical methods for concentration calculations are no long applicable. The CO₂ spectrum from the QASoft® database is shown in Figure 3-9 and from this 100 ppm spectra the maximum absorbance level is 0.65 absorbance units at the k_2 (667 cm⁻¹) characteristic frequency and 0.40 absorbance units at the k_3 (2350 cm⁻¹) characteristic frequency. Because of this strong absorption, the Beer-Lambert law is not applicable to monitor CO₂ concentrations above 460 ppm near the k_2 (667 cm⁻¹) characteristic frequency and 750 ppm near the k_3 (2350 cm⁻¹) characteristic frequency. It is standard industry practice to monitor CO₂ high concentrations within a working wavenumber window of detection of 2390-2379 cm⁻¹ to avoid significant absorption of the IR source.

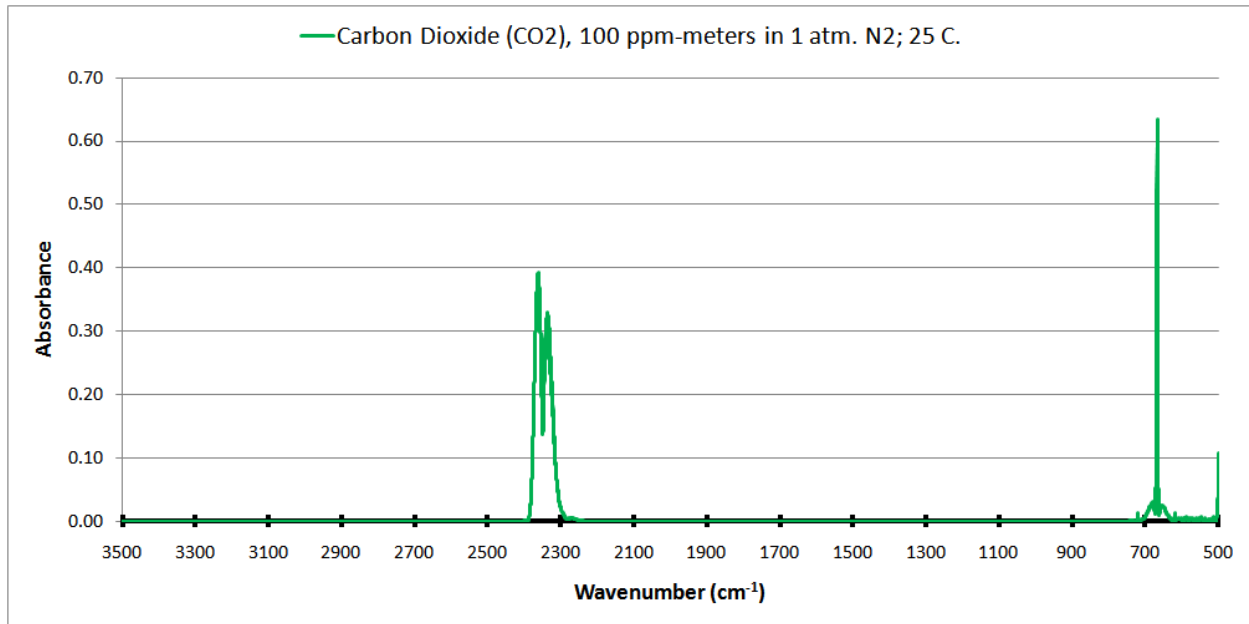


Figure 3-9: IR Absorbance Spectra for CO₂ from QASoft® Database

Water (H₂O)

H₂O is a heterogeneous, bent, tri-atomic molecule that has three characteristic modes of vibration ($3N-6 = 3 \times 3 - 6 = 3$) [31]. Figure 3-10 shows the H₂O second fundamental vibrational mode, k_2 that results when the hydrogen atoms bend their O-H bonds. The characteristic vibrational frequency associated with this motion is 1595 cm^{-1} . The first and third fundamental vibrational modes are outside the spectral window used for analysis in this research. This second fundamental vibrational mode, along with its associated rotational modes is a major source of interference for a number of characteristic IR spectra of interest, spanning from approximately 2000 cm^{-1} to 1300 cm^{-1} . The H₂O spectrum from the QASoft® database is shown in Figure 3-11.

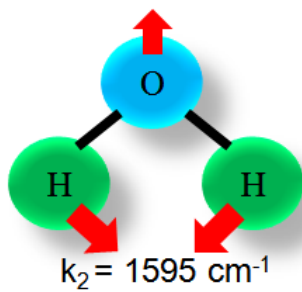


Figure 3-10: H₂O Second Fundamental Vibrational Frequency, $k_2 = 1595 \text{ cm}^{-1}$

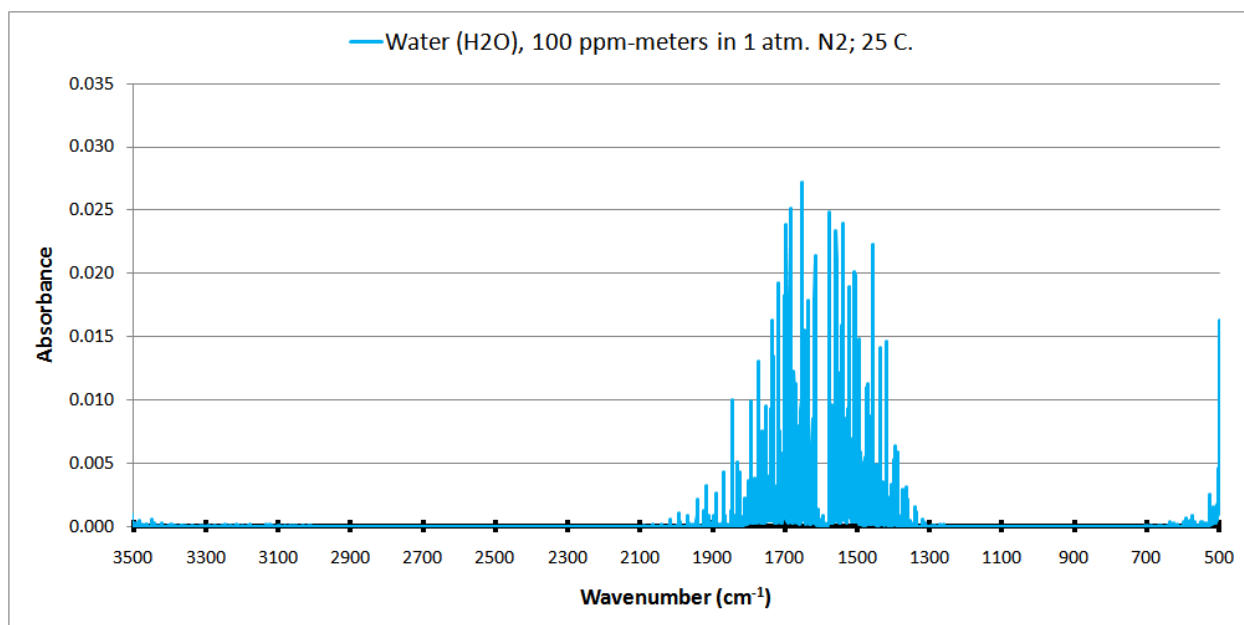


Figure 3-11: IR Absorbance Spectra for H₂O from QASoft® Database

Chapter 4: Principal Component Analysis (PCA)

PCA Theoretical Background

PCA is a technique that has been exploited quite extensively and successfully in the area of applied chemistry for a wide variety of functions, ranging from surface enhanced Raman scattering [37, 38], to X-ray photoelectron spectroscopy (XPS) [39], liquid chromatography [40-42], and FTIR spectroscopy [43-45]. This versatile technique allows for a large number of variables in a data set, such as absorbance values for given wavenumbers in the case of IR spectra, to be reduced to simple primary, i.e. principal, components. These principal components are orthogonal and have been shown to retain a significant amount of the original data set variation [46, 47]. A further principal component reduction process allows for the use of only the first few uncorrelated and ordered principal components for determining the simplified internal structure of the original data [43].

Typically, the data matrix used in PCA, $[X]_{(n \times p)}$, is a data set consisting of n samples taken at p measurements points. Using the singular value decomposition (SVD) theorem of matrix algebra, $[X]$ can be written as a product of three terms as shown in Equation 4.1, where the product $[U][L]$, is most commonly referred to as the scores matrix, $[S]_{(n \times p)}$, and $[V]_{(p \times p)}$ as the loadings matrix.

$$[X] = [U][\Lambda][V]^T \quad (4.1)$$

The first step in solving for [S] and [V] of the original data set is to calculate the mean adjusted data matrix, $[X_M]_{(n \times p)}$, by subtracting the columns means from each column component in [X]. Mean subtraction is necessary to ensure that the first principal component describes the direction of maximum variance instead of corresponding to the mean of the data [48]. From $[X_M]$, a variance-covariance matrix, $[Z]_{(p \times p)}$, can be constructed where the diagonals, Z_{ij} ($i = j$), represent the variance of the data points, *Var*, at a given wavenumber and the Z_{ij} ($i \neq j$) components represent the covariance, *Cov*, of particular wavenumbers among the samples as shown in Figure 4-1.

$$\begin{bmatrix} \text{Var}(X_{M11}) & \text{Cov}(X_{M12}) & \text{Cov}(X_{M13}) & \dots & \text{Cov}(X_{M1p}) \\ \text{Cov}(X_{M21}) & \text{Var}(X_{M22}) & \text{Cov}(X_{M23}) & \dots & \text{Cov}(X_{M2p}) \\ \text{Cov}(X_{M31}) & \text{Cov}(X_{M32}) & \text{Var}(X_{M33}) & \dots & \text{Cov}(X_{M3p}) \\ \dots & \dots & \dots & \dots & \dots \\ \text{Cov}(X_{Mp1}) & \text{Cov}(X_{Mp2}) & \text{Cov}(X_{Mp3}) & \dots & \text{Var}(X_{Mpp}) \end{bmatrix}$$

Figure 4-1: Variance-Covariance Matrix, $[Z]_{(p \times p)}$ Calculated from Mean Centered Data Matrix, $[X_M]_{(n \times p)}$

The magnitude of each eigenvalue, λ , in Equation 4.2 indicates the relative contribution of the corresponding eigenvector to the variance of the original data. In Equation 4.2, I is the identity matrix. The eigenvalues are arranged in order from largest to smallest and the measure of reconstruction accuracy, γ , is provided by the relative contribution of the retained eigenvalues to the sum of squares of eigenvalues as shown in Equation 4.3, where p^* is the number of retained eigenvalues.

$$|Z - \lambda I| = 0 \tag{4.2}$$

$$\gamma = \frac{\sum_k^{p^*} \lambda_k^2}{\sum_k^p \lambda_k^2} \quad (4.3)$$

From each of the ordered eigenvalues, corresponding eigenvectors or loadings matrix, $[V]_{(p \times p)}$, can be found by solving Equation 4.4 for each of the p eigenvalues from Equation 4.2. In Equation 4.4, $V_{i,j}$ corresponds to the eigenvector associated with the i^{th} eigenvalue. These eigenvectors are the coefficients required to transform the original variables into the principal component variable space.

$$\begin{bmatrix} Z_{11} - \lambda_i & Z_{12} & \dots & Z_{1p} \\ Z_{21} & Z_{22} - \lambda_i & \dots & Z_{2p} \\ \dots & \dots & \dots & \dots \\ Z_{p1} & Z_{p2} & \dots & Z_{pp} - \lambda_i \end{bmatrix} \begin{bmatrix} V_{i1} \\ V_{i2} \\ \dots \\ V_{ip} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (4.4)$$

From these eigenvectors, the individual elements of the new variable of $[S]$ are calculated from Equation 4.5.

$$[S]_{(n \times p)} = [X_M]_{(n \times p)} [V]_{(p \times p)} \quad (4.5)$$

The columns of $[V]$ are then arranged from largest to smallest and a cut-off for the number of retained or significant eigenvalues, p^* , is made when the cumulative variance explained from each of the p eigenvalues is within the experimental error associated with the measurement process, which is typically 5% for FTIR experimental data. The first p^* columns

of $[V]$ and $[S]$ are used as a reduced loadings matrix, $[V^*]_{(p \times p^*)}$, and reduced scores matrix, $[S^*]_{(n \times p^*)}$, respectively, for further calculations. The error matrix, $[E]$, associated with the use of the reduced loadings and reduced scores matrices can be calculated from Equation 4.6, where $[X^*]$ is the reconstructed data matrix defined in Equation 4.7.

$$[X]_{(n \times p)} = [X^*]_{(n \times p)} + [E]_{(n \times p)} \quad (4.6)$$

$$[X^*]_{(n \times p)} = [S^*]_{(n \times p^*)} [V^*]^T_{(p^* \times p)} + [X_M]_{(n \times p)} \quad (4.7)$$

This residual error is typically very small when the reconstruction accuracy, γ , of the original data explained is high (Equation 4.3).

Principal Component Regression (PCR)

PCR is a mathematical technique that determines component concentrations of a prediction data set based on multivariate regression of a calibration data set. In most multivariate regression techniques though, there are correlations within the set of variables on which the measured response is dependent, and these correlations add redundancy to the regression model that can cause numerical instability in estimating regression coefficients [48]. PCR is employed when the responses of one variable are dependent on a set of other variables as shown in Equation 4.8 for the mean adjusted values, where $[b]$ is the vector of estimates of regression coefficients to be determined [49]. The advantage of PCR over other multivariate regression models is that through PCA, the number of significant components has been determined and the analyzed variables are orthogonal, which by definition do not have correlations.

$$[Y_M] = a + [X_M][b] \quad (4.8)$$

The vector [b] is defined as the product of the reduced loading, [V*], and the y-loadings term, [q], shown in Equation 4.9. The [q] term is defined in Equation 4.10, where [D]_(p×p) is a diagonal matrix that has each diagonal element (*i = j*) equal to the inverse of the *i*th eigenvalue.

$$[b] = [V^*][q] \quad (4.9)$$

$$[q] = [D][S^*][Y] \quad (4.10)$$

With the values for [b] calculated, the constant *a* from Equation 4.8 can be found using Equation 4.11, where *Y_M* is the mean value of the dependent variable and *X_M* is the mean value of the independent variable at a given measurement.

$$a = Y_M - X_M[b] \quad (4.11)$$

With the regression analysis completed with a calibration data set, the regression analysis results can be applied to a prediction data set. In each case, root mean square error (RMSE) values can be computed to provide a measure of how well the PCR technique performs. RMSE is defined in Equation 4.12.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\text{predicted}_i - \text{actual}_i)^2}{n}} \quad (4.12)$$

Proportionality Constant Calculation (PCC)

For PCC, the total integrated intensity of an IR absorbance band, I_i , is assumed to depend linearly by a proportionality constant, k_i , on the amount of the component concentration, X_i , in a mixture as shown in Equation 4.13.

$$X_i = k_i I_i \quad (4.13)$$

The mixture of r components with their respected concentrations is defined in Equation 4.14, where X_T is the total amount (i.e. total volume) of the mixture analyzed.

$$X_T = X_1 + X_2 + \dots + X_r \quad (4.14)$$

Based on the linearity assumption, one can determine the proportionality constant for each of the individual components in a calibration data set in which the amount is known and the peaks associated with the particular component can be isolated from the peaks associated with the mixture. This isolation of peaks associated with a particular component is an ideal task for PCA that reduces mixtures down to principal components. As with PCR, for both the calibration and prediction data sets, calculation of RMSE using Equation 4.12 can provide a measure of how well the PCC technique performs.

With the proportionality constants known for each component, a relationship to calculate the prediction data set component concentrations can be derived. In some cases where the overall volume is not constant, the PCC technique can be expanded to perform the analysis on

volume and integrated area fractions instead of components as shown in Equation 4.15 for a 2-component system.

$$\%X_1 = \frac{X_1}{X_1 + X_2} = \frac{k_1 I_1}{k_1 I_1 + k_2 I_2} \quad (4.15)$$

Equation 4.15 can be further simplified to Equation 4.16 where the ratio of proportionality constants for the two components calculated from a calibration data set are utilized to determine the volume fraction of a component in prediction data set mixtures where the total volume of the system cannot be kept constant.

$$\%X_1 = \frac{X_1}{X_1 + X_2} = \frac{I_1}{I_1 + \left(\frac{k_2}{k_1}\right) I_2} \quad (4.16)$$

Application to FTIR Spectroscopy Data

In the case of FTIR spectroscopy data for CH₂O and C₃H₄O, a simplified data set that has overlap between the two components can be constructed to highlight the mathematics involved within PCA, PCR, and PCC. Figure 4-2 shows the complete spectra of CH₂O and C₃H₄O that the simplified spectra values are shown in Table 4-1 and displayed graphically in Figure 4-3 that consists of seven wavenumbers with their corresponding absorbance values.

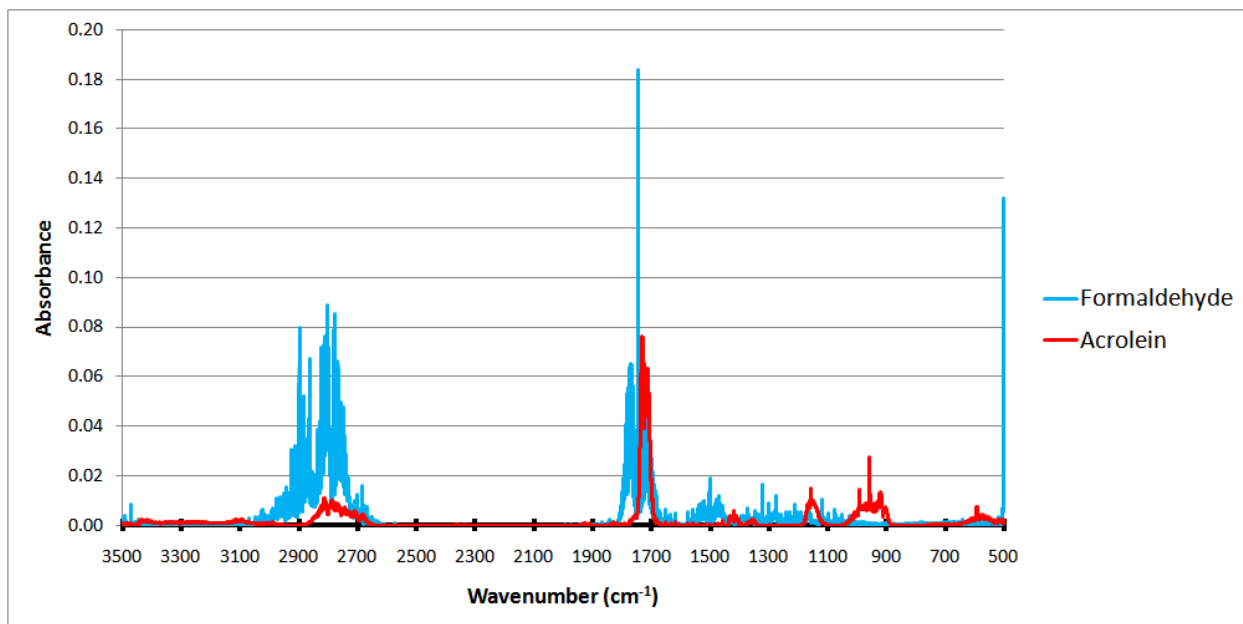


Figure 4-2: Formaldehyde (CH₂O) and Acrolein (C₃H₄O) Complete Pure Spectra

Table 4-1: CH₂O/C₃H₄O Simplified Pure Spectra for Illustration of PCA Application to FTIR Spectroscopy Data

		Wavenumber (cm ⁻¹)						
Component		2897	2863	2813	2810	2802	2791	2778
CH ₂ O	Absorbance	0.551	0.448	0.277	0.437	0.533	0.144	0.369
C ₃ H ₄ O	Absorbance	0.000	0.009	0.078	0.079	0.042	0.072	0.055

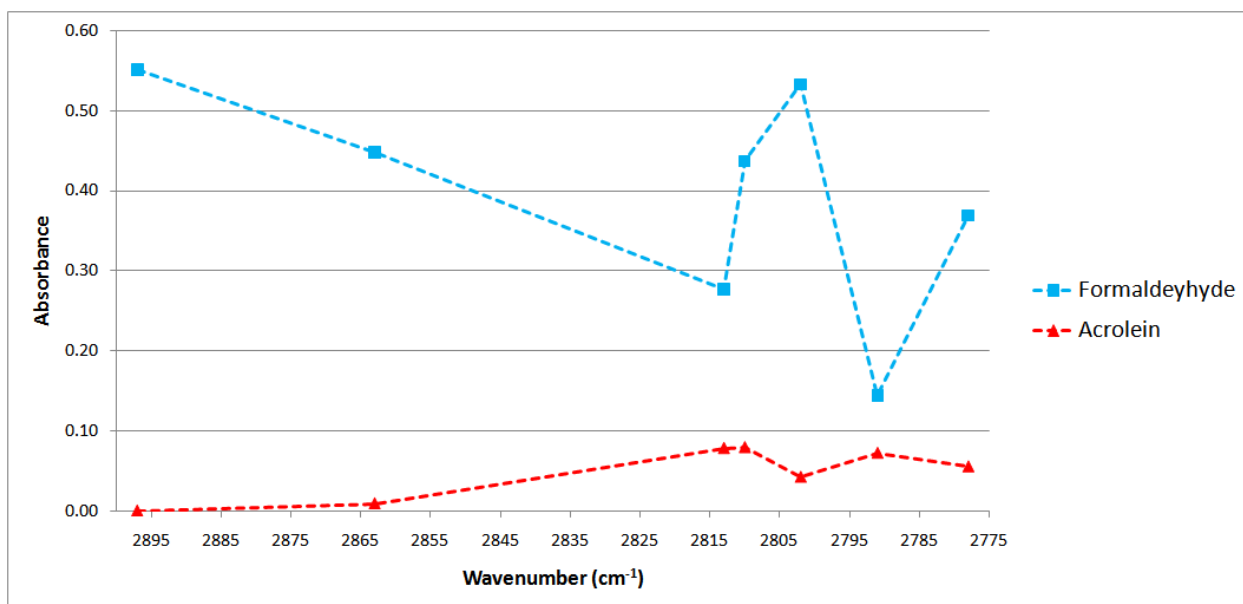


Figure 4-3: CH₂O/C₃H₄O Simplified Pure Spectra for Illustration of PCA Application to FTIR Spectroscopy Data

From these pure spectra, a set of 10 samples ($n = 10$) at 7 different wavenumbers ($p = 7$) were simulated with concentrations given in Table 4-2. Five spectra of the ten spectra used to produce a calibration data set $[X_C]_{(10 \times 7)}$ are shown in Figure 4-4. A set of five samples, with concentrations given in Table 4-3, were simulated as shown in Figure 4-5 to produce a prediction data set $[X_P]_{(10 \times 7)}$ for PCR. In the case of the PCC, a single data set $[X]_{(15 \times 7)}$ consisting of the calibration and prediction data sets were combined and PCA was simultaneously performed on all 15 samples to extract out the pure components in each simulated spectra. The mean centered data matrix is shown in Figure 4-6.

Table 4-2: CH₂O/C₃H₄O Spectra Compositions for Calibration Data Set, $[X_C]_{(10 \times 7)}$, Ordered from Lowest to Highest Amount of CH₂O in the Gas Mixture

Sample #	Amount of CH ₂ O (x100 ppm)	Amount of C ₃ H ₄ O (x100 ppm)
1	0.00	10.00
2	0.33	6.67
3	0.45	4.75
4	0.50	7.75
5	0.50	8.00
6	0.67	3.33
7	0.75	5.00
8	0.80	0.50
9	0.95	9.50
10	1.00	0.00

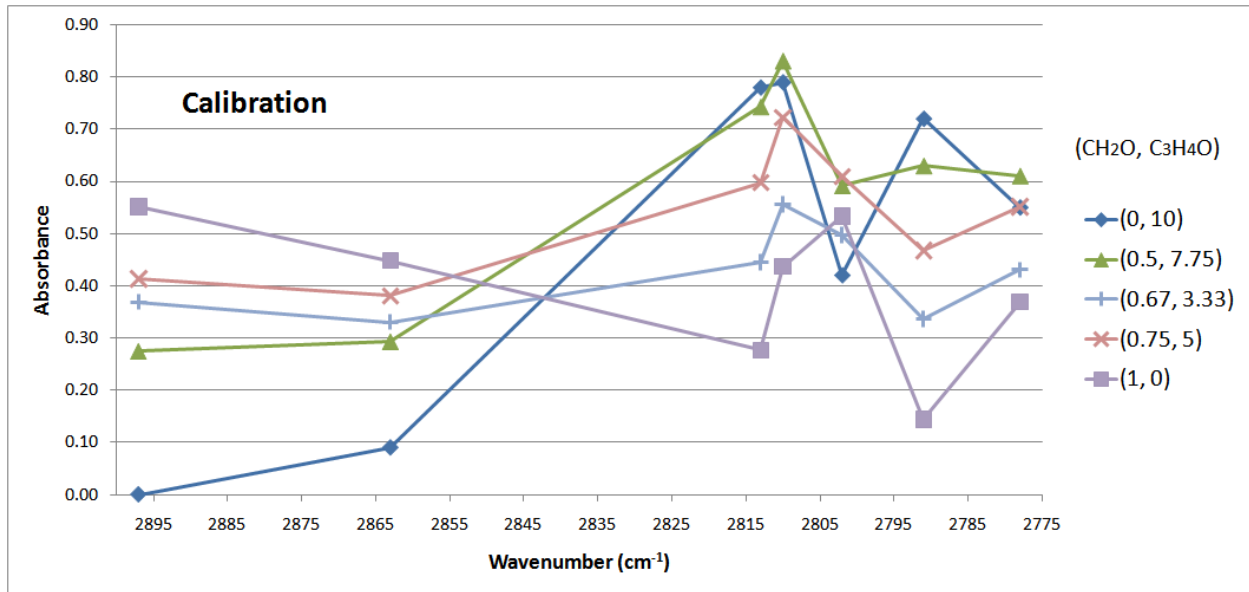


Figure 4-4: CH₂O/C₃H₄O Calibration Data Set, [X_C]_(10 x 7); Note – Only 5 of 10 Calibration Spectra Shown

Table 4-3: CH₂O/C₃H₄O Spectra Compositions for Prediction Data Set, [X_P]_(5 x 7), Ordered from Lowest to Highest Amount of CH₂O in the Gas Mixture

Sample #	Amount of CH ₂ O (x100 ppm)	Amount of C ₃ H ₄ O (x100 ppm)
11	0.17	4.25
12	0.25	8.75
13	0.40	9.75
14	0.60	6.00
15	0.70	1.25

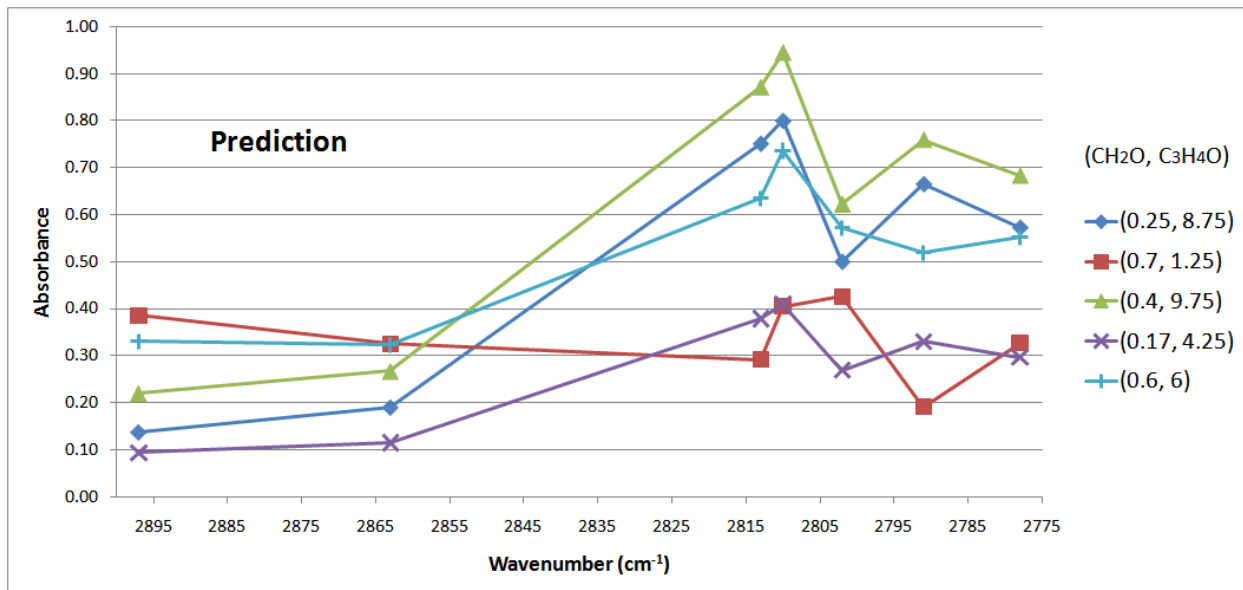


Figure 4-5: CH₂O/C₃H₄O Prediction Data Set, [X_P]_(5 x 7)

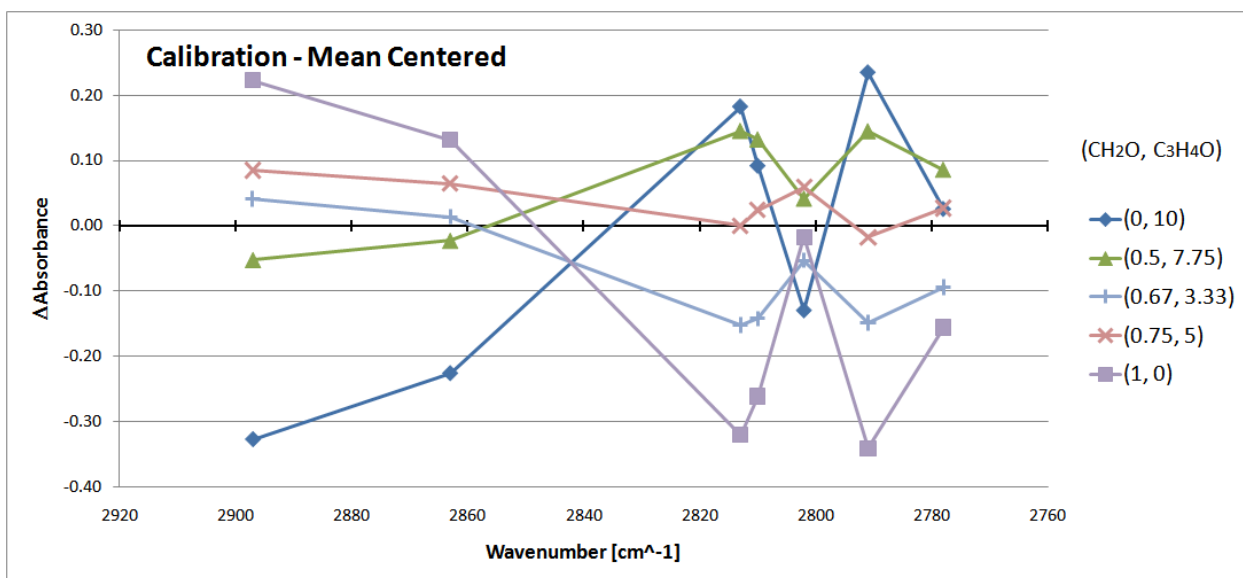


Figure 4-6: CH₂O/C₃H₄O Calibration Data Set, Mean Centered, [X_M]_(10 x 7); Note – Only 5 of 10 Calibration Spectra Shown

From the covariance-variance matrix shown in Figure 4-7, solving Equation 4.2 yields the eigenvalues, ordered from largest to smallest, shown in Figure 4-8. As shown in the plot of

eigenvalues as a function of principal components, Figure 4-9, the number of significant eigenvalues obtained from PCA is two. The first principal component explains 77.1% of the total calibration data set variance and the second principal component explains of 22.9% of the total calibration data set variance. With these two principal components, 100.0% of the original data variance can be explained. Figure 4-9 is commonly referred to as a SCREE plot that shows the eigenvalues as a function of each principal component.

$$\begin{bmatrix} 0.0278 & 0.0177 & -0.0106 & -0.0037 & 0.0177 & -0.0149 & 0.0004 \\ 0.0177 & 0.0145 & -0.0023 & 0.0027 & 0.0111 & -0.0057 & 0.0040 \\ -0.0106 & -0.0023 & 0.0552 & 0.0473 & 0.0194 & 0.0480 & 0.0317 \\ -0.0037 & 0.0027 & 0.0473 & 0.0521 & 0.0229 & 0.0444 & 0.0322 \\ 0.0117 & 0.0111 & 0.0194 & 0.0229 & 0.0207 & 0.0155 & 0.0174 \\ -0.0149 & -0.0057 & 0.0480 & 0.0444 & 0.0155 & 0.0526 & 0.0292 \\ 0.0004 & 0.0040 & 0.0317 & 0.0322 & 0.0174 & 0.0292 & 0.0250 \end{bmatrix}$$

Figure 4-7: CH₂O/C₃H₄O Variance-Covariance Matrix, [Z]_(7 x 7)

$$\begin{bmatrix} 1.91 \times 10^{-1} \\ 5.67 \times 10^{-2} \\ 1.31 \times 10^{-9} \\ 6.36 \times 10^{-10} \\ 6.70 \times 10^{-11} \\ 2.72 \times 10^{-11} \\ 0 \end{bmatrix}$$

Figure 4-8: CH₂O/C₃H₄O Eigenvalues from Solving Equation 4-2

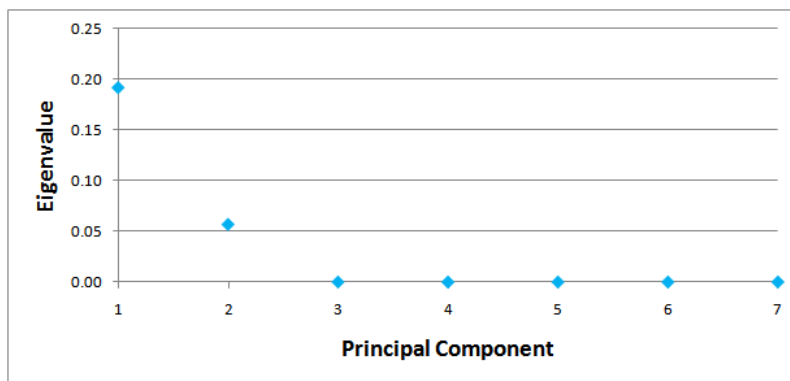


Figure 4-9: SCREE Plot Indicating Eigenvalues for Each Principal Component; Principal Components 1 and 2 Explain 77.1% and 22.9% of the Total Calibration Data Set Variance

Next, Equation 4.4 can be solved to find the eigenvectors corresponding to the calculated eigenvalues. This yields the loadings matrix, [V] shown in Figure 4-10. With [V], Equation 4.5 can be solved to determine the scores matrix, [S], which is shown in Figure 4-11. Since the number of significant principal components was calculated to be two, only the first two columns of both [V] and [S] are needed to represent the 100% of the variance found in the calibration data set for [V*] and [S*]. The residual error from using two principal components is found by solving Equation 4.6 for [E], which for this simulated data set is for all practical purposes zero.

$$\begin{bmatrix}
 0.0842 & 0.6834 & 0.0397 & -0.3288 & -0.2144 & -0.4439 & 0.4175 \\
 0.0018 & 0.5065 & -0.5853 & 0.5681 & 0.1528 & 0.2341 & 0 \\
 -0.5355 & -0.0821 & -0.3906 & -0.4629 & 0.5595 & -0.0366 & 0.1593 \\
 -0.5184 & 0.1109 & 0.5267 & 0.2983 & 0.0194 & 0.3442 & 0.4835 \\
 -0.2297 & 0.4319 & 0.1328 & -0.3900 & -0.1913 & 0.4707 & -0.5769 \\
 -0.5113 & -0.2143 & -0.3667 & 0.0805 & -0.7224 & -0.1729 & 0.0027 \\
 -0.3510 & 0.1576 & 0.2717 & 0.3276 & 0.2454 & -0.6139 & -0.4835
 \end{bmatrix}$$

Figure 4-10: CH₂O/C₃H₄O Loadings Matrix, [V]_(7x7)

-0.2720	-0.4462	0	0	0	0	0	0	0	0
0.6159	0.0894	0	0	0	0	0	0	0	0
-0.2646	-0.0440	0	0.0001	0	0	0	0	0	0
0.0068	-0.2146	-0.0001	0	0	0	0	0	0	0
-0.8184	0.3871	0	0	0	0	0	0	0	0
0.2135	-0.1432	0	0	0	0	0	0	0	0
0.2806	0.0259	0.0001	0	0	0	0	0	0	0
-0.0197	0.1274	0	0	0	0	0	0	0	0
-0.3017	-0.0394	0	0	0	0	0	0	0	0
0.5595	0.2576	0	0	0	0	0	0	0	0

Figure 4-11: CH₂O/C₃H₄O Scores Matrix, [S]_(10 x 10)

With [V*] and [S*], PCR can be conducted and the vector of estimates of regression coefficients, [b], shown in Figure 4-12 can be found by solving Equation 4.9. The intercept, *a*, can be found by solving Equation 4.11 and for this example is calculated to be -4.2×10^{-4} . Applying this analysis, a plot can be created that shows the calibration data set used to determine the concentrations in the prediction data as shown in Figure 4-13. With the PCR technique, the RMSE (Equation 4.12) for both the calibration set (RMSEC) and for the prediction data set (RMSEP) is found to be 1.1×10^{-4} ppm.

0.8606
0.6291
-0.1839
0.0588
0.5008
-0.3443
0.1418

Figure 4-12: CH₂O/C₃H₄O Estimates of Regression Coefficients from Calibration Data Set, [b]_(7 x 1)

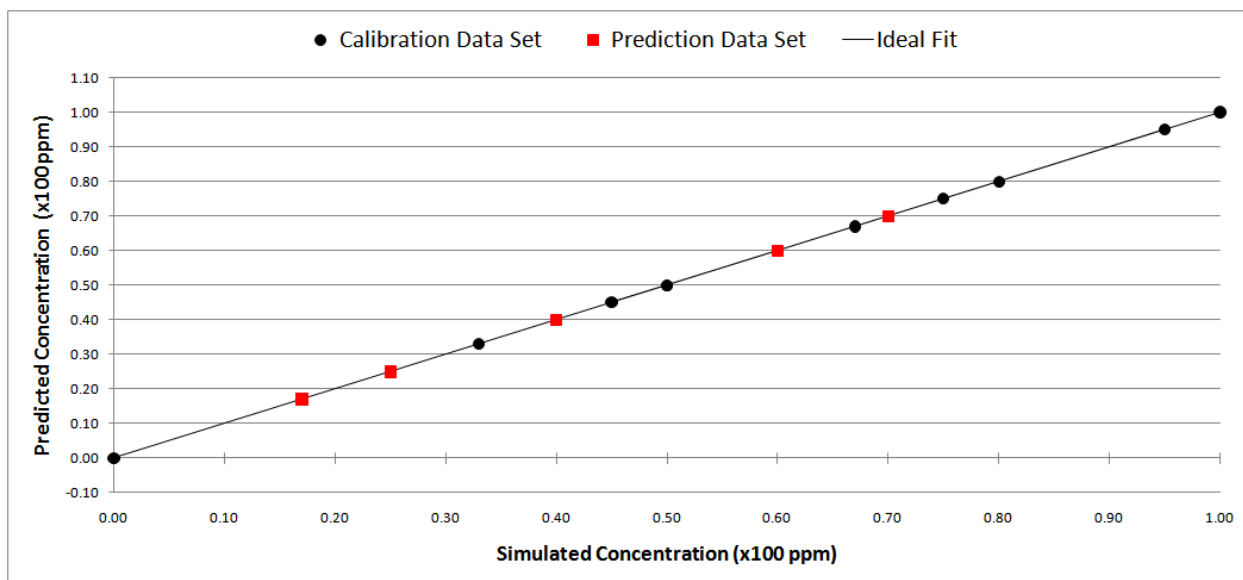


Figure 4-13: Principal Component Regression (PCR) for CH₂O Concentrations in CH₂O/C₃H₄O Gas Mixtures;
 RMSE Calibration = 1.1×10^{-4} ppm, RMSE Prediction = 1.1×10^{-4} ppm

PCA on the entire set of collected spectra, both calibration and prediction sets, is capable of isolating the peaks associated with each particular component as shown in Figures 4-14 and 4-15. Since the principal component loadings, *V-1* and *V-2*, are abstract representations of information within the original data set, it is acceptable and in most cases unavoidable to have negative elements due to the orthogonality requirement of PCA. The benefit of Figures 4-14 and 4-15 are that *V-1* and *V-2* can be easily identifiable as corresponding to the pure spectra of C₃H₄O and CH₂O, respectively.

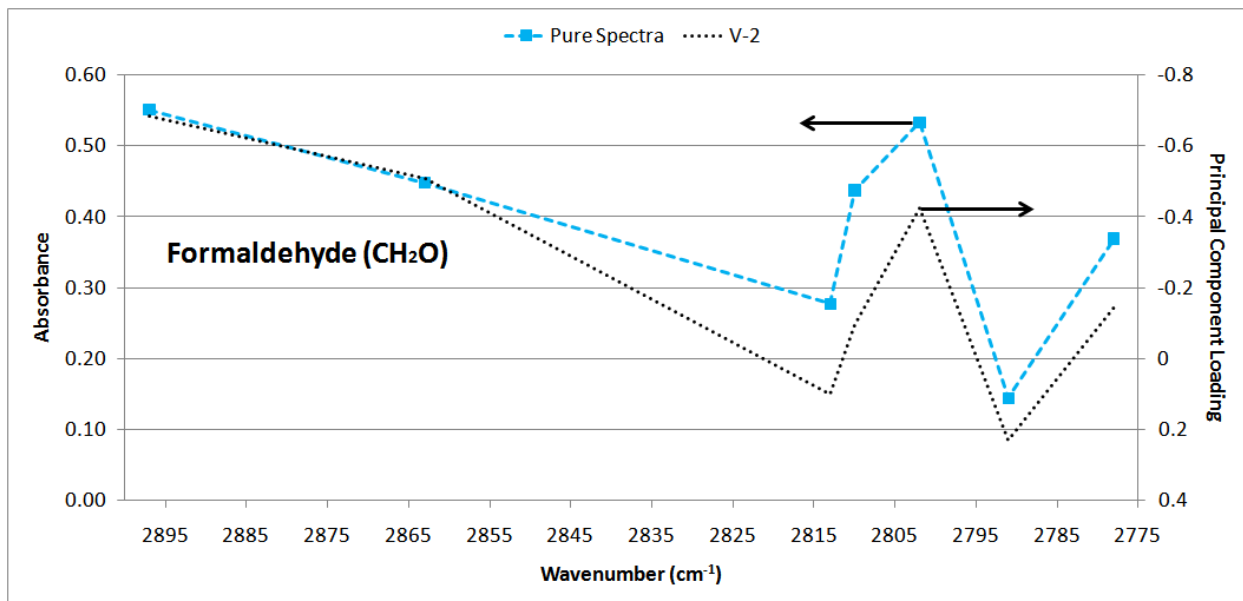


Figure 4-14: PCA Separated CH₂O Spectra in CH₂O/C₃H₄O Gas Mixtures

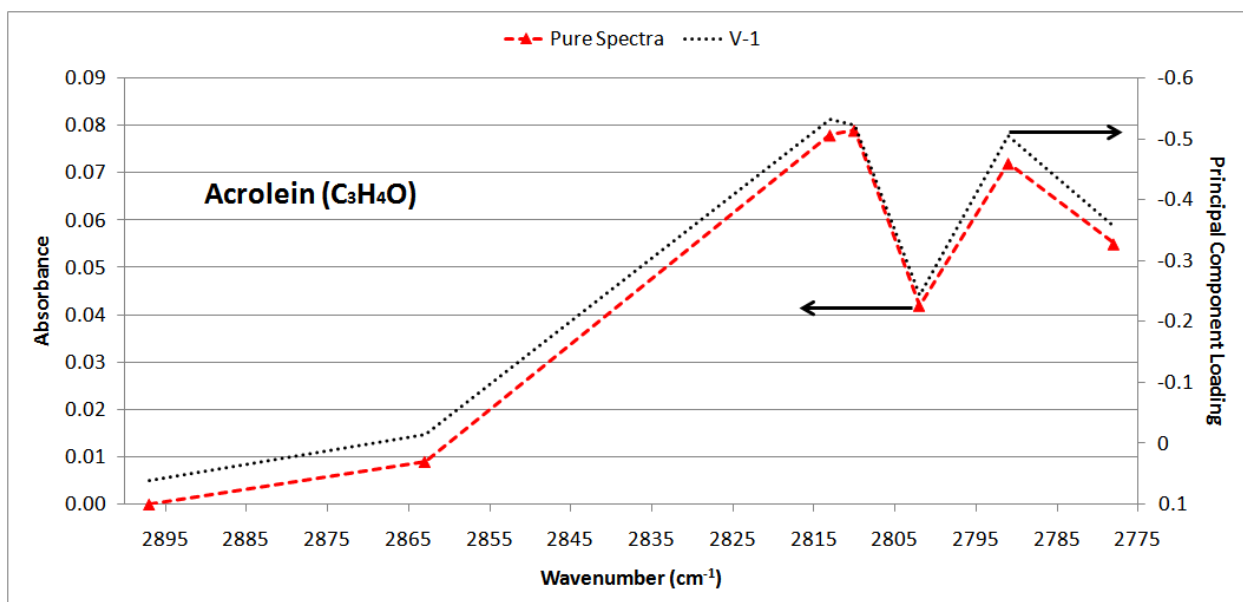


Figure 4-15: PCA Separated C₃H₄O Spectra in CH₂O/C₃H₄O Gas Mixtures

With the components separated, the contribution of each component to the sample mixtures can be determined as shown as in Figures 4-16 and 4-17 for the C₃H₄O calibration and

prediction data sets, respectively, and Figures 4-18 and 4-19 for CH₂O calibration and prediction data set, respectively.

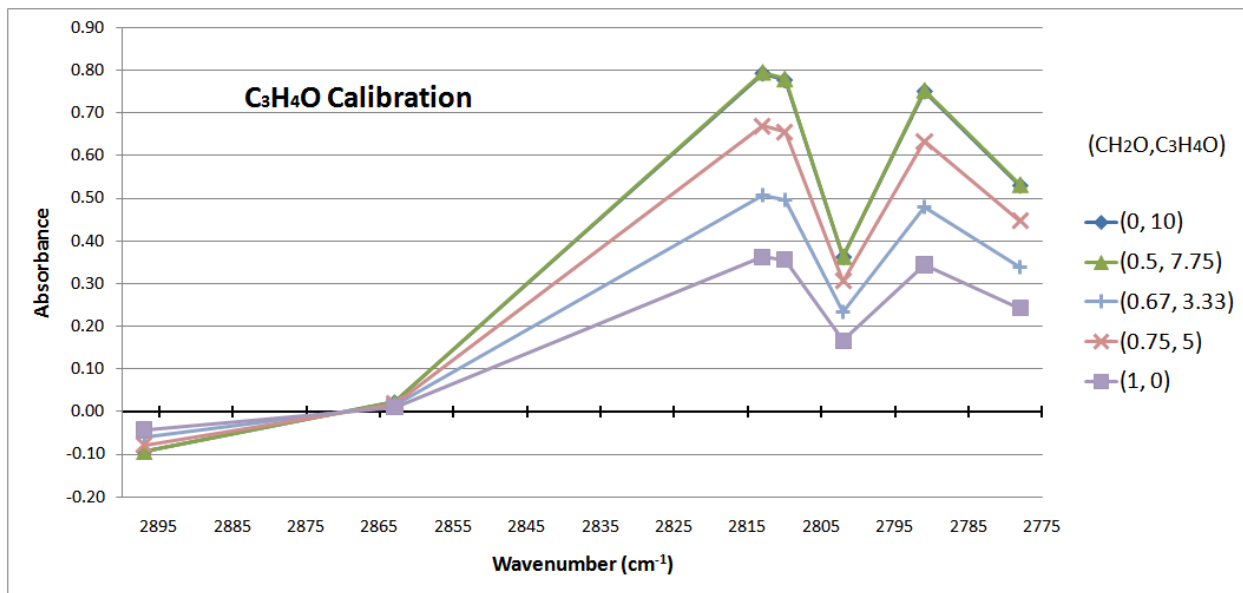


Figure 4-16: PCC C₃H₄O Calibration Data Set; Note – Only 5 of 10 Calibration Spectra Shown

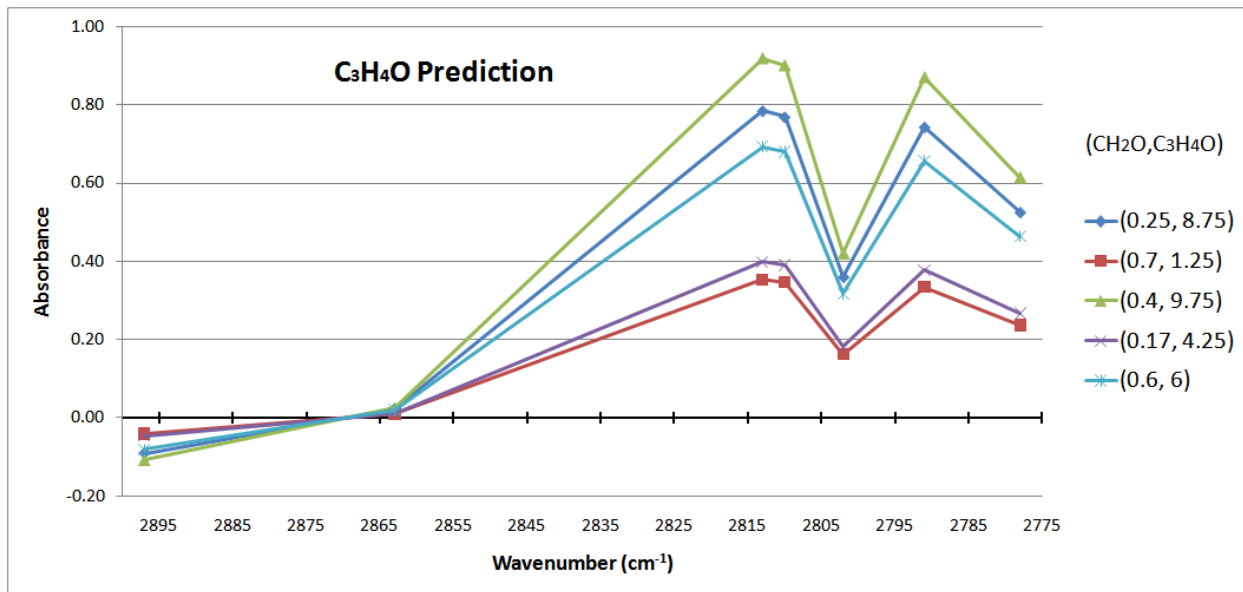


Figure 4-17: PCC C₃H₄O Prediction Data Set

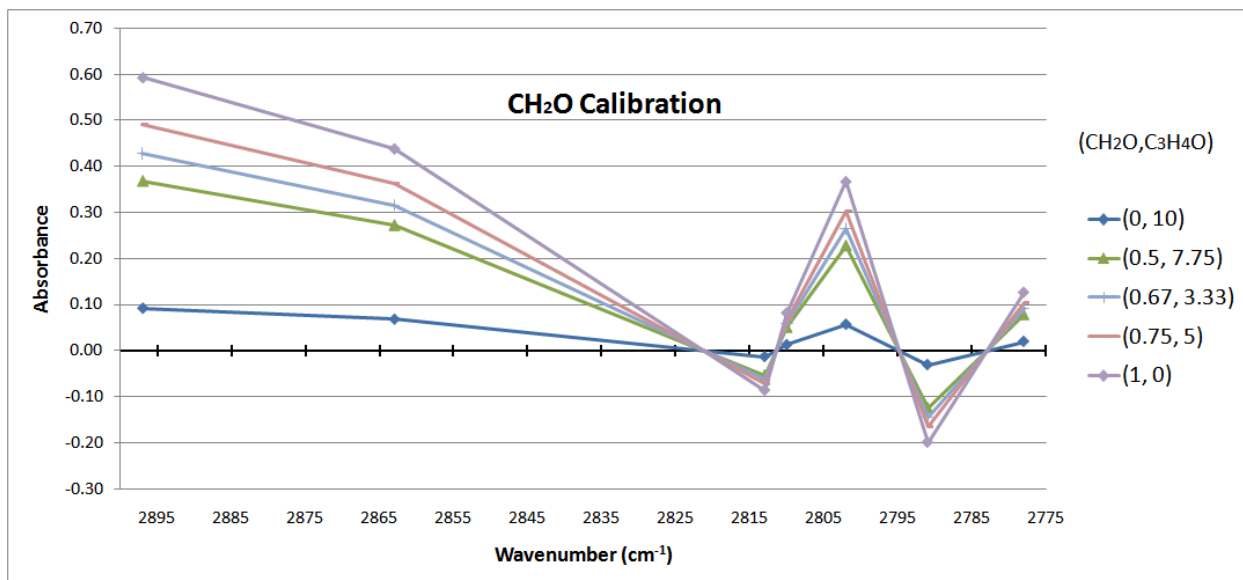


Figure 4-18: PCC CH₂O Calibration Data Set; Note – Only 5 of 10 Calibration Spectra Shown

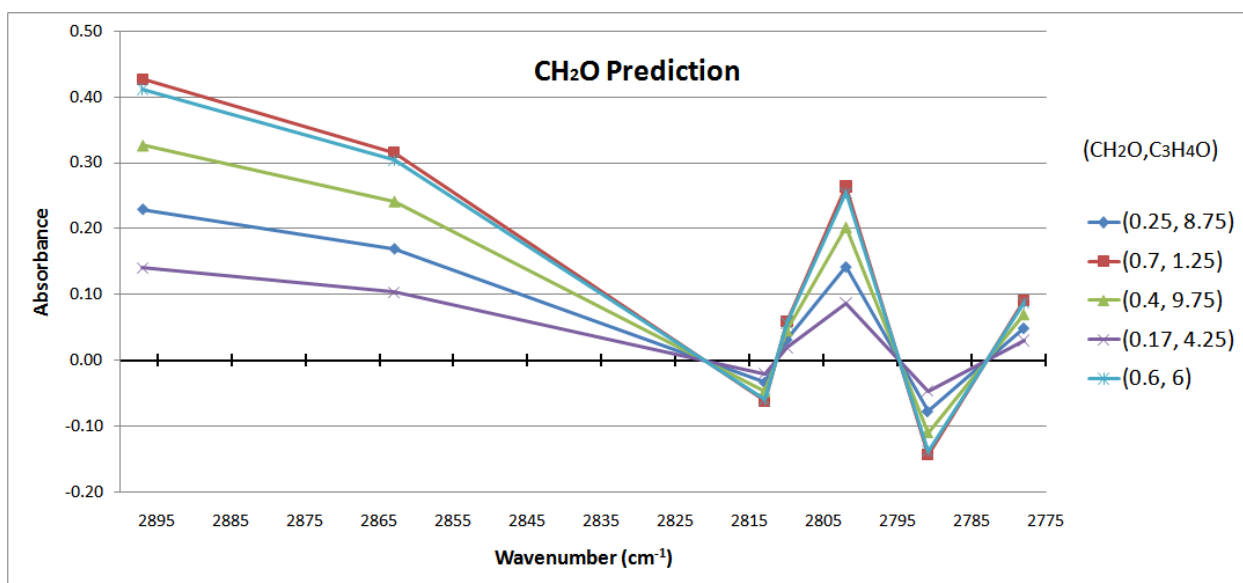


Figure 4-19: PCC CH₂O Prediction Data Set

As shown in Figure 4-20 for CH₂O, a baseline correction is necessary to remove the generated noise into the data set due to the mathematics in the PCA technique. With the baseline correction made (Figure 4-21), and based on the linearity assumption, the proportionality

constant can be determined from Equation 4.13 for each of the individual components in the calibration data set from the isolated peaks. This calculated proportionality constant, equal to the inverse of the slope of the line of best fit shown in Figure 4-21, can then be used to calculate the concentration of the mixtures in the prediction data set that are shown in Figure 4-22.

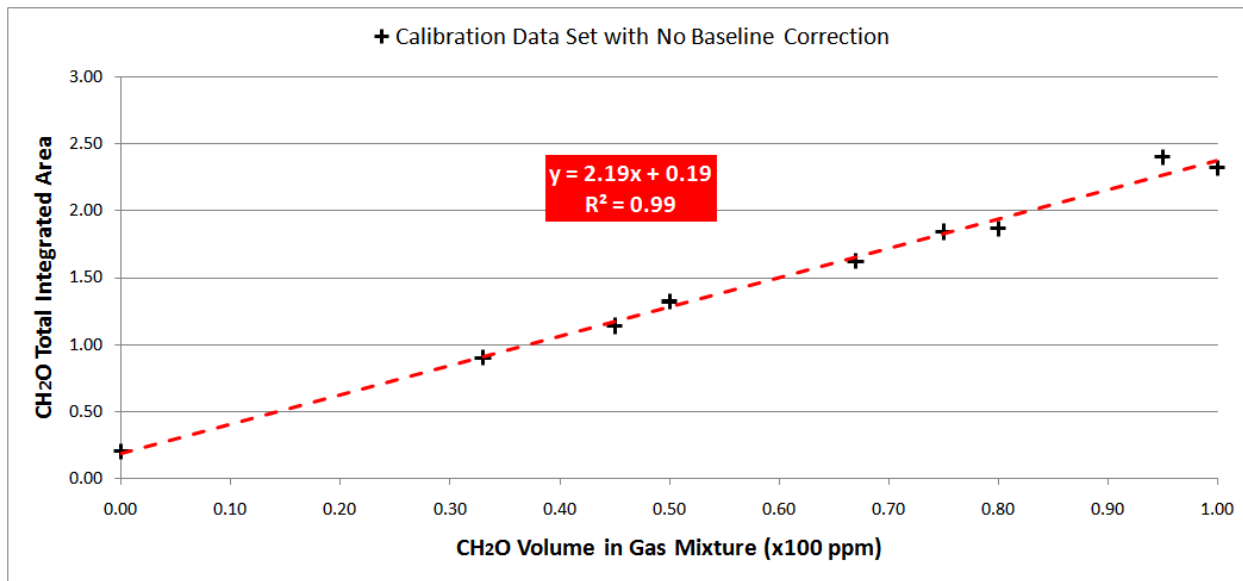


Figure 4-20: PCC CH₂O Calibration Data Set with No Baseline Correction

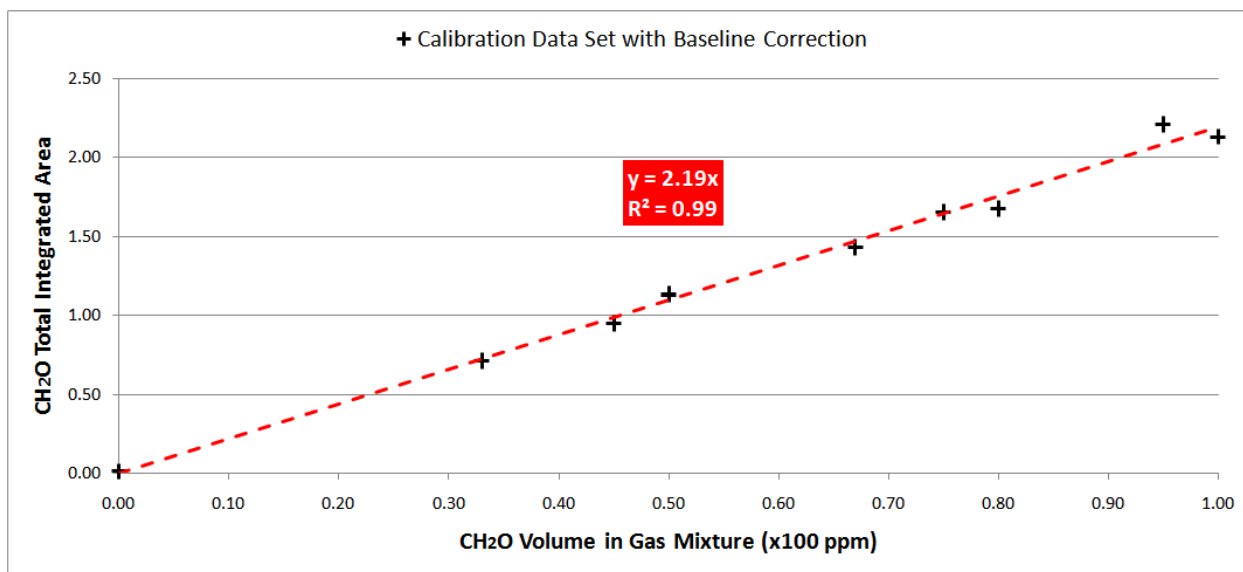


Figure 4-21: PCC Calibration Data Set with Baseline Correction

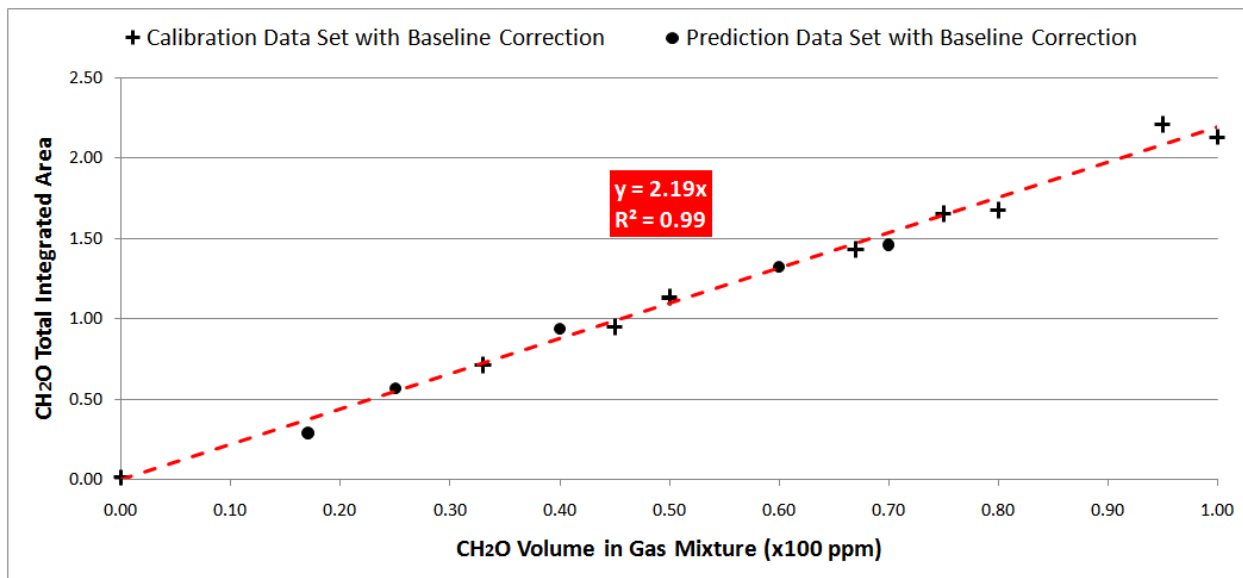


Figure 4-22: PCC CH₂O Prediction Data Set; RMSE Calibration = 2.6×10^{-2} ppm, RMSE Prediction = 2.6×10^{-2} ppm

The comparison of errors associated with both the PCR and PCC techniques shown in Table 4-4 indicates that the PCC technique could be a viable alternative quantitative analysis method to the PCR technique. The PCC technique in most cases provides a more tangible solution method that directly relates total area under an absorbance as a function of wavenumber curve proportionally to the volume of the component in the mixture, while the PCR technique is solely a mathematical multivariate regression technique. In addition, the PCC technique can be further expanded, as will be shown in Chapter 5, to analyze data sets that do not contain a constant volume of the mixtures across the entire sample space and outperform the PCR technique in terms of RMSEC and RMSEP.

Table 4-4: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques

Analysis Technique	RMSE Calibration	RMSE Prediction
PCR	1.1×10^{-4}	1.1×10^{-4}
PCC	2.6×10^{-2}	2.6×10^{-2}

Chapter 5: PCA Application to FTIR Spectroscopy Data of Vapor Phase Hydrogen Peroxide (VPHP) Aircraft Cabin Decontamination Events

Discussion of Analyzed Data Sets

Data set 1 was comprised of experimentally obtained IR spectra from H₂O₂ aqueous solution mixtures. Eight concentrations for the calibration data set were prepared by diluting a 35 wt.% H₂O₂ solution with de-ionized water at approximately 5 wt.% intervals. The H₂O₂ aqueous solution mixture samples for calibration in data set 1 were made from the solution in the Steris 1000ED Bio-decontamination Unit (Mentor, OH, USA) that uses VAPROX® (35 wt.% H₂O₂) as the sterilant. A single sample, for the prediction data set, was taken from an experimental VPHP run where conditions were known to produce significant and observable condensation of the H₂O₂ within a sample chamber. The IR spectra for data set 1 was obtained using the FTIR spectrometer over a wavenumber scan range from 2000 cm⁻¹ to 1200 cm⁻¹ with a spectral resolution of 4 cm⁻¹.

Data set 2 was comprised of a calibration data set with 15 spectra of H₂O₂ concentrations in aqueous solutions prepared by diluting a 70 wt.% H₂O₂ solution offered by Armeka Canada Inc. with de-ionized water at approximately 5 wt. % intervals thus giving a calibration set that spans 0-70 wt.% H₂O₂ in solution. The prediction data set for the second set of H₂O₂ experiments consisted of 5 samples taken from different experimental VPHP runs using the Steris 1000ED Bio-decontamination Unit that produced condensation of the H₂O₂ within a

sample chamber. The IR spectra for data set 2 was obtained using the FTIR spectrometer over a wavenumber scan range from 1800 cm^{-1} to 1200 cm^{-1} with a spectral resolution of 4 cm^{-1} .

The purpose of data set 2 was to confirm the findings from data set 1. Data set 1 consisted of a restricted range for the calibration data set because the Steris 1000ED Bio-decontamination Unit operates with 35 wt.% H_2O_2 in aqueous solution. Use of this 35 wt.% H_2O_2 solution though is capable of producing condensates that are much higher in H_2O_2 concentration depending on the particular operating conditions as discussed in [24]. Data set 2 contains a more complete calibration data set that may not always be available in engineering applications and shows that the performance of the PCA technique in regards to data set 1 is sufficient when a complete calibration data is not feasible or possible to obtain. Sample collection for the prediction data sets 1 and 2 was performed by Mobbassar Hassan Sk (Ph.D. Graduate Student, Materials Engineering, Auburn University).

To confirm the PCR and PCC techniques for prediction in each of the data sets, a titration process (performed by Mobbassar Hassan Sk) on the prediction data sets for the H_2O_2 aqueous solutions [50]. First, a 5 N aqueous H_2SO_4 solution, where N is the number of protons (H^+) in a molecule of the acid, was made using 36 N H_2SO_4 (Fisher Scientific, Pittsburg, PA, USA, lot no. 094134) and a 0.05 M (moles/L) KMnO_4 solution was made from solid KMnO_4 (Fisher Scientific). Next, 50 ml of the 5 N aqueous H_2SO_4 solution was taken in a flask and an exactly weighed sample of liquid H_2O_2 was added to it followed by thorough mixing. Then, the 0.05 M KMnO_4 solution was taken in a burette and added to the solution mixture drop wise with constant stirring of the solution mixture. The end of titration process was easily identifiable by the permanent change of color of the solution mixture into pale pink. Once this point was

reached, the volume (measured in mL) of KMnO_4 solution consumed in the titration process was noted. The weight percentage of H_2O_2 in the solution was calculated using Equation 5.1.

$$\text{wt.\%H}_2\text{O}_2 = \frac{\text{KMNO}_4(\text{strength} = 0.05\text{M}) \text{ consumed [mL]} \times 0.425175}{\text{sample}(\text{H}_2\text{O}_2\text{solution})\text{wt}} \quad (5.1)$$

PCA Results and Discussions

Data Set 1

From the experimental FTIR spectral analysis of the H_2O_2 in aqueous solution mixture, a set of 8 samples ($n = 8$) at 801 different wavenumbers ($p = 801$) were collected as shown in Figure 5-1 to produce a calibration data set $[\text{X}_\text{C}]_{(8 \times 801)}$. A single sample was collected as shown in Figure 5-2 to produce a prediction data set $[\text{X}_\text{P}]_{(1 \times 801)}$ for PCR. In the case of the PCC, a single data set $[\text{X}]_{(9 \times 801)}$ consisting of the calibration and prediction data sets were combined and PCA was simultaneously performed on all 9 samples to extract out the pure components in each spectra.

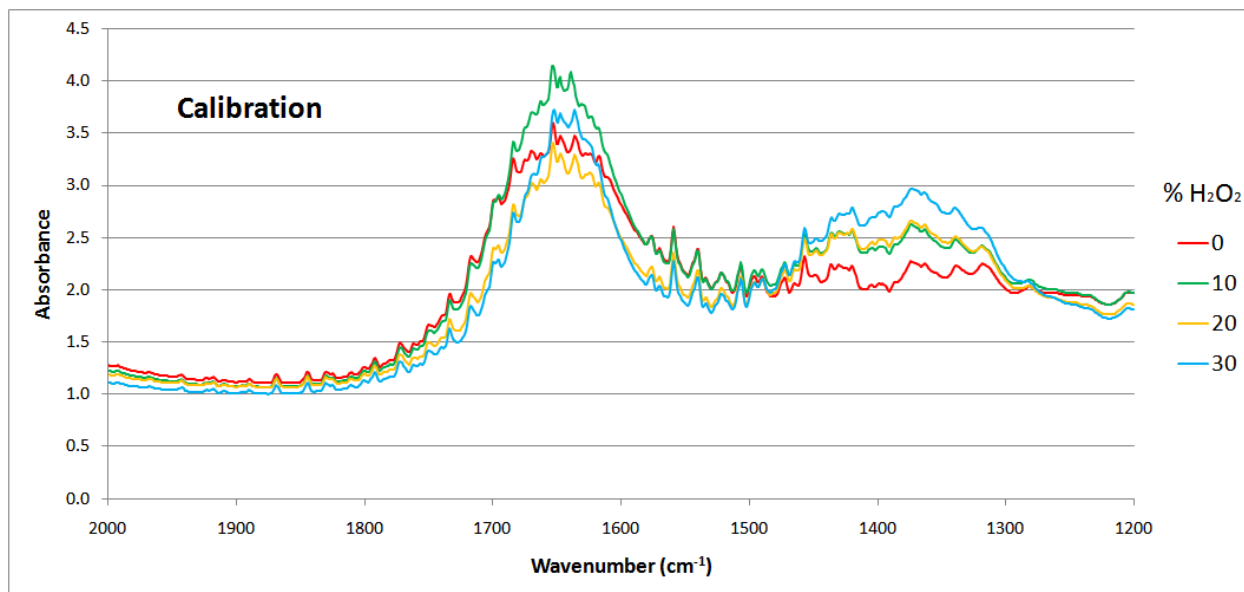


Figure 5-1: H₂O₂ in Aqueous Solution Calibration Data Set 1, [X_C]_(8 x 801); Note – Only 4 of 8 Calibration Spectra Shown, 0%, 10%, 20%, 30% H₂O₂

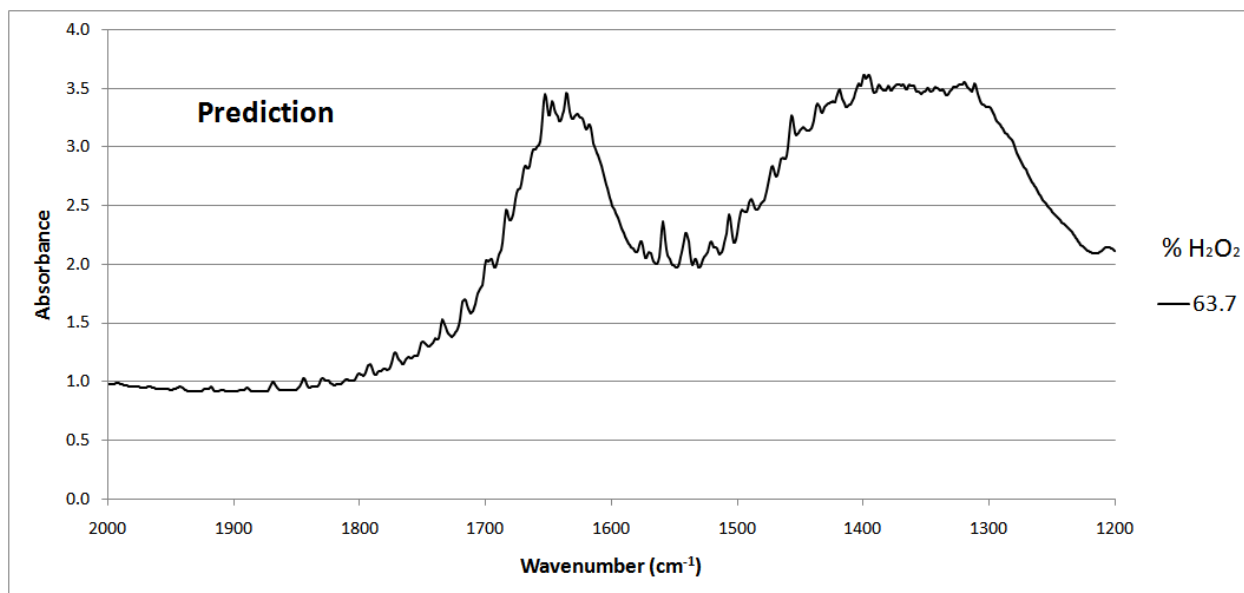


Figure 5-2: H₂O₂ in Aqueous Solution Prediction Data Set 1, [X_P]_(1 x 801); 63.7% H₂O₂

As shown in the plot of eigenvalues as a function of principal components (Figure 5-3), the number of significant eigenvalues obtained from PCA is two. The first principal component

explains 67.5% of the total calibration data set variance and the second principal component explains 28.5% of the total calibration data set variance. With these two principal components, 96.0% of the original data variance can be explained.

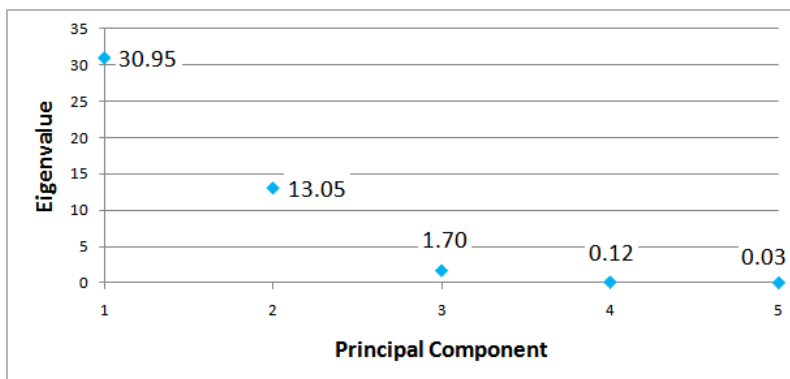


Figure 5-3: SCREE Plot Indicating Eigenvalues for Each Principal Component for H₂O₂ in Aqueous Solution Data Set 1; Principal Components 1 and 2 Explain 67.5% and 28.5% of the Total Calibration Data Set Variance

Next, Equation 4.4 can be solved to find the eigenvectors corresponding to the calculated eigenvalues and this yields the loadings matrix, [V]. With [V], Equation 4.5 can be solved to determine the scores matrix, [S]. Since the number of significant principal components was calculated to be two, only the first two columns of both [V] and [S] are needed to represent the 96.0% of the variance found in the calibration data set for [V*] and [S*]. Figure 5-4 shows the plot of the reduced loadings matrix [V*] for the first two principal components. V-1 represents the variable in the original data set contributing the most variance within the spectra, the H₂O component and V-2 represents the variable in the original data set contributing the second most variance, the H₂O₂ component. V-2 produces two partitions, with the positive loadings representing the bands for H₂O₂ and the negative loadings representing the bands for H₂O.

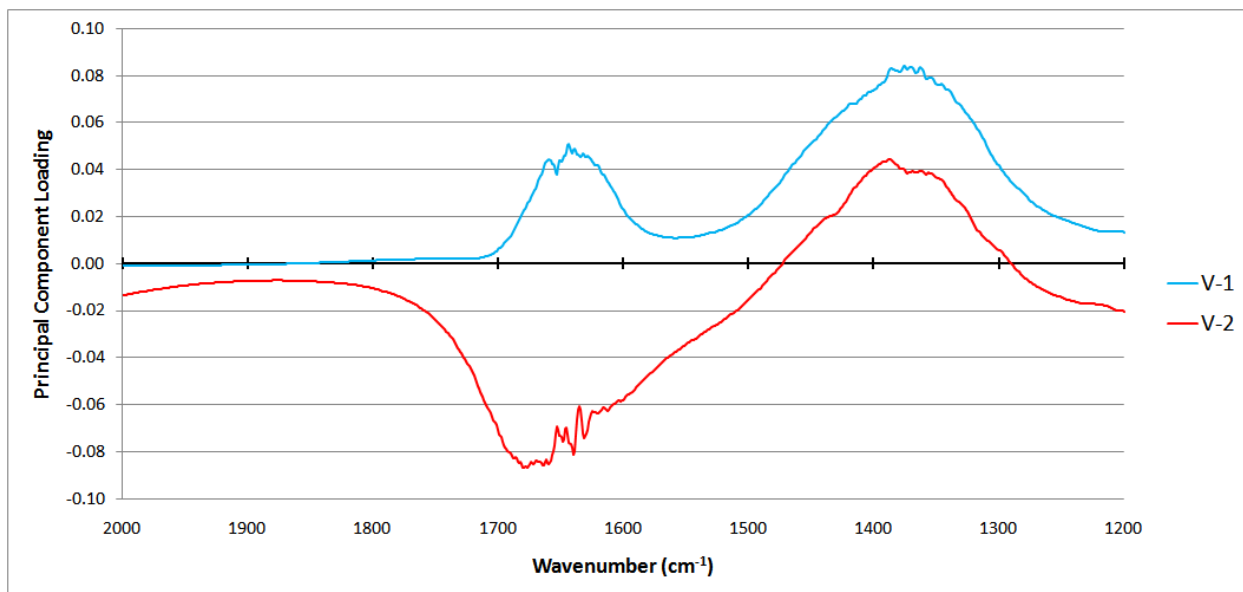


Figure 5-4: H₂O₂ in Aqueous Solution Mixtures – Reduced Principal Component Loadings, [V*]; V-1 represents the variable in the original data set contributing the most variance within the spectra, the H₂O component, V-2 represents the variable in the original data set contributing the second most variance, the H₂O₂ component

With [V*] and [S*], PCR can be conducted and the vector of estimates of regression coefficients for wt. % H₂O₂, [b_{H₂O₂}], shown graphically by wavenumber in Figure 5-5, can be found by solving Equation 4.9. The intercept, $a_{H_2O_2}$, can be found by solving Equation 4.11 and for H₂O₂ calibration data set 1 it is calculated to be 12.7. Applying this analysis, a plot can be created that shows the calibration data set used to determine the concentrations in the prediction data set as shown in Figure 5-6. With the PCR technique, the RMSE (Equation 4.12) for the calibration data set 1 is found to be 2.1 wt.% of H₂O₂, while the RMSE for the prediction data set 1 is found to be 12.0 wt.% of H₂O₂.

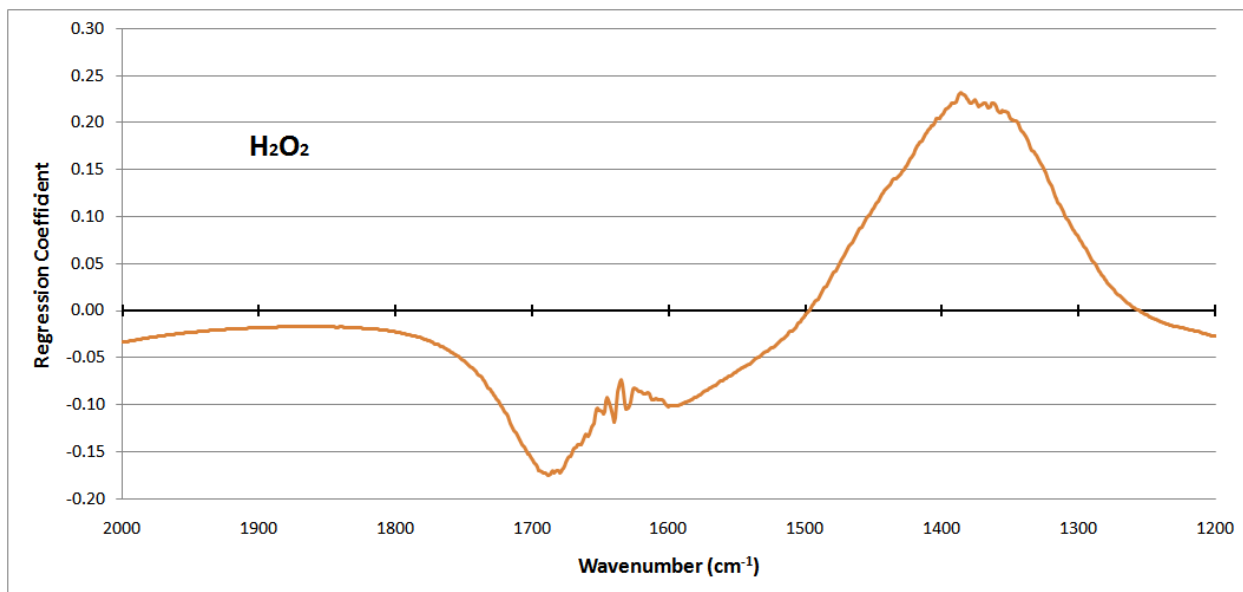


Figure 5-5: H_2O_2 in Aqueous Solution Mixtures – Estimates of Regression Coefficients for wt. % of H_2O_2 from Calibration Data Set 1 as Function of Wavenumber, $[b_{H_2O_2}]_{(801 \times 1)}$

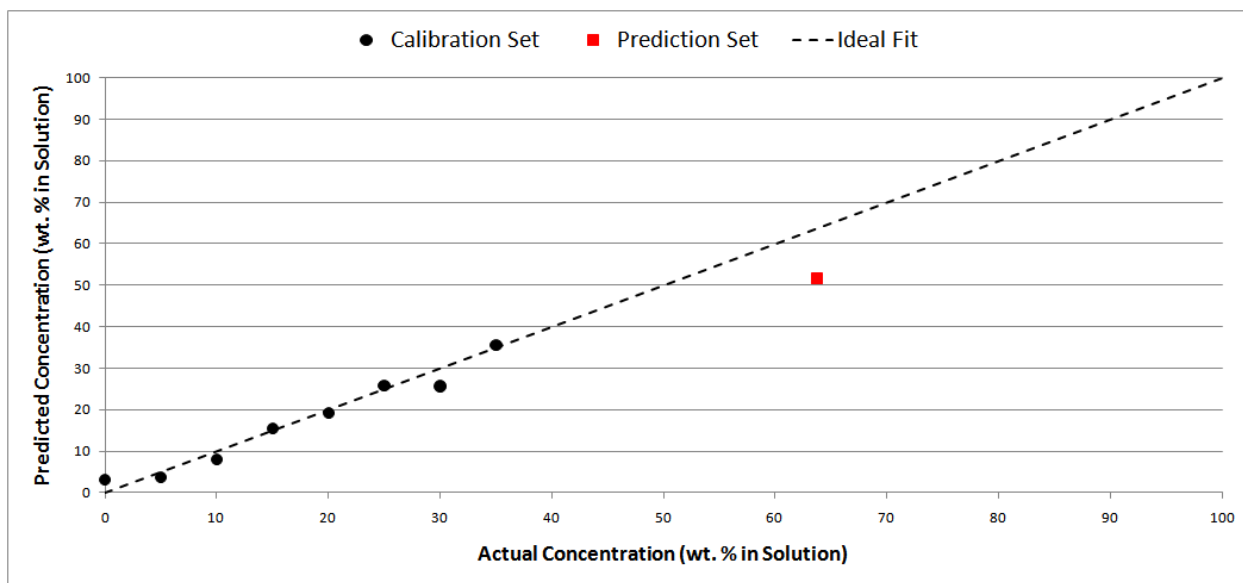


Figure 5-6: Principal Component Regression (PCR) – H_2O_2 Concentrations for H_2O_2 in Aqueous Solution Mixtures; RMSE Calibration = 2.1 wt.%, RMSE Prediction = 12.0 wt.%

In some cases, to improve the PCR analysis for H_2O_2 the analysis can utilize more of the identified principal components. This is not always possible because in some cases, the representation of redundant data can cause numerical instability within the regression analysis. To show that for this experimental data the PCR technique is not dramatically improved and can actually become worse with the use of more principal components, Figure 5-7 was produced that shows both RMSEC and RMSEP as a function of increasing number of principal components used to represent the original data set. This analysis indicates that no significant improvements can be made in terms of RMSEC for the H_2O_2 data set with the use of more principal components than was calculated to be necessary to explain 96.0% of the original data variance.

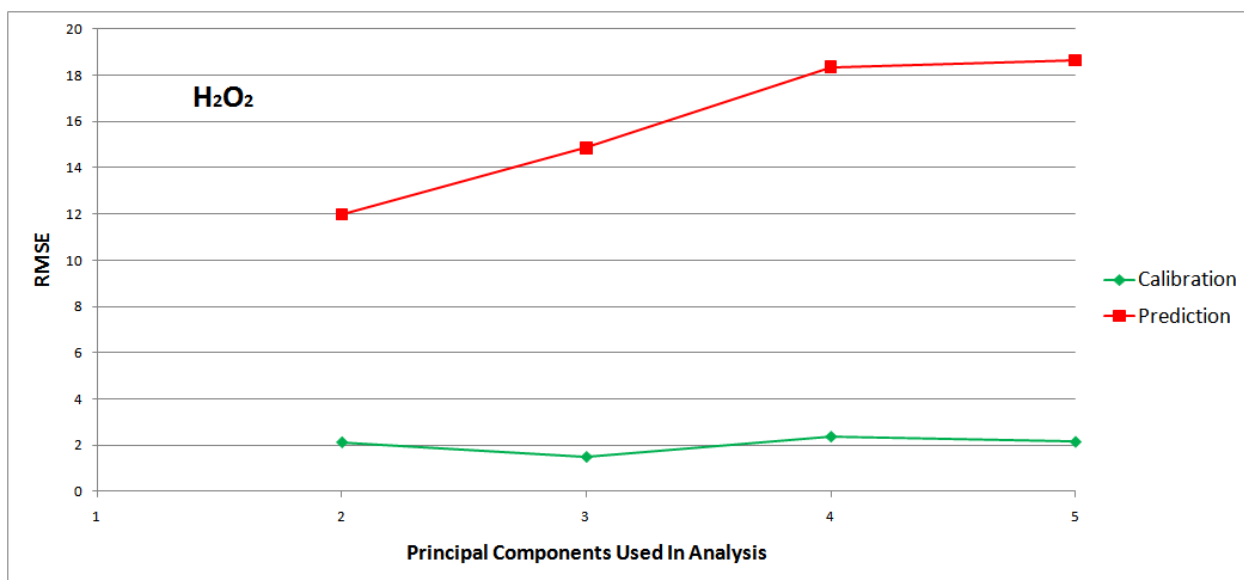


Figure 5-7: H_2O_2 RMSE Calibration and RMSE Prediction as a Function of the Number of Principal Components Used to Represent the Original Data Set for H_2O_2 in Aqueous Solution Mixtures Data Set 1

PCA on the entire set of collected spectra, both calibration and prediction sets, is capable of isolating the peaks associated with each particular component. Since the principal component loadings, $V-1$ and $V-2$, are abstract representations of information within the original data set, it

is acceptable and in most cases unavoidable to have negative elements due to the orthogonality requirement of PCA. With the components separated, the contribution of each component to the sample mixtures can be determined as shown as in Figure 5-8 and 5-9 for the H_2O_2 calibration and prediction data sets, respectively, and Figures 5-10 and 5-11 for H_2O calibration and prediction data set, respectively.

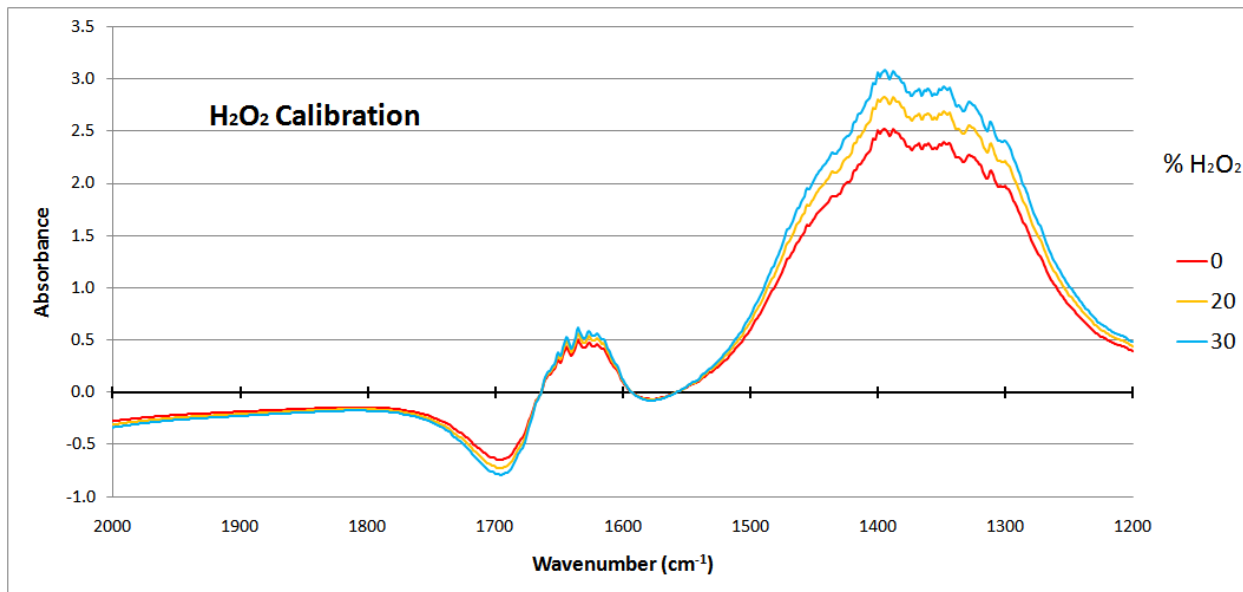


Figure 5-8: PCC H_2O_2 Calibration Data Set 1; Note – Only 3 of 8 Calibration Spectra Shown, 0%, 20%, 30% H_2O_2

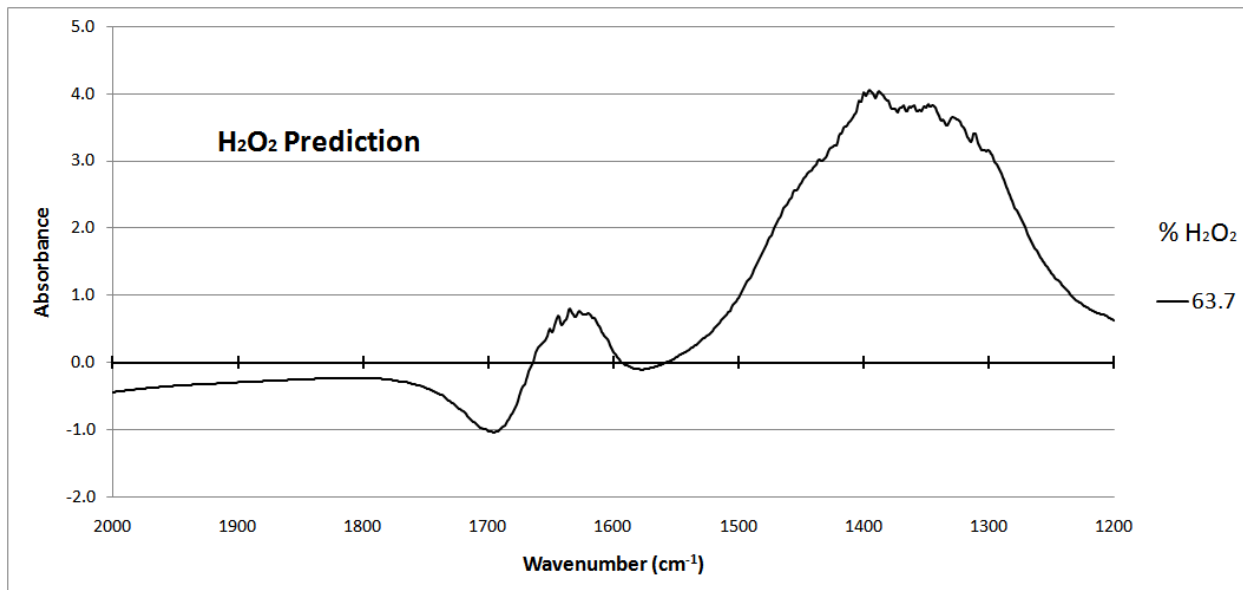


Figure 5-9: PCC H₂O₂ Prediction Data Set 1; 63.7% H₂O₂

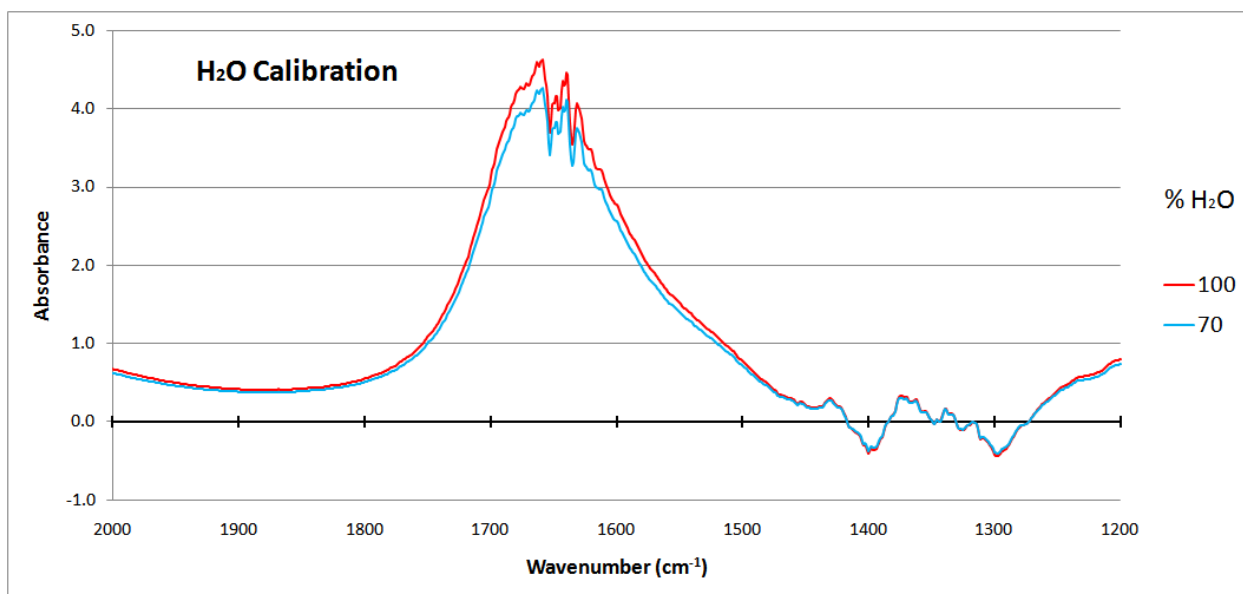


Figure 5-10: PCC H₂O Calibration Data Set 1; Note – Only 2 of 8 Calibration Spectra Shown, 70%, 100% H₂O

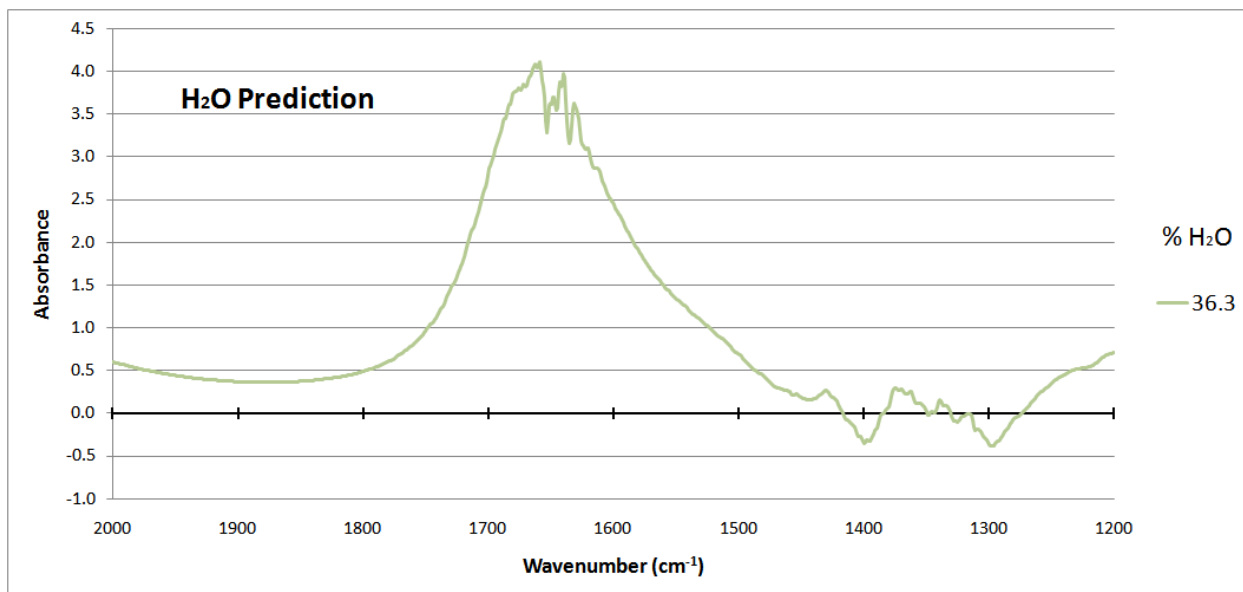


Figure 5-11: PCC H₂O Prediction Data Set 1; 36.3% H₂O

As evident in Figure 5-12 and 5-13 for the H₂O₂ and H₂O calibration data sets, a baseline correction is necessary to remove the generated noise into the data set due to the mathematics in the PCA technique as well as the noise from the experiment. In addition, it is also clear that the samples contain variable total amounts of solution and it will be necessary to correct for this by using the relationship in Equation 4.14 for the prediction data set.

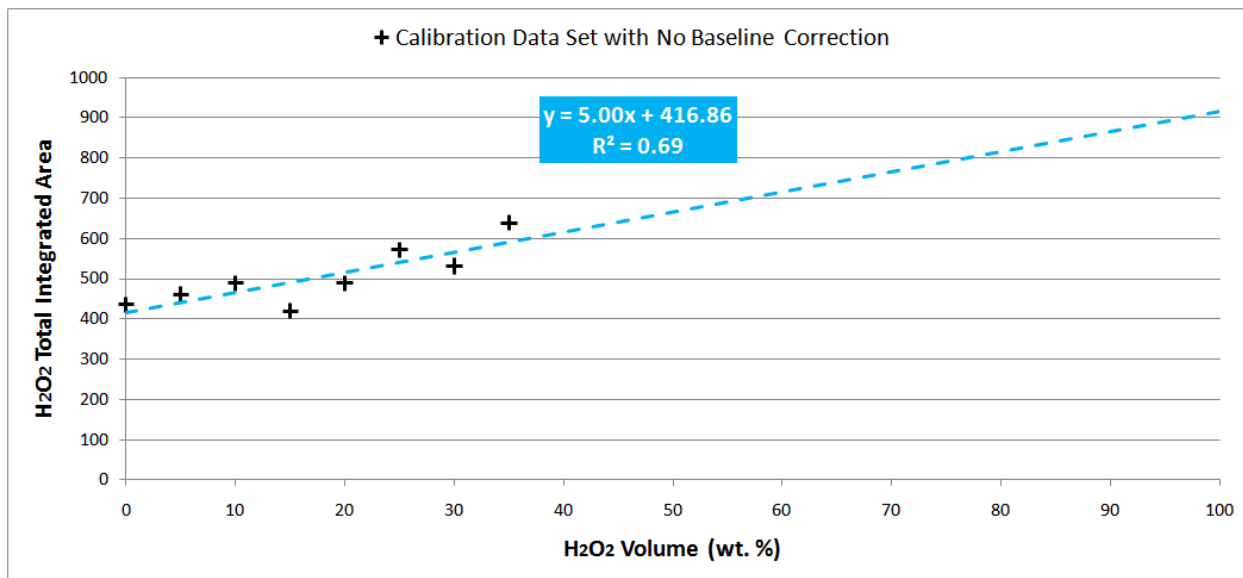


Figure 5-12: PCC H₂O₂ Calibration Data Set 1 with No Baseline Correction

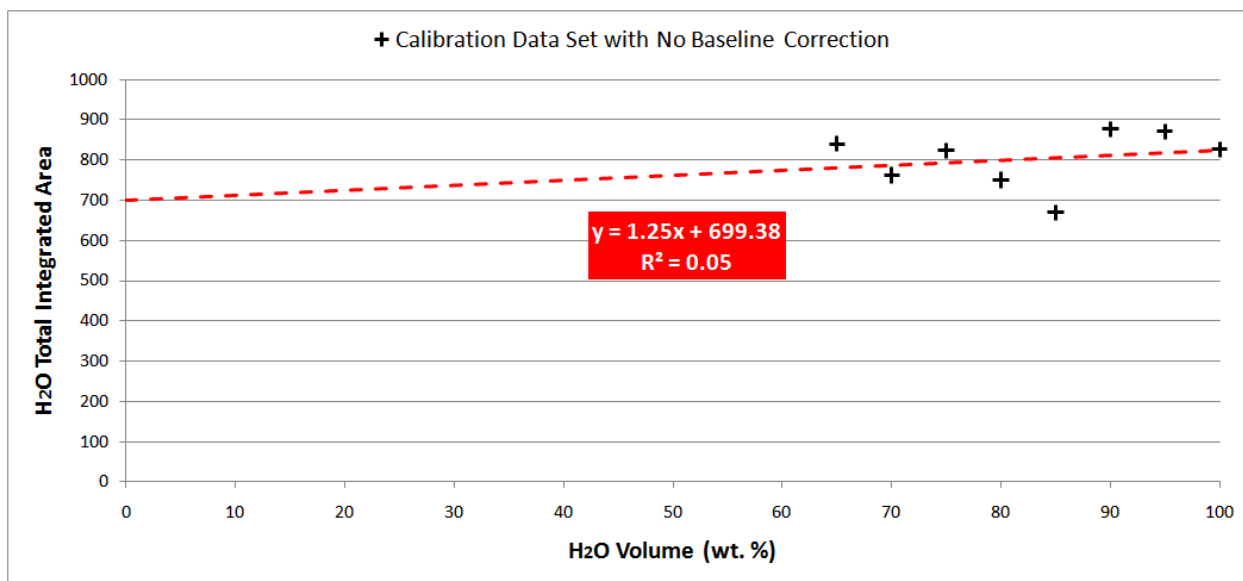


Figure 5-13: PCC H₂O Calibration Data Set 1 with No Baseline Correction

With the baseline correction made (Figures 5-14 and 5-15), and based on the linearity assumption, the proportionality constant can be determined from Equation 4.13 for each of the individual components in the calibration data set from the isolated peaks. For the baseline

correction, it was found that subtraction of the calculated baseline value from the fourth data point (15 wt.% H₂O₂) in the calibration data set produces a negative value. Because of this, the proportionality constant calculation uses a calibration data set with seven components. This may indicate that some error occurred during the FTIR sampling procedure for the 15 wt.% H₂O₂ sample.

The calculated proportionality constants ($k_{\text{H}_2\text{O}_2} = 0.19$, $k_{\text{H}_2\text{O}} = 0.67$), equal to the inverse of the slope of the lines of best fit shown in Figures 5-14 and 5-15, can then be used to calculate the concentration of the mixtures in the PCC model as shown in Figure 5-16. This model requires the utilization of the relationship presented in Equation 4.14 to correct for the total amount of solution varying throughout both the calibration and prediction sample spaces. With the PCC technique, the RMSE (Equation 4.12) for the calibration data set 1 is found to 4.1 wt.%, while the RMSE for the prediction data set 1 is found to be 5.2 wt.%.

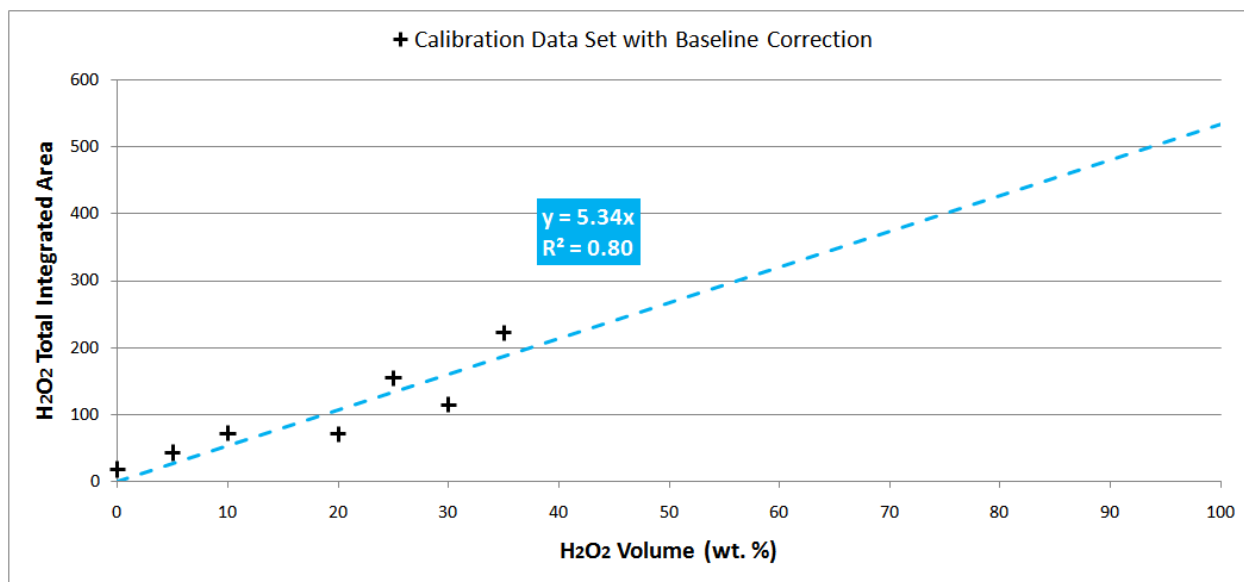


Figure 5-14: PCC H₂O₂ Calibration Data Set 1 with Baseline Correction

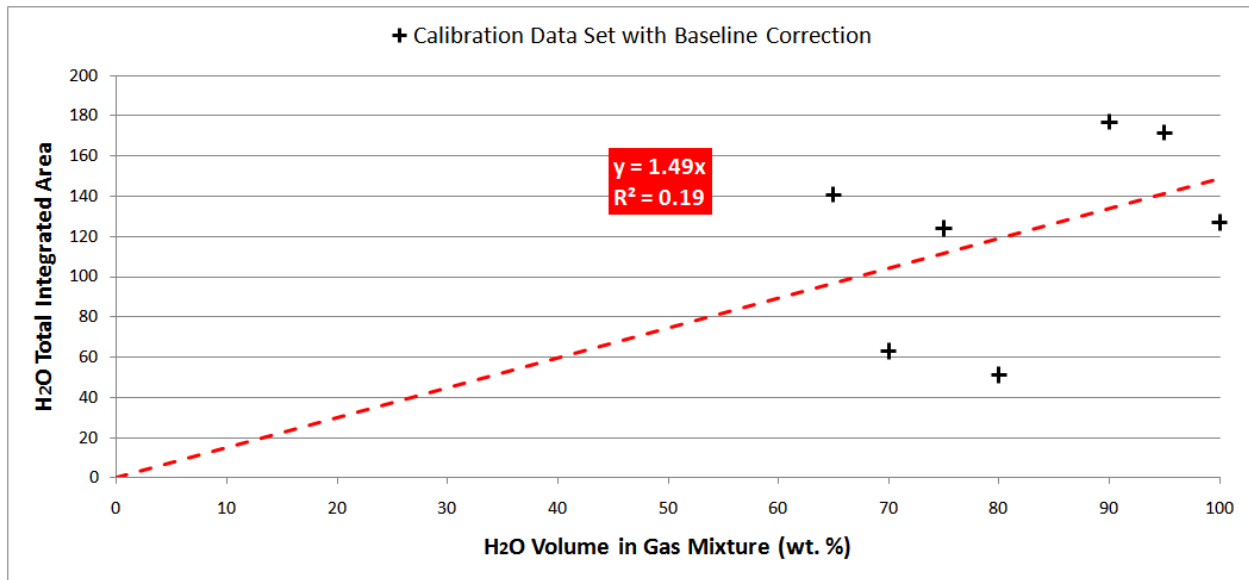


Figure 5-15: PCC H₂O Calibration Data Set 1 with Baseline Correction

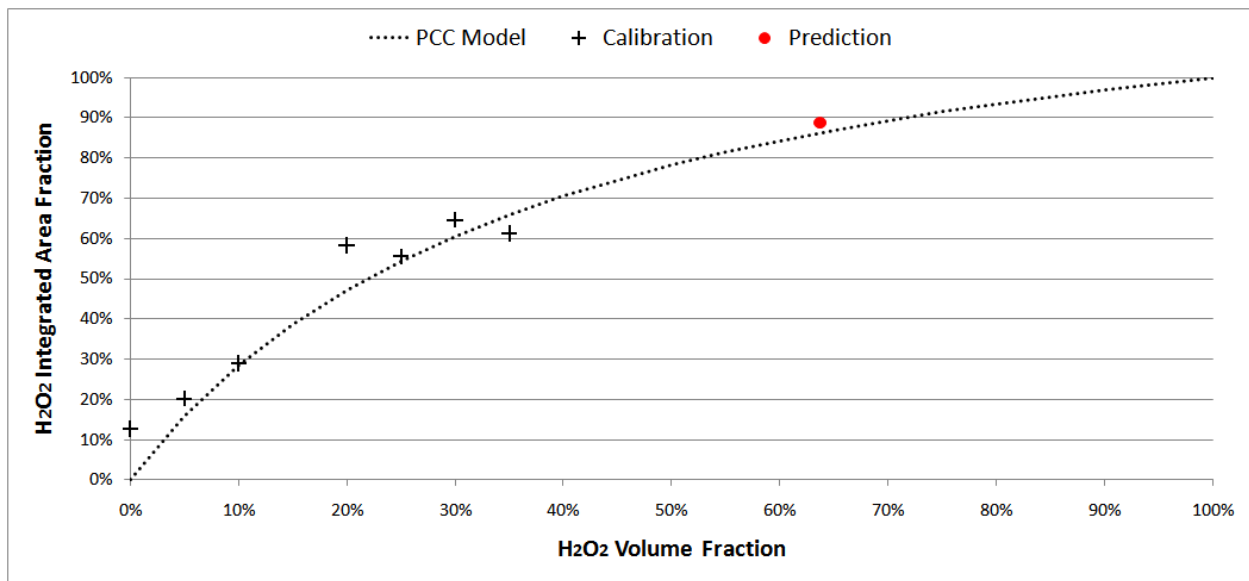


Figure 5-16: PCC Model for wt.% of H₂O₂ in Aqueous Solution with Variable Total Amount in Data Set 1; $k_{\text{H}_2\text{O}_2} = 0.19$, $k_{\text{H}_2\text{O}} = 0.67$; RMSE Calibration = 4.1 wt.%, RMSE Prediction = 5.2 wt.%

As shown in Table 5-1, comparison of the PCR and PCC technique for analysis of data sets that do not contain constant volume of the mixtures across the entire sample space indicates

that the PCC technique performs significantly better in terms of prediction of the unknown concentration. This finding will be further explored by an additional data set presented next that contains 15 spectra of H₂O₂ concentrations in aqueous solutions for the calibration data set and 5 spectra for the prediction data set.

Table 5-1: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; H₂O₂ in Aqueous Solution Data Set 1

Analysis Technique	RMSE Calibration (wt.%)	RMSE Prediction (wt.%)
PCR	2.1	12.0
PCC	4.1	5.2

Data Set 2

From the experimental FTIR spectral analysis of the H₂O₂ in aqueous solution mixture, a set of 15 samples ($n = 15$) at 601 different wavenumbers ($p = 601$) were collected to produce a calibration data set $[X_C]_{(15 \times 601)}$ of wt.% H₂O₂ in aqueous solution from 0-70% at 5% intervals. A set of 5 samples shown in Table 5-2 were collected to produce a prediction data set $[X_P]_{(5 \times 601)}$ for PCR. In the case of the PCC, a single data set $[X]_{(20 \times 601)}$ consisting of the calibration and prediction data sets were combined and PCA was simultaneously performed on all 20 samples to extract out the pure components in each spectra.

Table 5-2: H₂O₂ in Aqueous Solution Spectra Compositions for Prediction Data Set 2, $[X_P]_{(5 \times 601)}$

Sample #	H ₂ O ₂ Concentration (wt.%) from Titration
16	38.4
17	35.9
18	26.0
19	38.7
20	45.2

The number of significant eigenvalues obtained from PCA is two (Figure 5-17). The first principal component explains 61% of the total calibration data set variance and the second

principal component explains of 34% of the total calibration data set variance. With these two principal components, 95% of the original data variance can be explained.

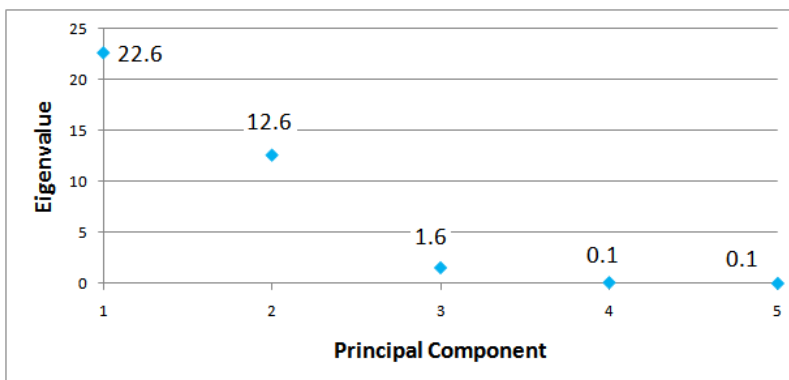


Figure 5-17: SCREE Plot Indicating Eigenvalues for Each Principal Component for H₂O₂ in Aqueous Solution Data Set 2; Principal Components 1 and 2 Explain 61% and 34% of the Total Calibration Data Set Variance

The plot of the calibration data set used to determine the concentrations in the prediction data set is shown in Figure 5-18. The RMSE for the calibration data set 2 is found to be 10.7 wt.%, while the RMSE for the prediction data set 2 is found to be 3.9 wt.%.

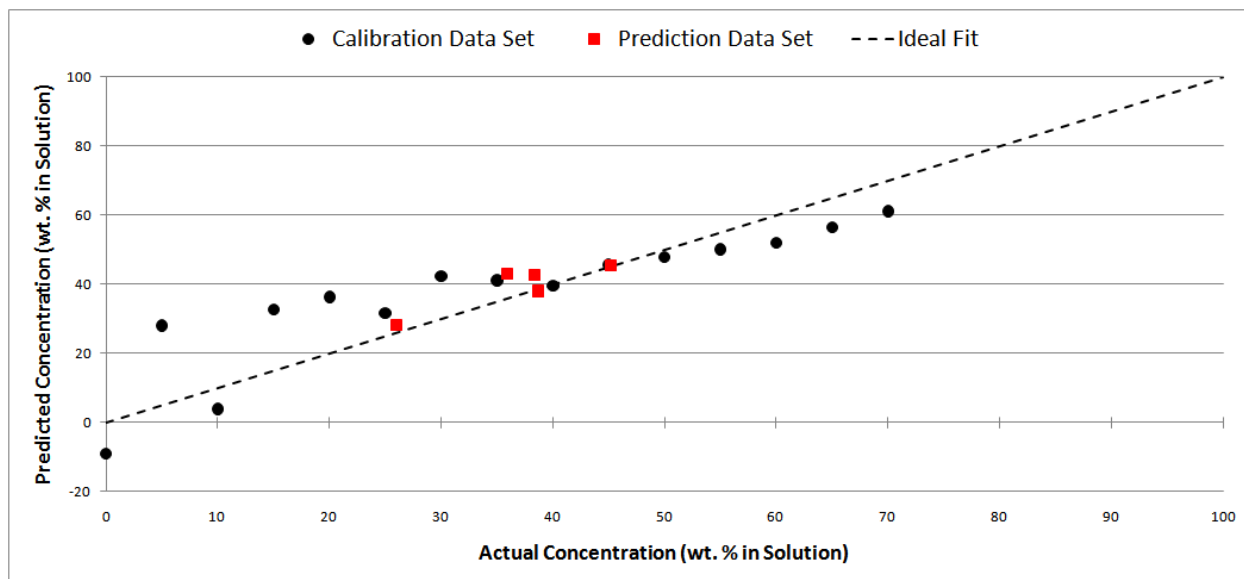


Figure 5-18: Principal Component Regression (PCR) – H₂O₂ Concentrations for H₂O₂ in Aqueous Solution with Variable Total Amount in Data Set 2; RMSE Calibration = 10.7 wt.%, RMSE Prediction = 3.9 wt.%

With the components separated from PCA on the entire set of collected spectra, the contribution of each component to the sample mixtures can be determined. Baseline corrections were made to the calibration data set spectra to account for non-zero values for the 0% H₂O₂ concentration spectra as well as from extrapolation to 0% H₂O concentration from the spectra values shown. The calculated proportionality constants ($k_{\text{H}_2\text{O}_2} = 0.65$, $k_{\text{H}_2\text{O}} = 0.40$) are used to calculate the concentration of the mixtures in the PCC model as shown in Figure 5-19. With the PCC technique, the RMSE for the calibration data set 2 is found to be 5.9 wt.%, while the RMSE for the prediction data is found to be 6.0 wt.%.

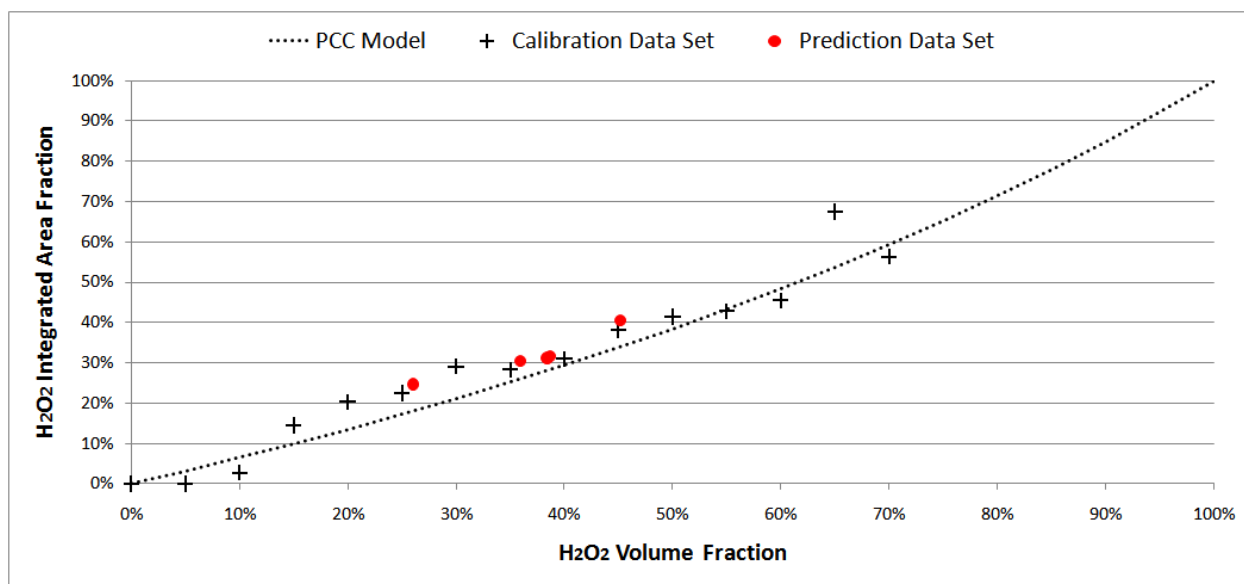


Figure 5-19: PCC Model for wt.% of H₂O₂ in Aqueous Solution with Variable Total Amount in Data Set 2; $k_{\text{H}_2\text{O}_2} = 0.65$, $k_{\text{H}_2\text{O}} = 0.40$; RMSE Calibration = 5.9 wt.%, RMSE Prediction = 6.0 wt.%

As shown in Table 5-3, comparison of the PCR and PCC technique for analysis of data sets that do not contain constant volume of the mixtures across the entire sample space indicates that the PCC technique performs significantly better overall for the calibration data set when compared on a RMSE basis. PCC compares reasonable well to the PCR technique in terms of

RMSE for the prediction data set. For the PCC technique, the RMSE were of 5.9 wt.% and 6.0 wt.% for the calibration and prediction data sets, respectively compared to the PCR technique, which had RMSE of 10.7 wt.% and 3.9 wt.% for the calibration and prediction data set, respectively. The PCC technique performs similarly in low (0-20 wt.%), medium (25-45 wt.%), and high ranges (50-70 wt.%) of H₂O₂ concentration with the RMSE of calibration for each of the segments for being 6.0 wt.%, 6.0 wt.%, and 5.7 wt.%, respectively. In contrast, the performance of the PCR technique in low H₂O₂ concentrations (15.7 wt.%) is worse than that for medium (6.8 wt.%) and high (7.0 wt.%) H₂O₂ concentration ranges.

Table 5-3: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; H₂O₂ in Aqueous Solution Data Set 2

Analysis Technique	RMSE Calibration (wt.%)	RMSE Prediction (wt.%)
PCR – Overall	10.7	3.9
PCR – Low (0-20 wt.% H₂O₂)	15.7	-
PCR – Med (25-45 wt.% H₂O₂)	6.8	-
PCR – High (50-70 wt.% H₂O₂)	7.0	-
PCC – Overall	5.9	6.0
PCC – Low (0-20 wt.% H₂O₂)	6.0	-
PCC – Med (25-45 wt.% H₂O₂)	6.0	-
PCC – High (50-70 wt.% H₂O₂)	5.7	-

This study indicates that FTIR spectroscopy used in conjunction with the discussed chemometric techniques has the potential to be utilized in determining the H₂O₂ concentrations in aqueous solutions from condensation events that may occur during a VPHP decontamination event, even when a complete calibration data set is not utilized, as was shown with data set 1.

Chapter 6: PCA Application to FTIR Spectroscopy Data of Potential Environment Air Contaminants within the Aircraft Cabin

Discussion of Analyzed Data Sets

In the simulation data sets, pure spectra from the QASoft® database are used. To form a simulated mixture, pure spectra are added together and different multiplication factors are applied to achieve a range of component concentrations. The simulated data sets consist of various spectra of targeted component systems for two and three component systems. The simulated data sets are comprised of the entire pure spectra of each of the components. To make the data sets more manageable, the analyzed data sets are comprised of the raw data averaged over ten wavenumber data points, thus reducing the size by an order of magnitude. This reduced data set is then used as input in a MATLAB® program to determine the PCA, PCR, and PCC characteristics of the data.

The first two-component simulation data set was a gas mixture of CH_2O and $\text{C}_3\text{H}_4\text{O}$. This system is of interest because of the strong overlap that the two components have in their respective IR spectra. The second two-component simulation data set was of gas mixtures consisting of CO and CO_2 . The experimental two-component data set consisted of mixtures of CO and CO_2 from gas cylinders. This system is of interest in that it is expected to be a combination of potential environmental air contaminants most commonly found in the aircraft cabin as well as being a data set that can be tested within both the Commercial Sensor Module and FTIR Gas Analysis Module in conjunction with simultaneous testing of commercial sensors.

For the three-component systems, a simulation data set consisting of a gas mixture of CH_2O , $\text{C}_3\text{H}_4\text{O}$, and H_2O was analyzed and like the two-component system, all three components have significant overlap within their respective spectra. The second three-component simulation data set was of gas mixtures consisting of CO , CO_2 , and H_2O . The experimental three-component data set consisted of mixtures of CO and CO_2 from gas cylinders with H_2O introduced into the system from evaporation in heated crucible linked to the FTIR Gas Analysis Module. The introduced was not directly controlled and was allowed to enter the system as it was evaporated to present variable water vapor levels when detecting known concentrations of CO/CO_2 gas mixtures. Even though the H_2O spectrum does not overlap with CO or CO_2 spectra, it is expected to be present at significantly higher concentrations than the target components that could be potential environmental air contaminants within the aircraft cabin. The pathlength of the gas cell for the experimental gas mixtures was kept constant at 2.24 cm.

PCA Results and Discussions: 2-Component Systems

$\text{CH}_2\text{O}/\text{C}_3\text{H}_4\text{O}$ Simulation Data Set

From the simulated FTIR spectral analysis of the $\text{CH}_2\text{O}/\text{C}_3\text{H}_4\text{O}$ gas mixtures, a set of 8 samples ($n = 8$) at 2,655 different wavenumbers ($p = 2,655$) with concentrations given in Table 6-1 were compiled as shown in Figure 6-1 to produce a calibration data set $[\text{X}_\text{C}]_{(8 \times 2655)}$. A set of 3 samples with concentrations given in Table 6-2 were compiled as shown in Figure 6-2 to produce a prediction data set $[\text{X}_\text{P}]_{(3 \times 2655)}$ for PCR. In the case of the PCC, a single data set $[\text{X}]_{(11 \times 2655)}$ consisting of the calibration and prediction data sets were combined and PCA was simultaneously performed on all 11 samples to extract out the pure components in each spectra.

Table 6-1: CH₂O/C₃H₄O Spectra Compositions for Calibration Data Set, [X_C]_(8 x 2655), Ordered from Low to High Concentration of CH₂O

Sample #	Amount of CH ₂ O (x100 ppm)	Amount of C ₃ H ₄ O (x100 ppm)
1	0.00	2.50
2	0.50	1.00
3	1.00	0.50
4	1.50	2.00
5	2.00	1.50
6	3.00	4.75
7	4.75	3.00
8	10.00	10.00

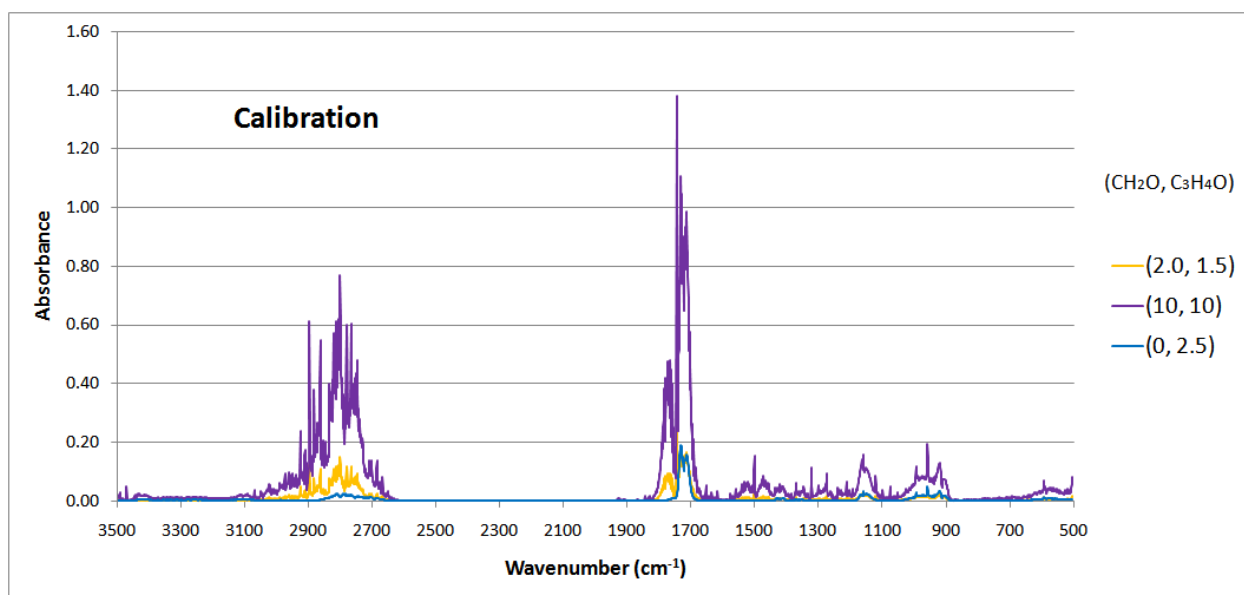


Figure 6-1: CH₂O/C₃H₄O Gas Mixtures Calibration Data Set, [X_C]_(8 x 2655); Note – Only 3 of 8 Calibration Spectra Shown

Table 6-2: CH₂O/C₃H₄O Spectra Compositions for Prediction Data Set, [X_P]_(3 x 2655), Ordered from Low to High Concentration of CH₂O

Sample #	Amount of CH ₂ O (x100 ppm)	Amount of C ₃ H ₄ O (x100 ppm)
9	1.33	8.00
10	2.50	0.00
11	8.00	1.33

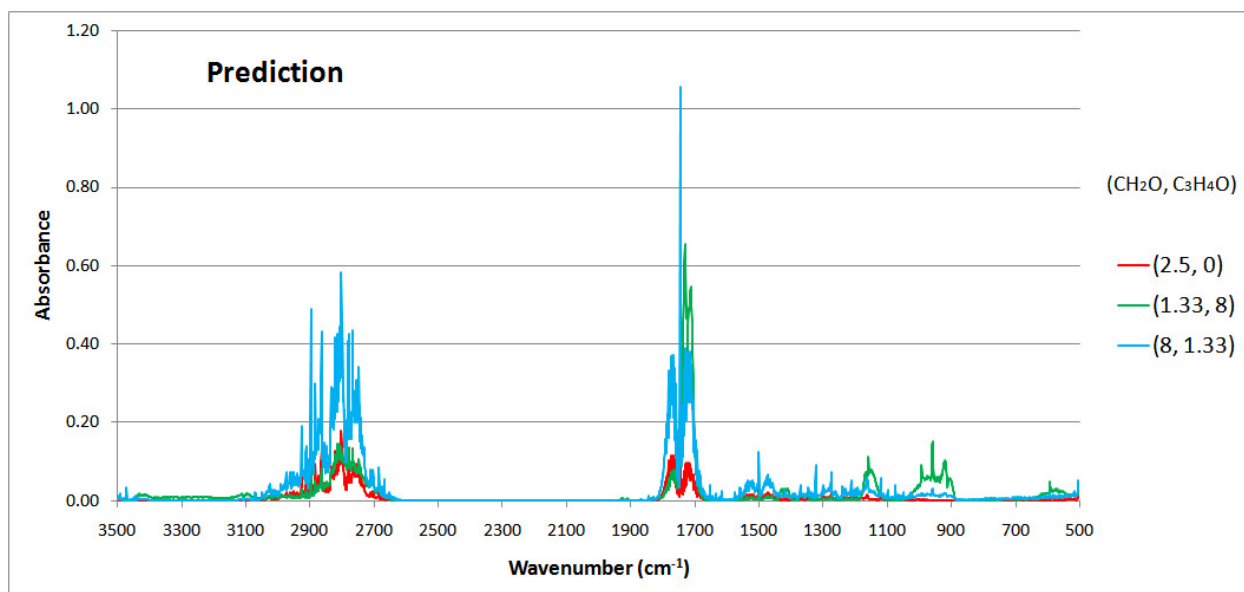


Figure 6-2: CH₂O/C₃H₄O Gas Mixtures Prediction Data Set, [X_P]_(3 x 2655)

As shown in the plot of eigenvalues as a function of principal components, the number of significant eigenvalues obtained from PCA is two (Figure 6-3). The first principal component explains 98.4% of the total calibration data set variance and the second principal component explains of 1.6% of the total calibration data set variance. With these two principal components, 100.0% of the original data variance can be explained.

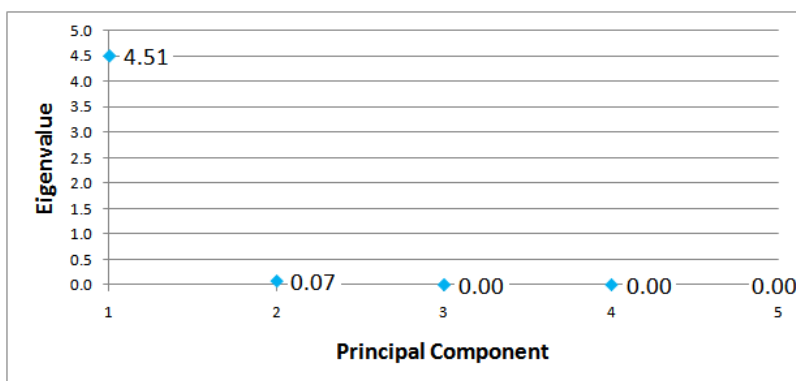


Figure 6-3: SCREE Plot Indicating Eigenvalues for Each Principal Component for CH₂O/C₃H₄O Gas Mixtures Data Set; Principal Components 1 and 2 Explain 98.4% and 1.6% of the Total Calibration Data Set Variance

Applying PCR, a plot can be created that shows the calibration data set used to determine the concentrations in the prediction data set as shown in Figures 6-4 and 6-5 for CH_2O and $\text{C}_3\text{H}_4\text{O}$, respectively. With the PCR technique, the RMSE for the calibration and prediction data sets in regards to the concentration of CH_2O and $\text{C}_3\text{H}_4\text{O}$ are found to be 0 ppm.

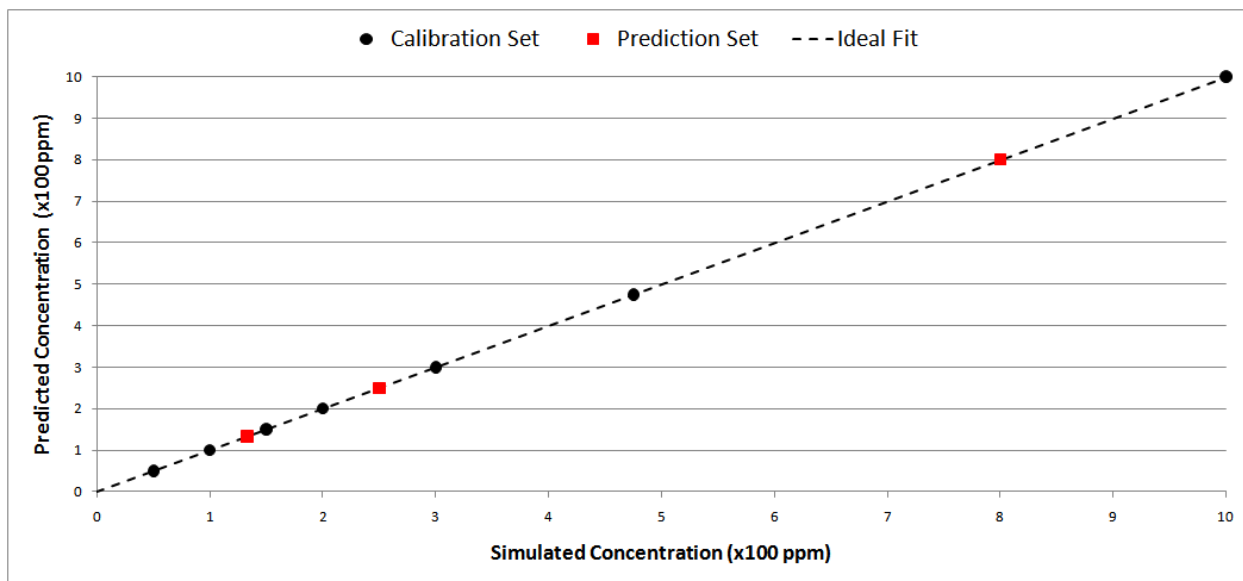


Figure 6-4: Principal Component Regression (PCR) – CH_2O Concentrations in $\text{CH}_2\text{O}/\text{C}_3\text{H}_4\text{O}$ Gas Mixtures; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm

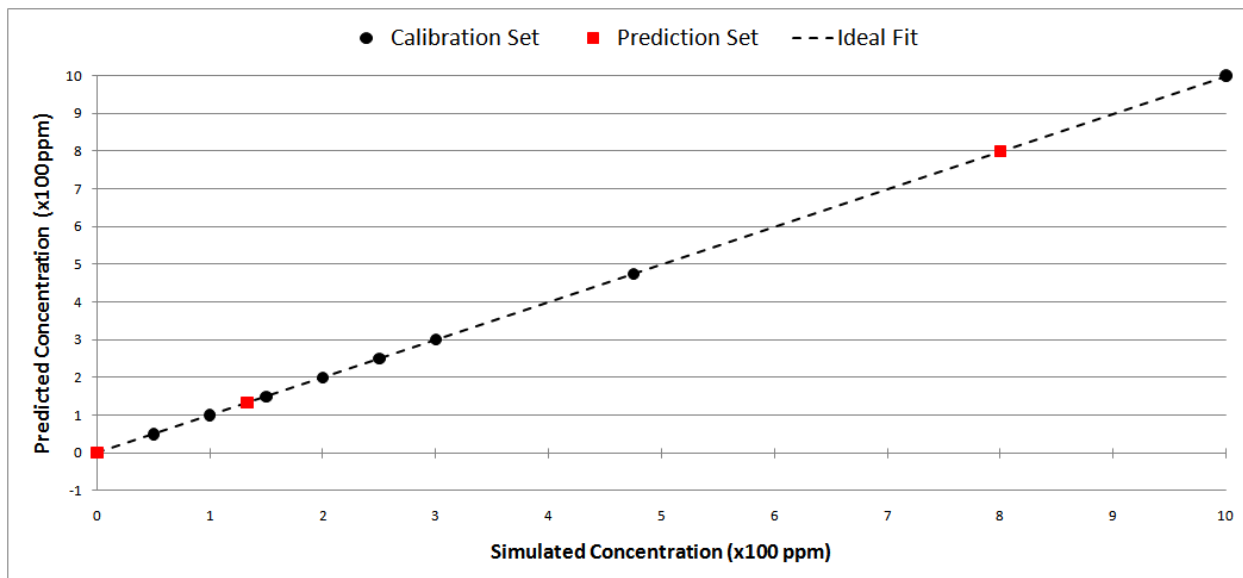


Figure 6-5: Principal Component Regression (PCR) – C_3H_4O Concentrations in CH_2O/C_3H_4O Gas Mixtures; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm

PCA on the entire set of collected spectra, both calibration and prediction sets, isolates the peaks associated with each particular component as shown in Figures 6-6 and 6-7. The benefit of these figures are that *V-1* and *V-2* can be easily identifiable as corresponding to the pure spectra of CH_2O and C_3H_4O and, respectively. With the components separated, the contribution of each component to the sample mixtures can be determined.

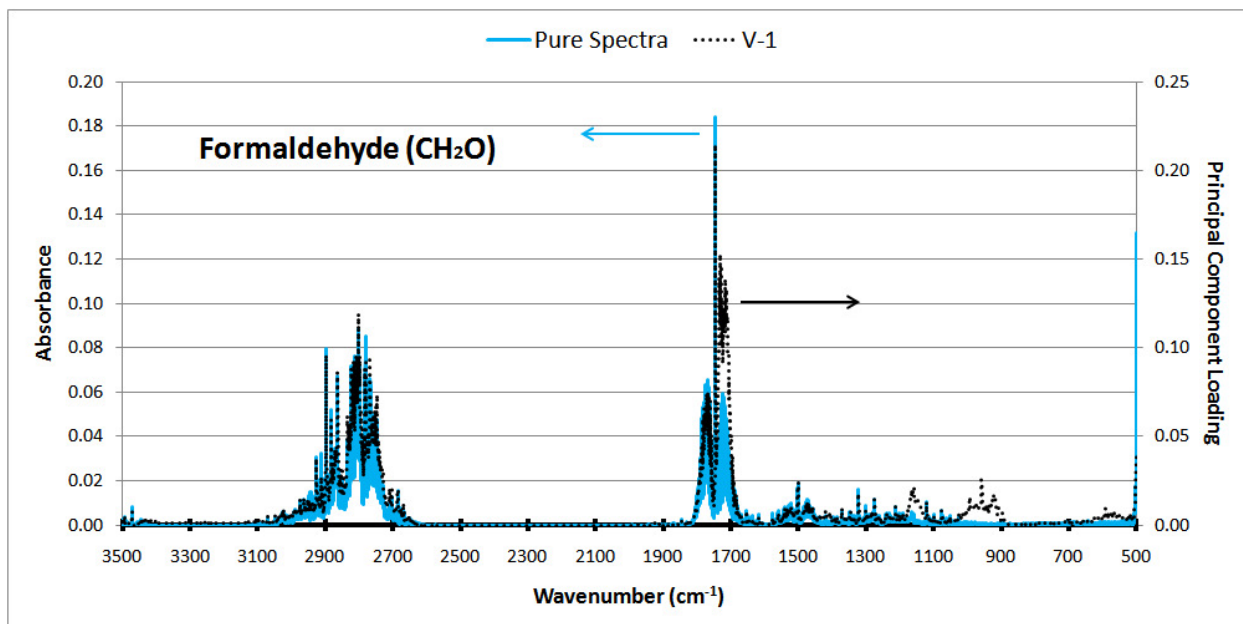


Figure 6-6: PCA Separated CH_2O Spectra in $\text{CH}_2\text{O}/\text{C}_3\text{H}_4\text{O}$ Gas Mixtures

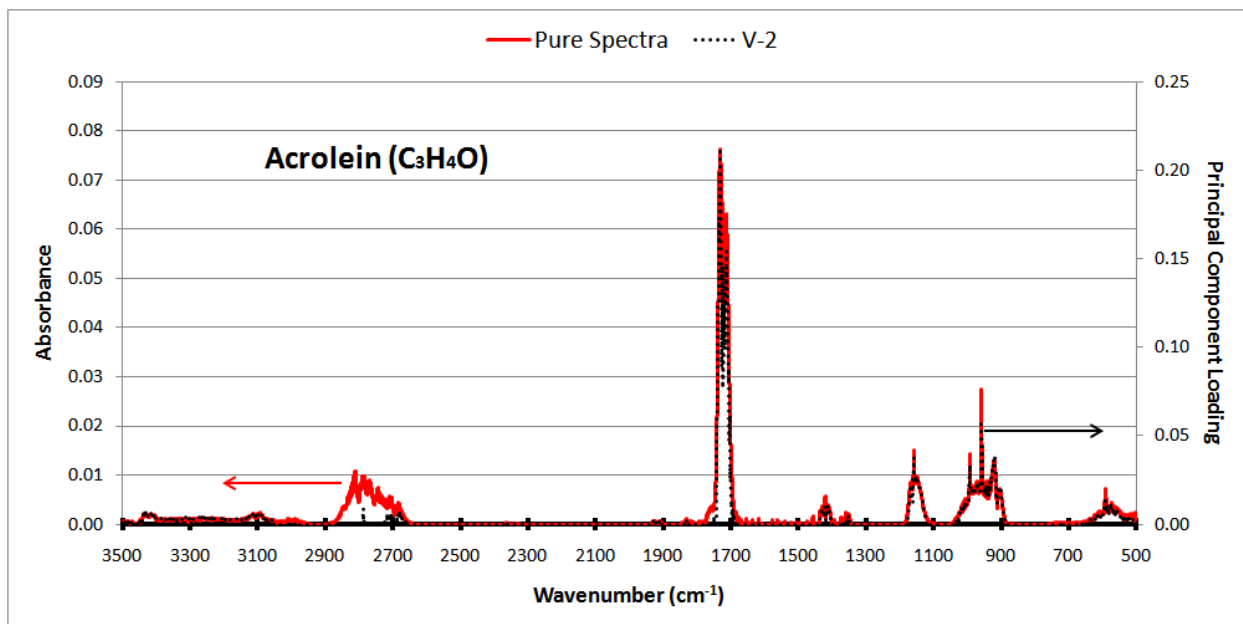


Figure 6-7: PCA Separated $\text{C}_3\text{H}_4\text{O}$ Spectra in $\text{CH}_2\text{O}/\text{C}_3\text{H}_4\text{O}$ Gas Mixtures

With the baseline correction made, and based on the linearity assumption, the proportionality constant is determined for each of the individual components in the calibration

data set from the isolated peaks. The calculated proportionality constants ($k_{\text{CH}_2\text{O}} = 0.08$, $k_{\text{C}_3\text{H}_4\text{O}} = 0.32$), equal to the inverse of the slope of the lines of best fit shown, can then be used to calculate the concentrations of the simulate gas mixtures for CH_2O and $\text{C}_3\text{H}_4\text{O}$ shown in Figures 6-8 and 6-9, respectively. With the PCC method, the RMSE for CH_2O and $\text{C}_3\text{H}_4\text{O}$ calibration data sets are found to be 1 ppm and 2 ppm, respectively, while the RMSE of the prediction data sets for CH_2O and $\text{C}_3\text{H}_4\text{O}$ are found to be 0 ppm and 2 ppm, respectively.

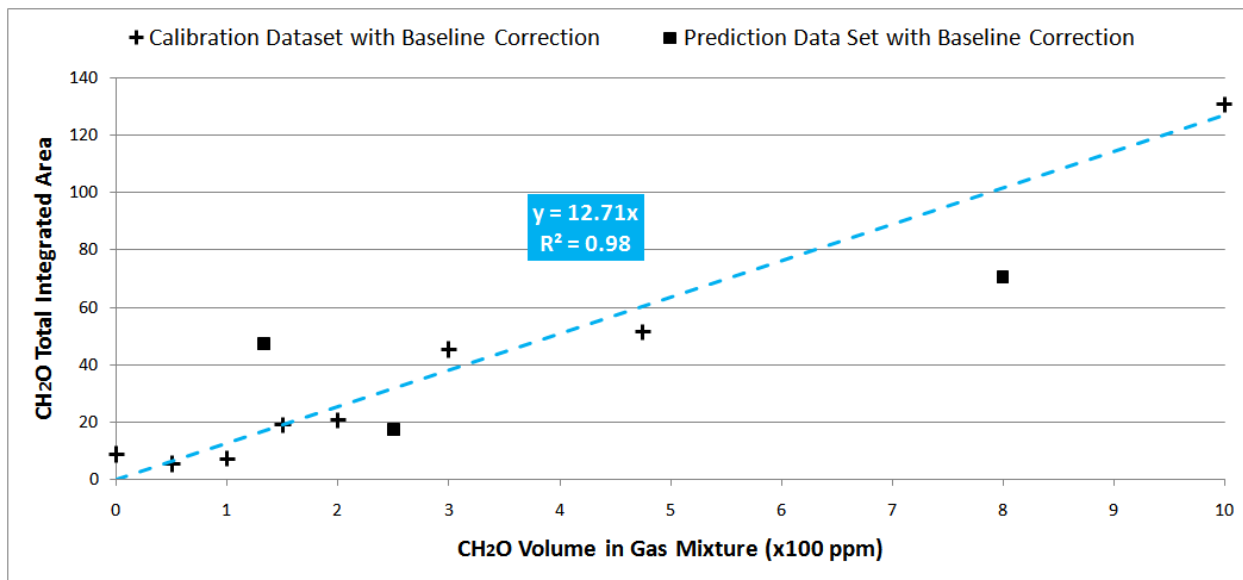


Figure 6-8: PCC CH_2O Calibration and Prediction Data Sets in $\text{CH}_2\text{O}/\text{C}_3\text{H}_4\text{O}$ Gas Mixtures; $k_{\text{CH}_2\text{O}} = 0.08$; RMSE Calibration = 1 ppm, RMSE Prediction = 2 ppm

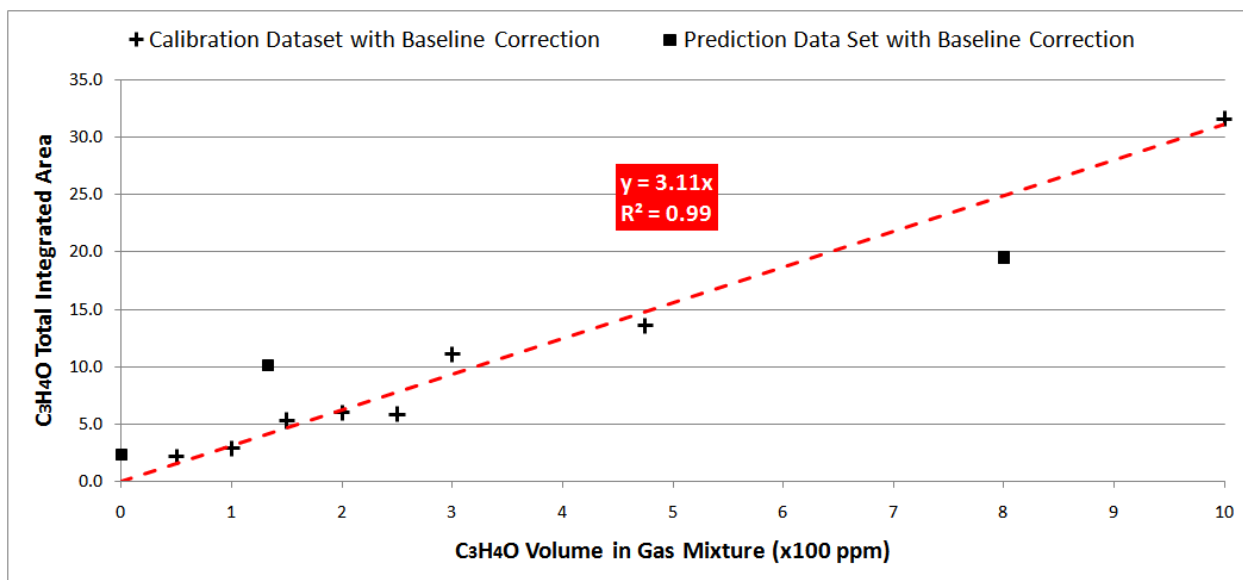


Figure 6-9: PCC C₃H₄O Calibration and Prediction Data Sets in CH₂O/C₃H₄O Gas Mixtures; $k_{C_3H_4O} = 0.32$; RMSE Calibration = 0 ppm, RMSE Prediction = 2 ppm

As shown in Table 6-3, the comparison of the PCR and PCC techniques indicate that the PCC technique could be a viable alternative quantitative analysis method to the PCR technique. The PCC technique for this example does not perform as well as PCR due to the strong overlap of the two components that make the separation of the two spectra more difficult for PCA. The PCA representation for CH₂O still contains small amounts of spectra that are due to C₃H₄O thus, the calculation for the area under the curve for each concentration introduces error within the PCC technique. In addition, the PCA representation for C₃H₄O does not contain the entire spectra for the wavenumber range of 2900-2700 cm⁻¹, where there is a strong overlap of the CH₂O spectra.

Table 6-3: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; CH₂O/C₃H₄O Gas Mixtures

Analysis Technique	RMSE Calibration (ppm)	RMSE Prediction (ppm)
CH ₂ O: PCR	0	0
CH ₂ O: PCC	1	2
C ₃ H ₄ O: PCR	0	0
C ₃ H ₄ O: PCC	0	2

CO/CO₂ Simulation Data Set

From the simulated FTIR spectral analysis of the CO/CO₂ gas mixtures, a set of 10 samples ($n = 10$) at 2,636 different wavenumbers ($p = 2,636$) with concentrations given in Table 6-4 were compiled as shown in Figure 6-10 to produce a calibration data set $[X_C]_{(10 \times 2636)}$. A set of 5 samples with concentrations given in Table 6-5 were compiled as shown in Figure 6-112 to produce a prediction data set $[X_P]_{(5 \times 2636)}$ for PCR. In the case of the PCC, a single data set $[X]_{(15 \times 2636)}$ consisting of the calibration and prediction data sets were combined and PCA was simultaneously performed on all 15 samples to extract out the pure components in each spectra.

Table 6-4: CO/CO₂ Spectra Compositions for Calibration Data Set, $[X_C]_{(10 \times 2636)}$

Sample #	Amount of CO (x100 ppm)	Amount of CO ₂ (x100 ppm)
1	0.00	5.00
2	8.00	3.00
3	6.00	1.00
4	4.00	0.50
5	2.00	1.50
6	1.00	3.50
7	3.00	0.25
8	5.00	0.00
9	7.00	2.00
10	9.00	4.00

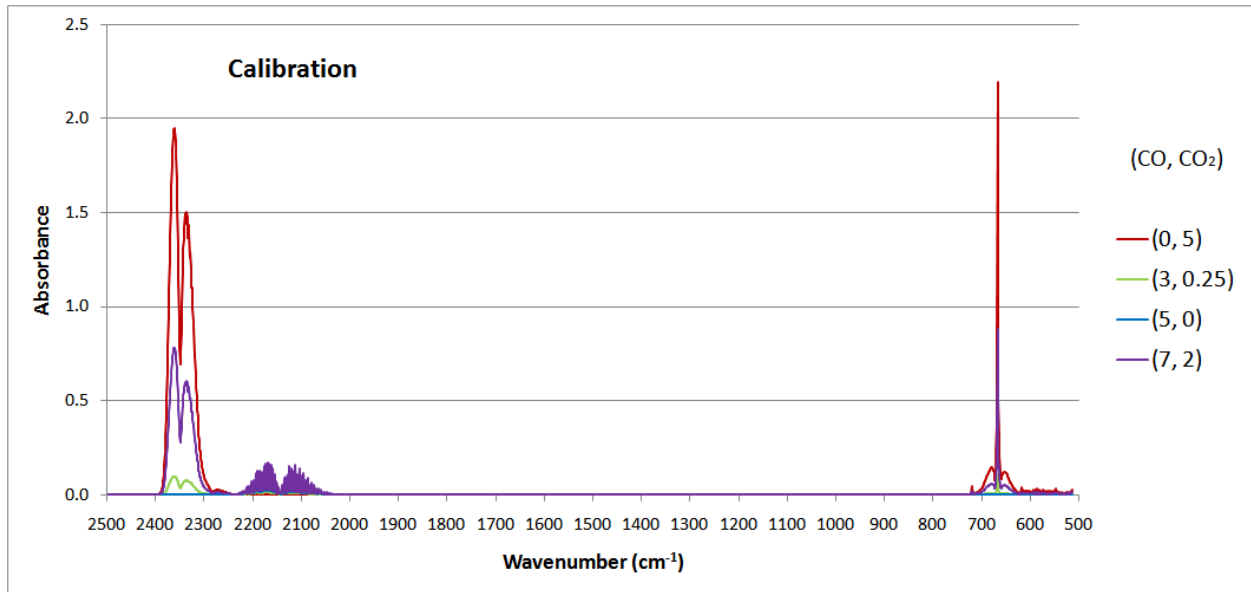


Figure 6-10: CO/CO₂ Gas Mixtures Calibration Data Set, $[X_C]_{(10 \times 2636)}$; Note – Only 4 of 10 Calibration Spectra Shown

Table 6-5: CO/CO₂ Spectra Compositions for Prediction Data Set, $[X_P]_{(5 \times 2636)}$

Sample #	Amount of CO (x100 ppm)	Amount of CO ₂ (x100 ppm)
11	1.25	0.40
12	5.67	0.80
13	2.25	2.20
14	4.33	1.20
15	8.75	0.60

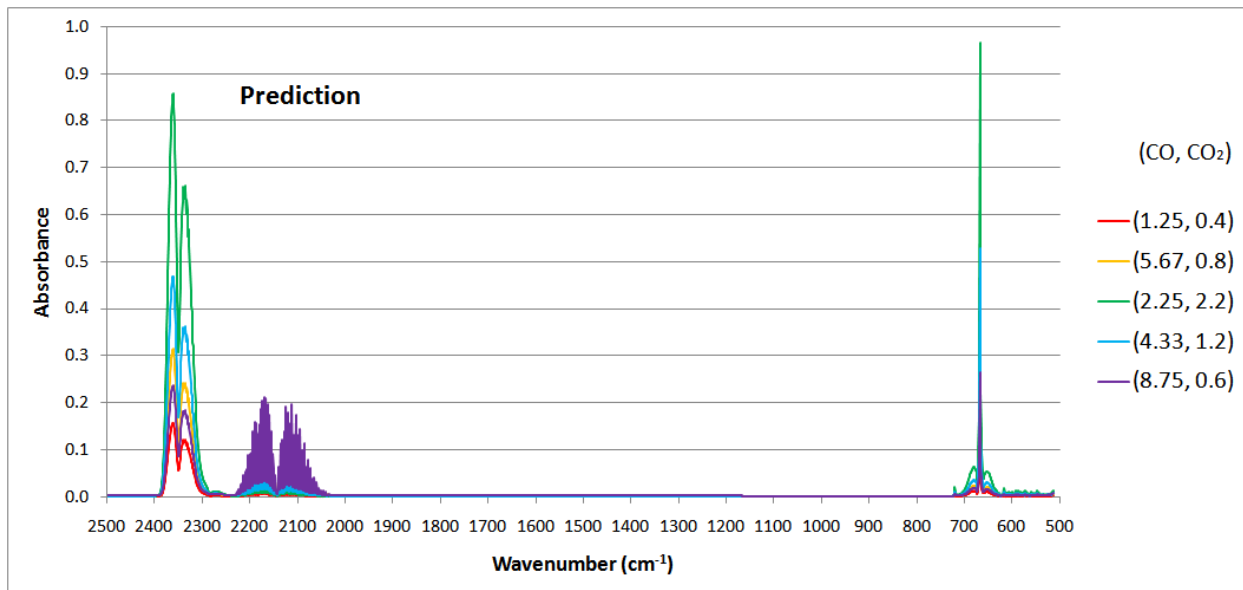


Figure 6-11: CO/CO₂ Gas Mixtures Prediction Data Set, [X_P]_(5 x 2636)

As shown in the plot of eigenvalues as a function of principal components, the number of significant eigenvalues obtained from PCA is two (Figure 6-12). The first principal component explains 99.1% of the total calibration data set variance and the second principal component explains 0.9% of the total calibration data set variance. With these two principal components, 100.0% of the original data variance can be explained.

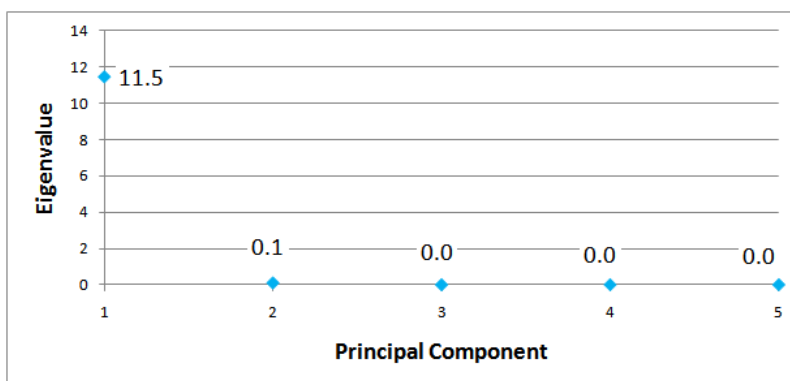


Figure 6-12: SCREE Plot Indicating Eigenvalues for Each Principal Component for CO/CO₂ Gas Mixtures Data Set; Principal Components 1 and 2 Explain 99.1% and 0.9% of the Total Calibration Data Set Variance

Applying PCR, a plot is created that shows the calibration data set used to determine the concentrations in the prediction data set as shown in Figures 6-13 and 6-14 for CO and CO₂, respectively. With the PCR technique, the RMSE for the calibration and prediction data set in regards to concentrations of CO and CO₂ is found to be 0 ppm.

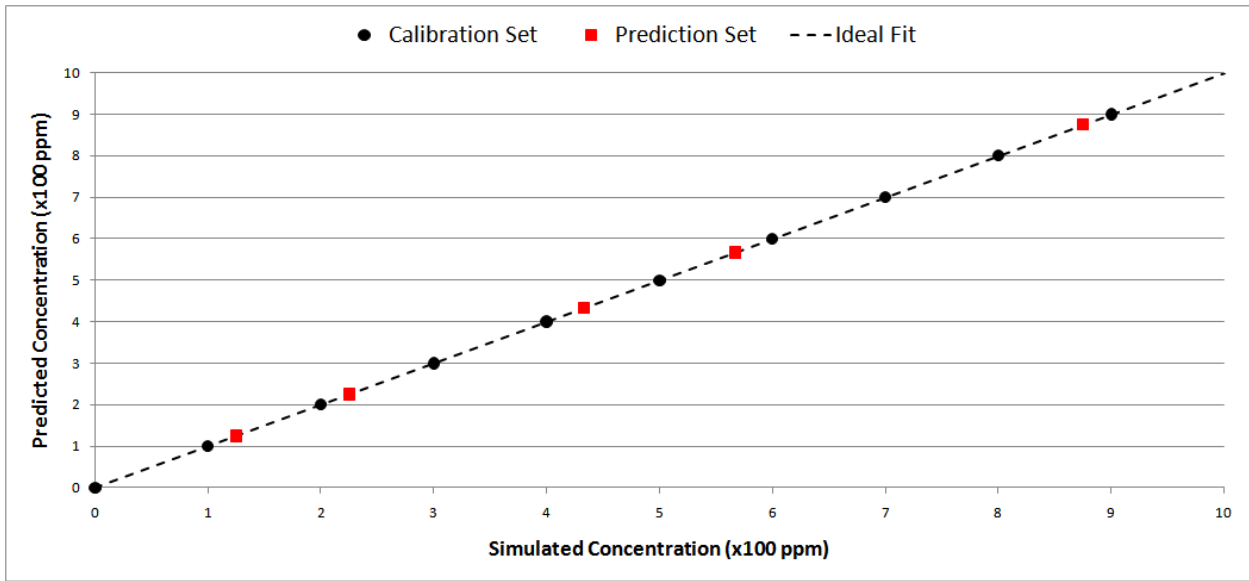


Figure 6-13: Principal Component Regression (PCR) – CO Concentrations in CO/CO₂ Gas Mixtures; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm

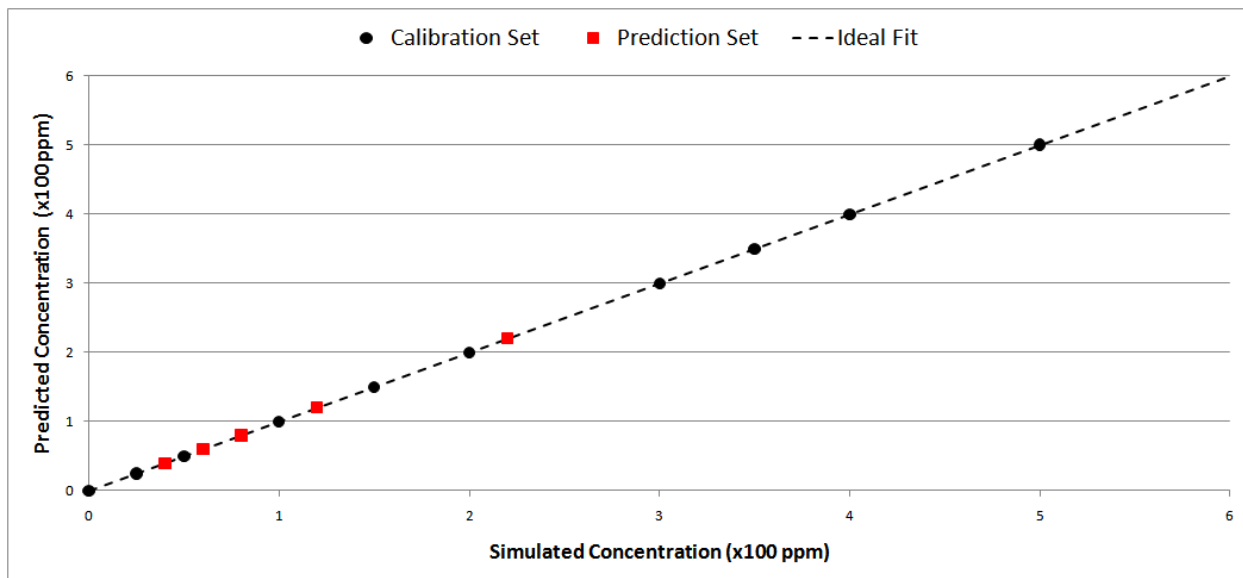


Figure 6-14: Principal Component Regression (PCR) – CO₂ Concentrations in CO/CO₂ Gas Mixtures; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm

Figures 6-15 and 6-16, produced from PCA on the entire set of compiled spectra, show that *V-2* and *V-1* can be easily identifiable as corresponding to the pure spectra of CO and CO₂ and, respectively. With the components separated, the contribution of each component to the sample mixtures can be determined for the CO calibration and prediction data sets, and for CO₂ calibration and prediction data set.

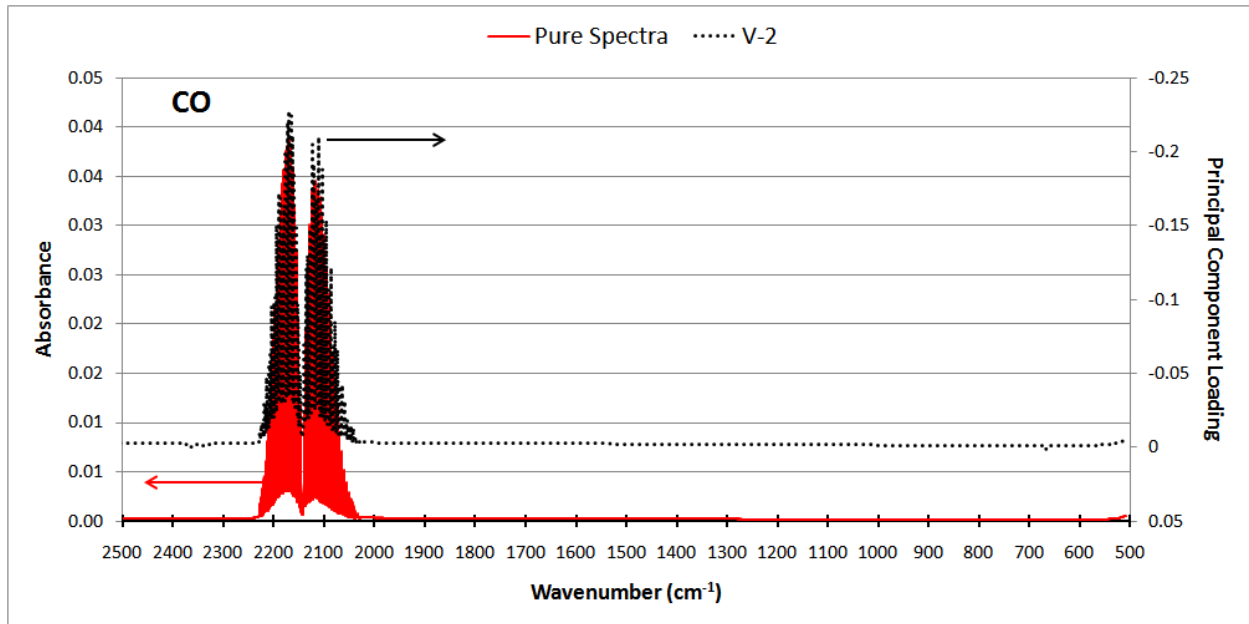


Figure 6-15: PCA Separated CO Spectra in CO/CO₂ Gas Mixtures

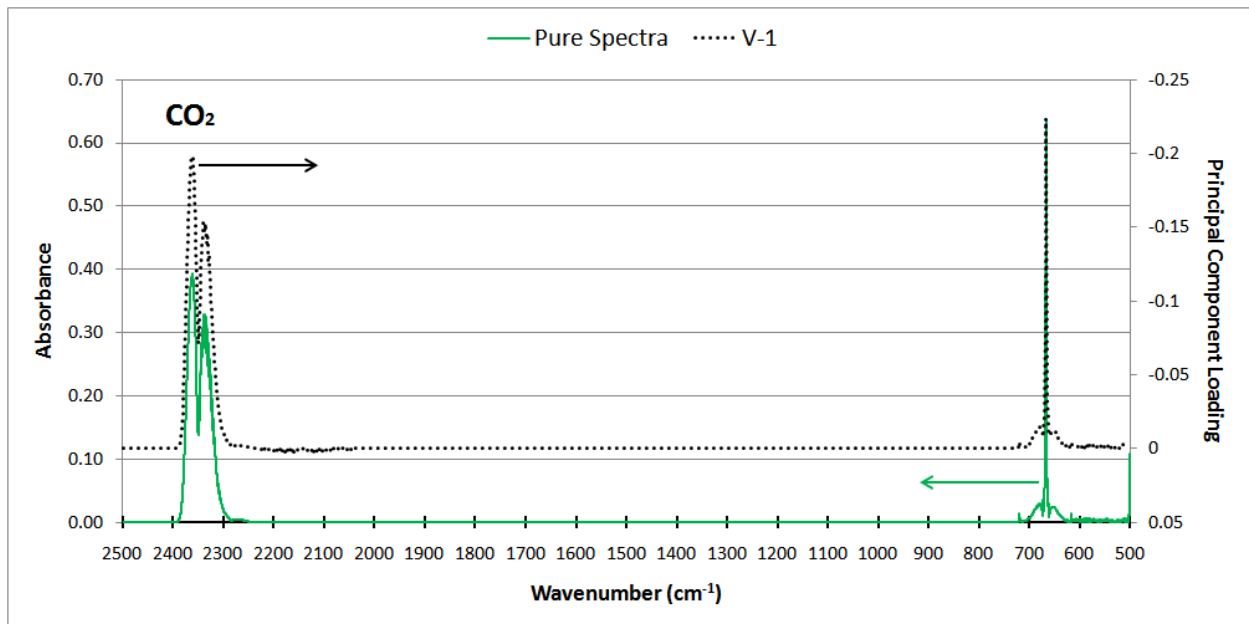


Figure 6-16: PCA Separated CO₂ Spectra in CO/CO₂ Gas Mixtures

The calculated proportionality constants ($k_{CO} = 0.75$, $k_{CO_2} = 0.06$), equal to the inverse of the slope of the lines of best fit shown, are then be used to calculate the concentration of the

mixtures for CO and CO₂ shown in Figures 6-17 and 6-18, respectively. With the PCC method, the RMSE for CO and CO₂ calibration and prediction data sets are found to be 0 ppm .

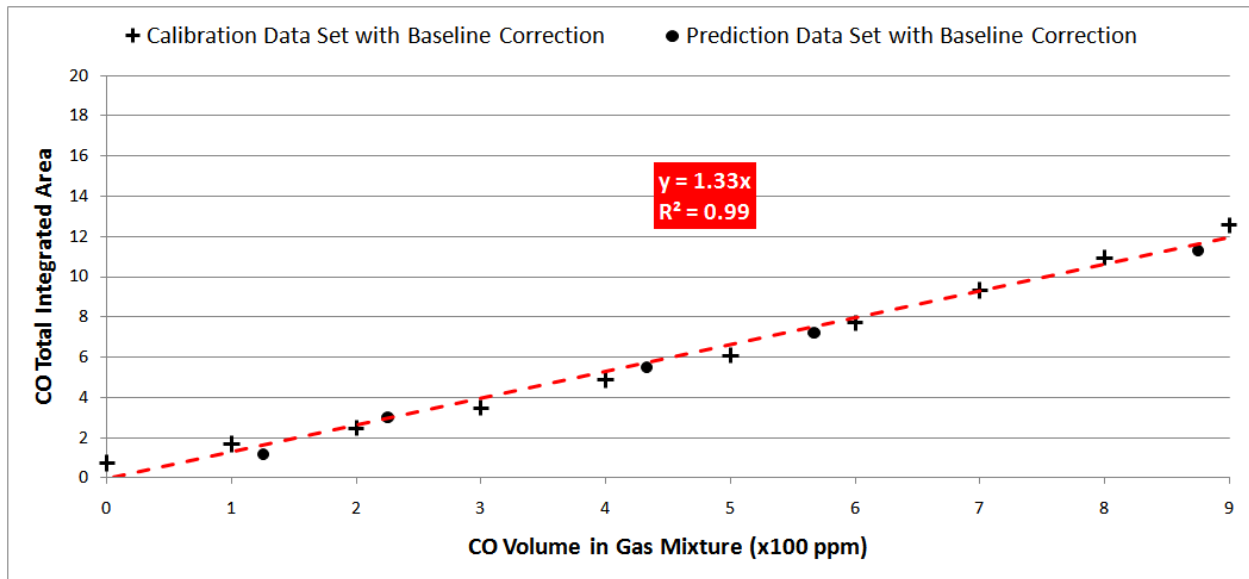


Figure 6-17: PCC CO Calibration and Prediction Data Sets in CO/CO₂ Gas Mixtures; $k_{CO} = 0.75$; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm

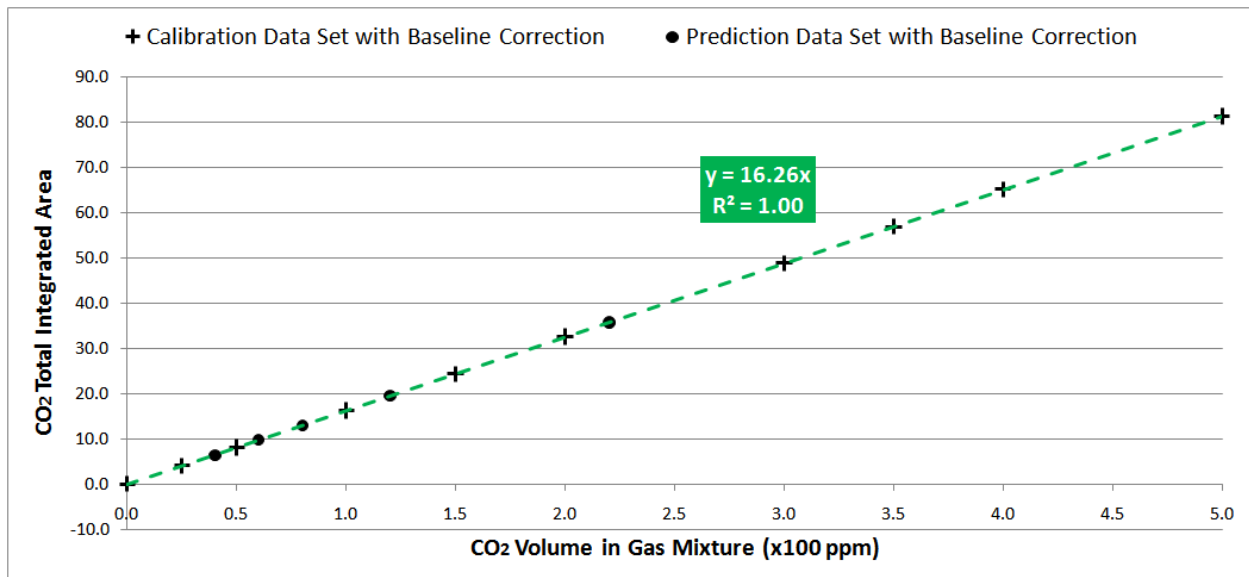


Figure 6-18: PCC CO₂ Calibration and Prediction Data Sets in CO/CO₂ Gas Mixtures; $k_{CO_2} = 0.06$; RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm

The values shown in Table 6-6 highlight that the PCC techniques compares very well with the PCR technique and although not zero for CO, the values are still statistically low enough to indicate PCC technique could be a viable alternative quantitative analysis method to the PCR technique. The PCC technique for CO in this example performs just as well as the PCR technique even with the strong absorption of CO₂ that lowers the impact the CO spectra has on the overall data set variation.

Table 6-6: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; CO/CO₂ Gas Mixtures

Analysis Technique	RMSE Calibration (ppm)	RMSE Prediction (ppm)
CO: PCR	0.0	0.0
CO: PCC	0.3	0.2
CO₂: PCR	0.0	0.0
CO₂: PCC	0.0	0.0

CO/CO₂ Experimental Data Set

From the experimental FTIR spectral analysis of the CO/CO₂ gas mixtures, a set of 8 samples ($n = 8$) at 5,001 different wavenumbers ($p = 5,001$) with concentrations given in Table 6-7 were collected as shown in Figure 6-19 to produce a calibration data set $[X_C]_{(8 \times 5001)}$. A set of 4 samples with concentrations given in Table 6-8 were collected as shown in Figure 6-20 to produce a prediction data set $[X_P]_{(4 \times 5001)}$ for PCR. In the case of the PCC, a single data set $[X]_{(12 \times 5001)}$ consisting of the calibration and prediction data sets were combined and PCA was simultaneously performed on all 12 samples to extract out the pure components in each spectra.

Table 6-7: CO/CO₂ Spectra Compositions for Experimental Calibration Data Set, $[X_C]_{(8 \times 5001)}$

Sample #	Amount of CO (ppm)	Amount of CO₂ (ppm)
1	258.0	3.1
3	653.8	3.3
4	737.8	2.7
5	756.5	6.6
6	735.9	13.1
8	741.2	103.1
10	750.5	178.8
12	777.5	210.6

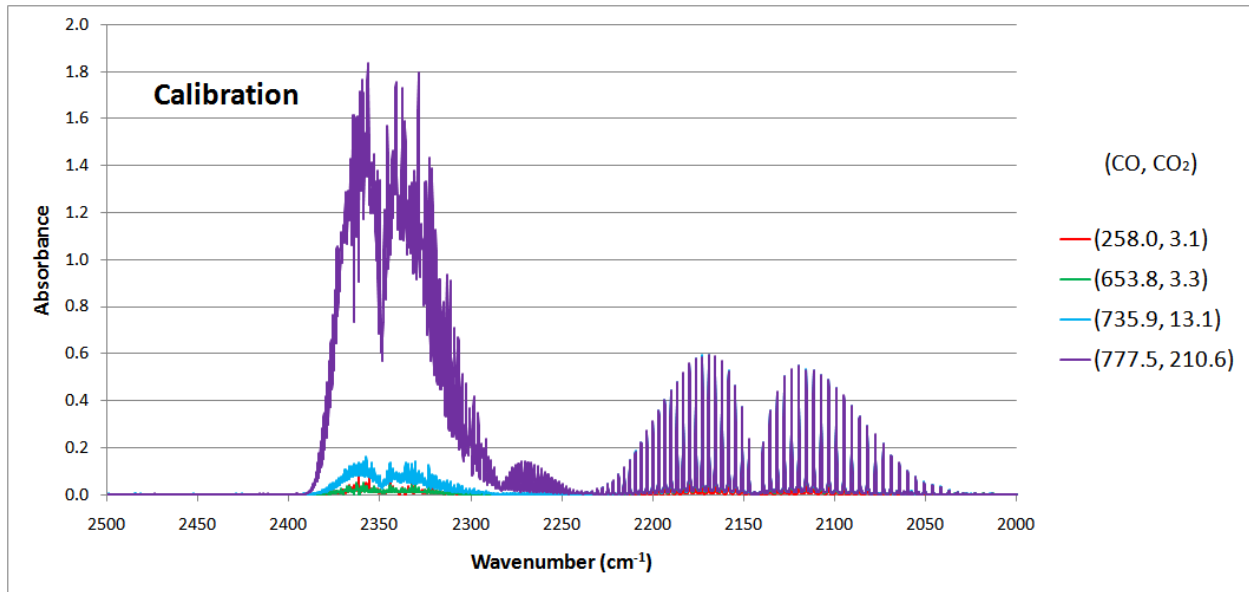


Figure 6-19: CO/CO₂ Gas Mixtures Experimental Calibration Data Set, $[X_C]_{(8 \times 5001)}$; Note – Only 4 of 8 Calibration Spectra Shown

Table 6-8: CO/CO₂ Spectra Compositions for Experimental Prediction Data Set, $[X_P]_{(8 \times 5001)}$

Sample #	Amount of CO (ppm)	Amount of CO ₂ (ppm)
2	441.2	1.4
7	737.8	55.5
9	754.8	127.8
11	759.5	195.2

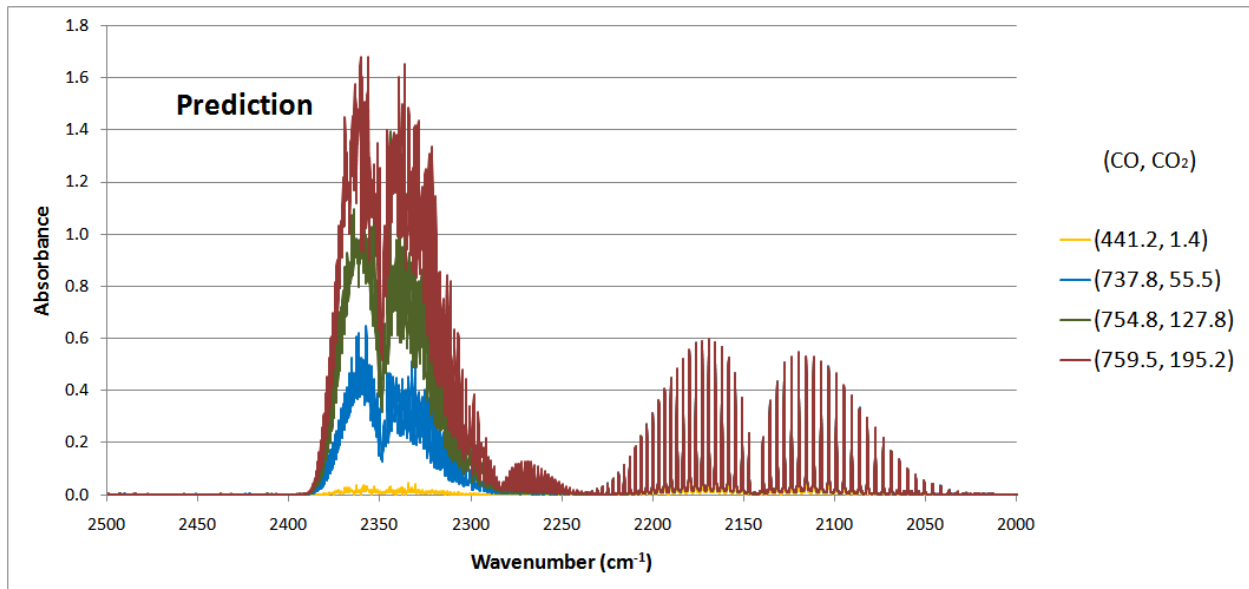


Figure 6-20: CO/CO₂ Gas Mixtures Prediction Data Set, $[X_P]_{(4 \times 5001)}$

As shown in the plot of eigenvalues as a function of principal components, the number of significant eigenvalues obtained from PCA is two (Figure 6-21). The first principal component explains 97.9% of the total calibration data set variance and the second principal component explains 0.9% of the total calibration data set variance. With these two principal components, 98.8% of the original data variance can be explained. This compares well with the explanation of variation that was found with the CO/CO₂ simulation data set, where the principal components corresponding to CO₂ and CO explained 99.1% and 0.9% of the data set variance, respectively. In an unknown mixture, the second component would typically not be analyzed since the first component explains at least 95% of the total data set variance as necessary for an experimental data set. The 95% explanation of the total data set variance level is associated with the typical error of the FTIR sampling procedure, which is 5%.

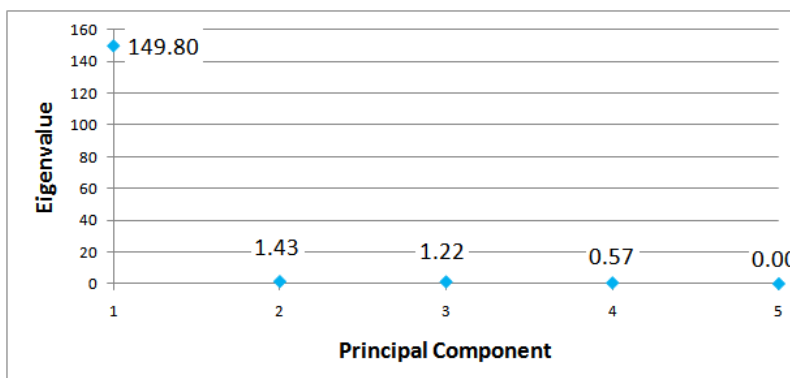


Figure 6-21: SCREE Plot Indicating Eigenvalues for Each Principal Component for CO/CO₂ Gas Mixtures Data Set; Principal Components 1 and 2 Explain 97.9% and 0.9% of the Total Calibration Data Set Variance

Applying PCR, a plot can be created that shows the calibration data set used to determine the concentrations in the prediction data set as shown in Figures 6-22 and 6-23 for CO and CO₂, respectively. With the PCR technique, the RMSEC for CO is found to be 92 ppm, while the

RMSEP for CO is found to be 49 ppm. The RMSEC and RMSEP values for CO₂ are found to be 1 ppm.

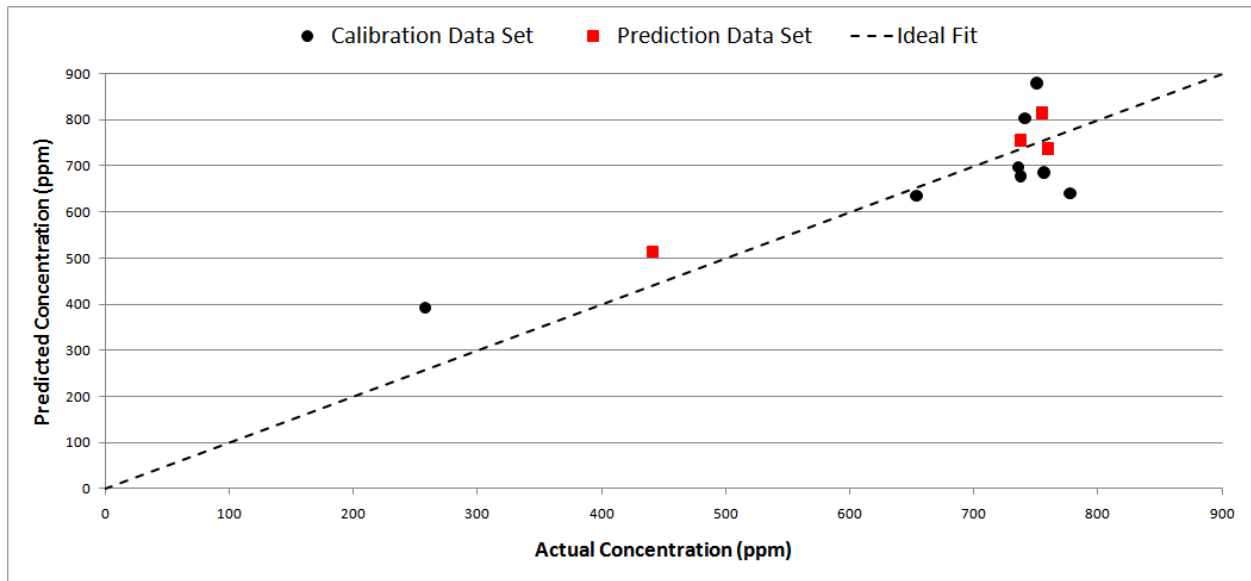


Figure 6-22: Principal Component Regression (PCR) – CO Concentrations in CO/CO₂ Gas Mixtures; RMSE Calibration = 92 ppm, RMSE Prediction = 49 ppm

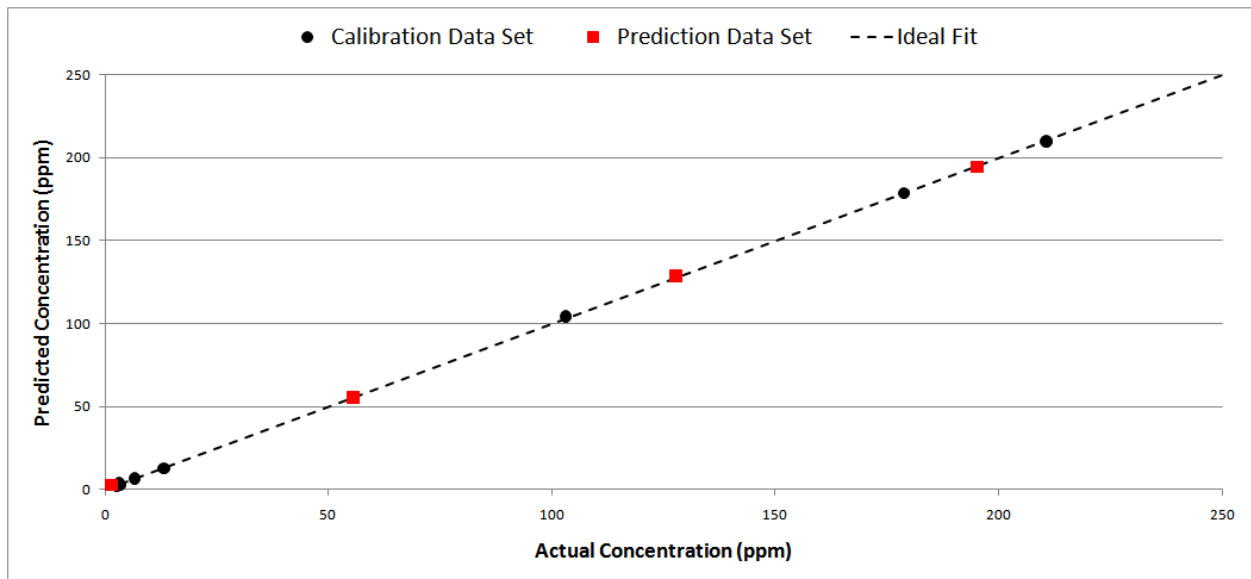


Figure 6-23: Principal Component Regression (PCR) – CO₂ Concentrations in CO/CO₂ Gas Mixtures; RMSE Calibration = 1 ppm, RMSE Prediction = 1 ppm

To improve the PCR analysis for CO, more principal components can be utilized. This is not always possible because in some cases, the representation of redundant data can cause numerical instability within the regression analysis. To show that for this experimental data the PCR technique is improved with the use of more principal components, Figure 6-24 was produced that shows both RMSE for calibration and prediction as a function of increasing number of principal components used to represent the original data set. This analysis indicates with the use of three principal components, which explains 99.6% of the original data variance, the RMSEC is reduced to 16 ppm and the RMSEP is reduced to 14 ppm as shown in Figure 6-25. Further analysis with four principal components, which cumulatively explain 100.0% of the original data variance, the RMSEC is found to be 8 ppm and the RMSEP is found to be 14 ppm as shown in Figure 6-26. The addition of the remaining principal components will not provide additional explanation of the data set variance.

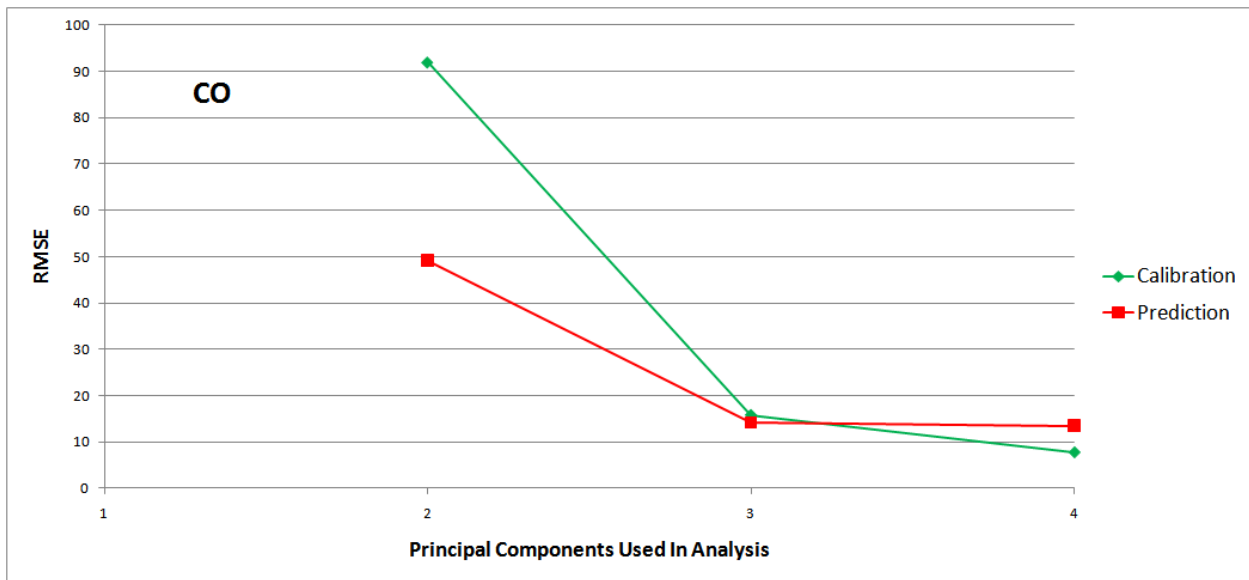


Figure 6-24: CO RMSE Calibration and RMSE Prediction as a Function of the Number of Principal Components Used to Represent the Original Data Set in CO/CO₂ Gas Mixtures

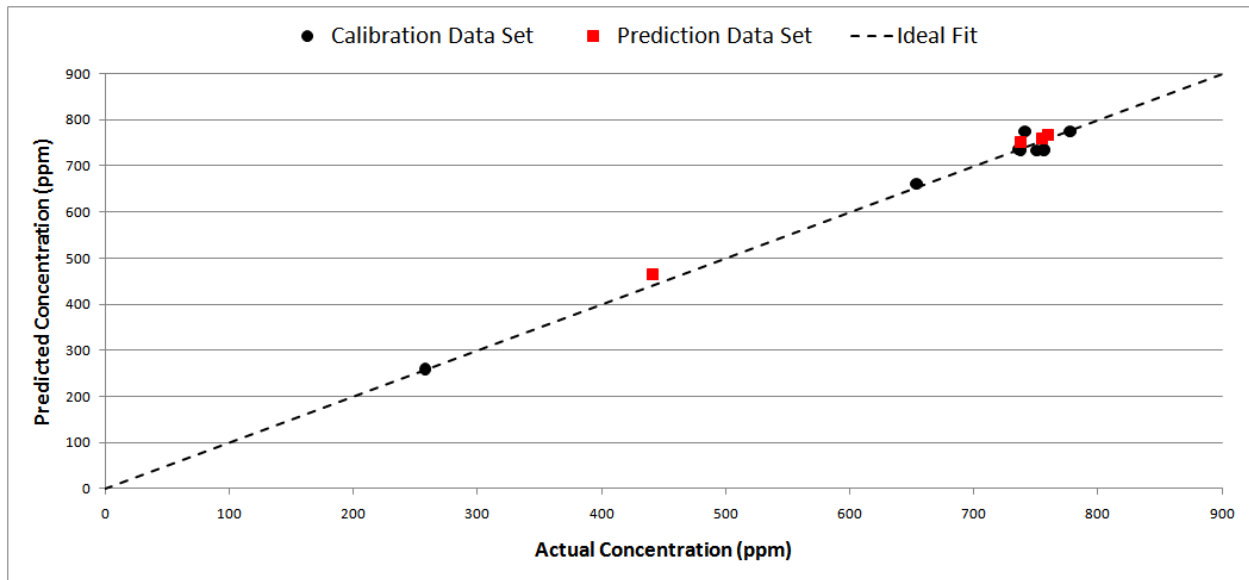


Figure 6-25: Principal Component Regression (PCR) with 3 Principal Components – CO Concentrations in CO/CO₂ Gas Mixtures; RMSE Calibration = 16 ppm, RMSE Prediction = 14 ppm

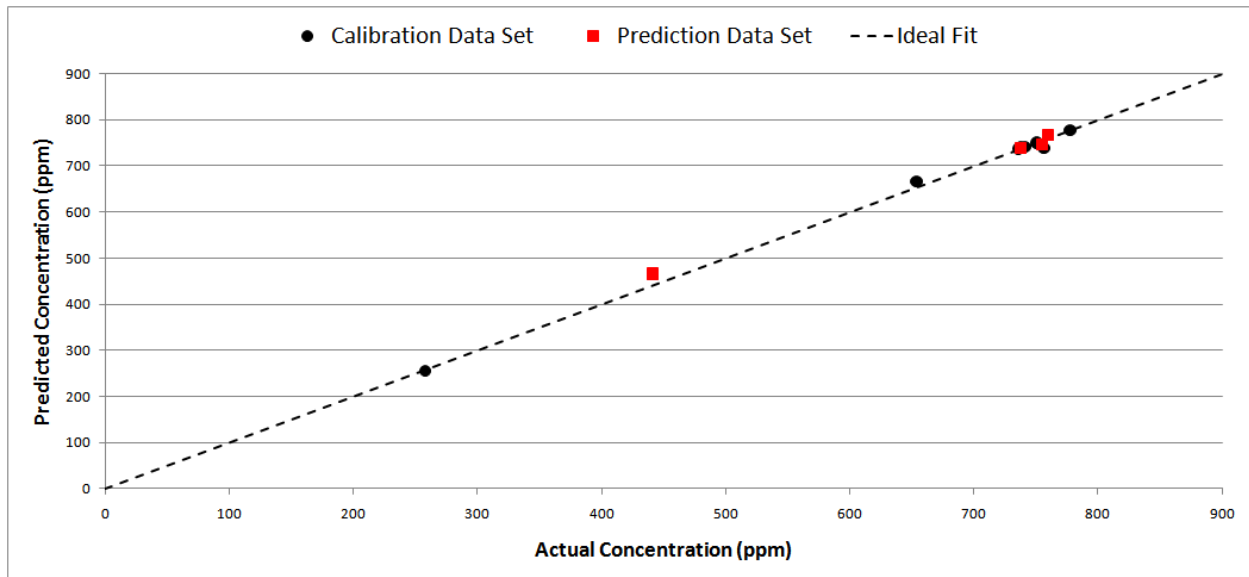


Figure 6-26: Principal Component Regression (PCR) with 4 Principal Components – CO Concentrations in CO/CO₂ Gas Mixtures; RMSE Calibration = 8 ppm, RMSE Prediction = 14 ppm

This analysis indicates that the spectra may contain beneficial data beyond the two principal components that explain 98.8% of the original calibration data set variation. This may be an indication that the variance for the IR spectra of the CO component in the CO/CO₂ mixed gas environment is partitioned among three principal components instead of just one. By utilizing more principal components, the quantification of the CO concentration is improved but there is no evidence to support four distinct components in the experimentally collected IR spectra since the system was purged with IR-inactive N₂ and only CO and CO₂ gas from certified gas cylinders were allowed into the FTIR Gas Analysis Module. The improved in using more principal components to calculated the regression coefficients for CO is highlighted in Figures 6-27 thru 6-29, where the contribution to the CO₂ peak region (2400 cm⁻¹ – 2300 cm⁻¹) is diminished.

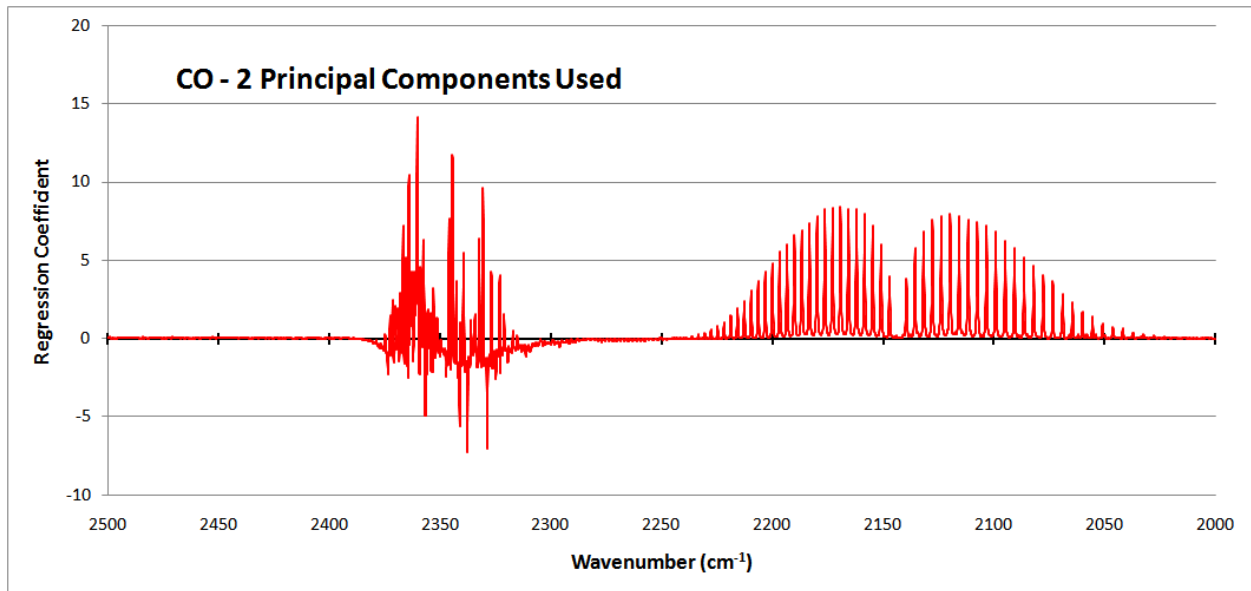


Figure 6-27: CO/CO₂ Gas Mixtures – Estimates of Regression Coefficients for Concentrations of CO from Calibration Data Set as Function of Wavenumber, [b_{CO}]_(5000 x 1); 2 Principal Components Used

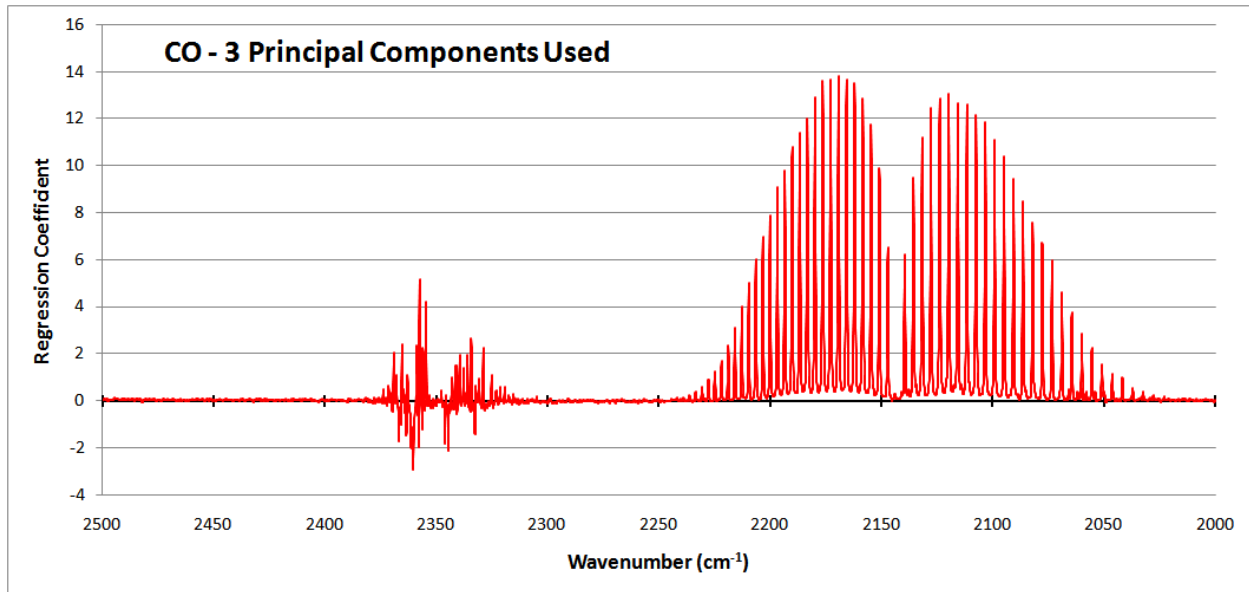


Figure 6-28: CO/CO₂ Gas Mixtures – Estimates of Regression Coefficients for Concentrations of CO from Calibration Data Set as Function of Wavenumber, $[b_{CO}]_{(5000 \times 1)}$; 3 Principal Components Used

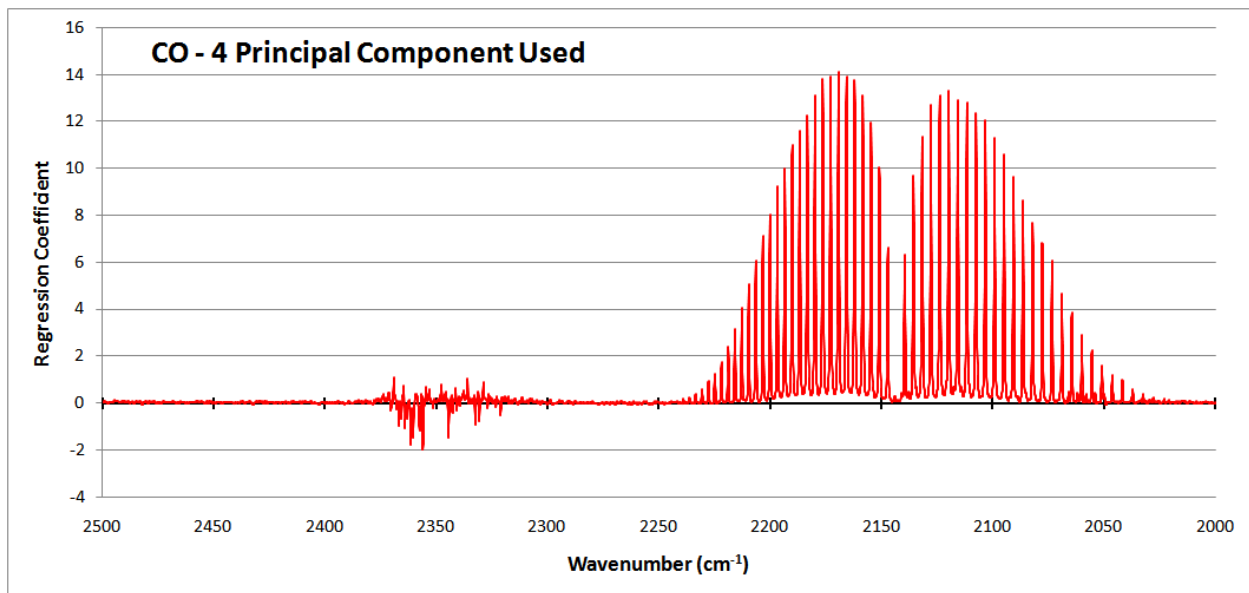


Figure 6-29: CO/CO₂ Gas Mixtures – Estimates of Regression Coefficients for Concentrations of CO from Calibration Data Set as Function of Wavenumber, $[b_{CO}]_{(5000 \times 1)}$; 4 Principal Components Used

Figures 6-30 and 6-31, derived from PCA on the entire set of collected experimental data set spectra, show that *V-2* and *V-1* can be identified as corresponding to the pure spectra of CO and CO₂ and, respectively. With the components separated, the contribution of each component to the sample mixtures can be determined for the CO calibration and prediction data sets, and for CO₂ calibration and prediction data set.

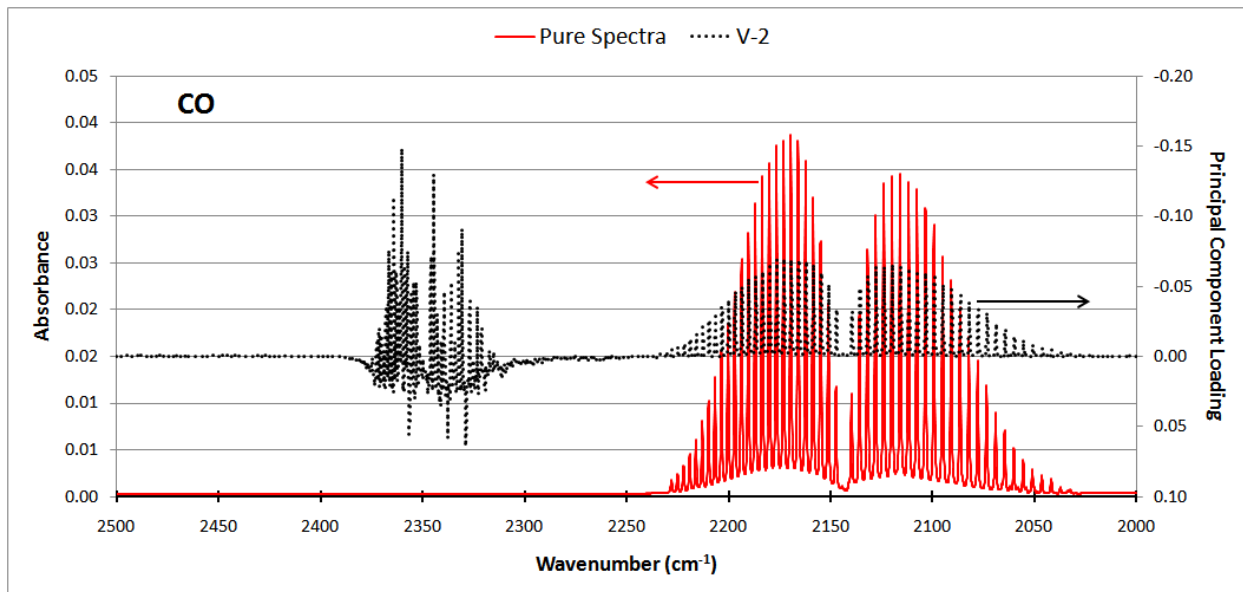


Figure 6-30: PCA Separated CO Spectra in CO/CO₂ Gas Mixtures Represented by *V-2*

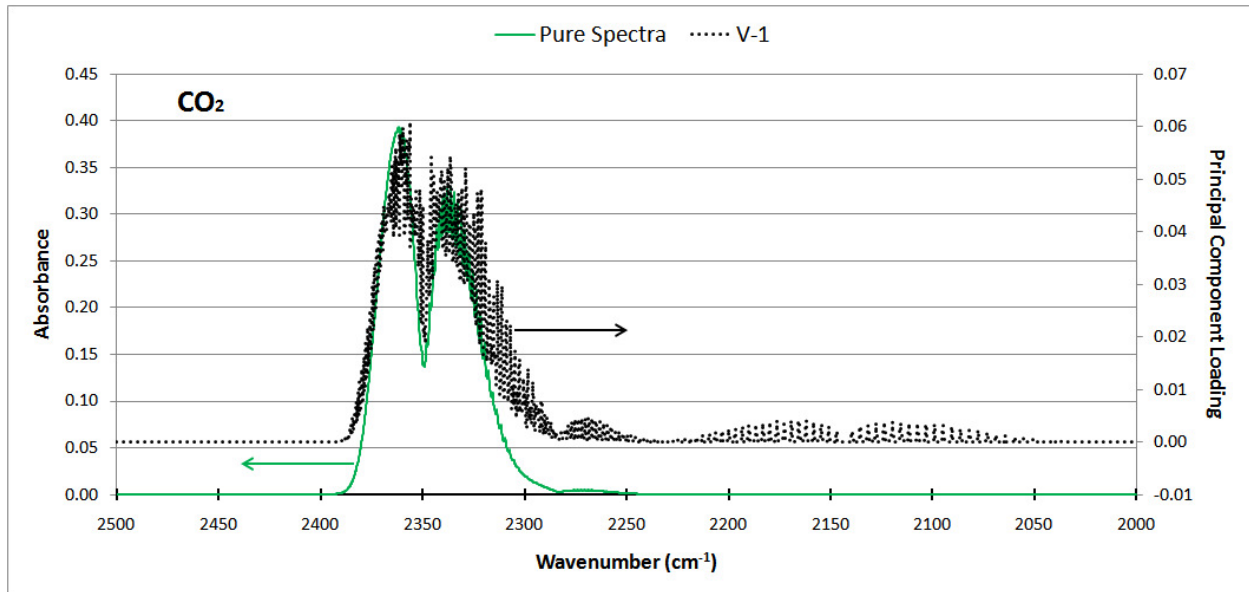


Figure 6-31: PCA Separated CO₂ Spectra in CO/CO₂ Gas Mixtures Represented by *V-1*

The calculated proportionality constants ($k_{\text{CO}} = 20$, $k_{\text{CO}_2} = 0.25$), equal to the inverse of the slope of the lines of best fit shown, are used to calculate the concentration of the mixtures for CO and CO₂ shown in Figures 6-32 and 6-33, respectively. With the PCC method, the RMSE for CO and CO₂ calibration data sets are found to be 204 ppm and 1 ppm, respectively, while the RMSE of the prediction data sets for CO and CO₂ are found to be 194 ppm and 1 ppm, respectively.

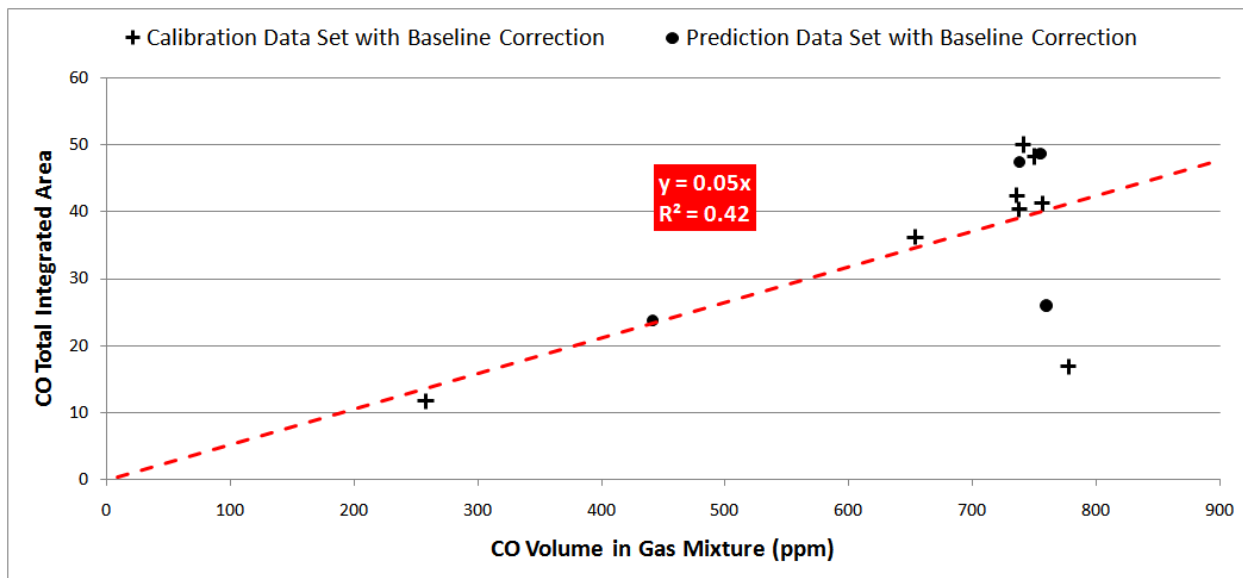


Figure 6-32: PCC CO Calibration and Prediction Data Sets in CO/CO₂ Gas Mixtures; $k_{CO} = 20$;

RMSE Calibration = 204 ppm, RMSE Prediction = 194 ppm

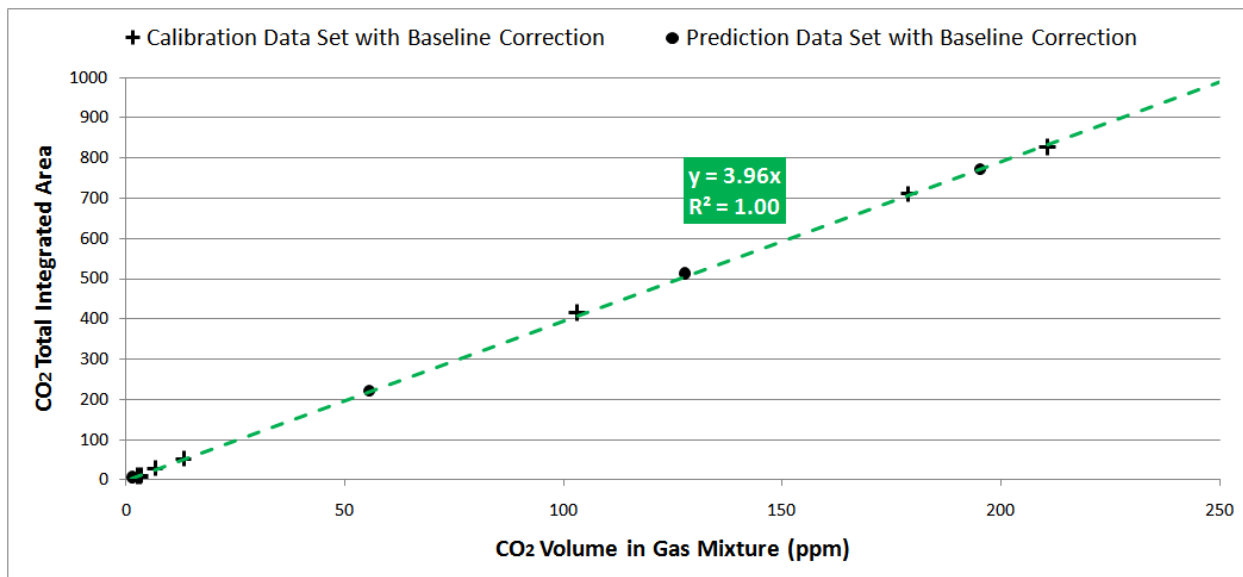


Figure 6-33: PCC CO₂ Calibration and Prediction Data Sets in CO/CO₂ Gas Mixtures; $k_{CO_2} = 0.25$; RMSE

Calibration = 1 ppm, RMSE Prediction = 1 ppm

As shown in Table 6-9 the values for CO₂ concentration calculations compare very well between the PCR and PCC techniques. The analysis using PCR is improved in quantitatively determining the CO concentration in the CO/CO₂ experimental gas mixtures by utilizing additional principal components. PCC for CO in this example does not perform well at all due to the strong absorption of CO₂ that lowers the impact the CO spectra has on the overall data set variation and thus makes it more difficult for the PCA process to extract the pure CO spectra using only 2 principal components. Because of this, the regression calculations for PCR and the calculation for the area under the curve for PCC, a significant amount of error is introduced in determining the CO concentration.

Table 6-9: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; CO/CO₂ Experimental Gas Mixtures

Analysis Technique	# of Principal Components Used in Analysis	RMSE Calibration (ppm)	RMSE Prediction (ppm)
CO: PCR	2	92	49
	3	16	14
	4	8	14
CO: PCC	2	204	194
CO ₂ : PCR	2	1	1
CO ₂ : PCC	2	1	1

PCA Results and Discussions: 3-Component Systems

CH₂O/C₃H₄O/H₂O Simulation Data Set

From the simulated FTIR spectral analysis of the pure CH₂O/C₃H₄O/H₂O gas mixtures shown in Figure 6-34, a set of 8 samples ($n = 8$) at 2,655 different wavenumbers ($p = 2,655$) with concentrations given in Table 6-10 were combined as shown in Figure 6-35 to produce a calibration data set $[X_C]_{(8 \times 2655)}$. A set of 3 samples with concentrations given in Table 6-11 were combined as shown in Figure 6-36 to produce a prediction data set $[X_P]_{(3 \times 2655)}$ for PCR. In the case of the PCC, a single data set $[X]_{(11 \times 2655)}$ consisting of the calibration and prediction data

sets were combined and PCA was simultaneously performed on all 11 samples to extract out the pure components in each spectra.

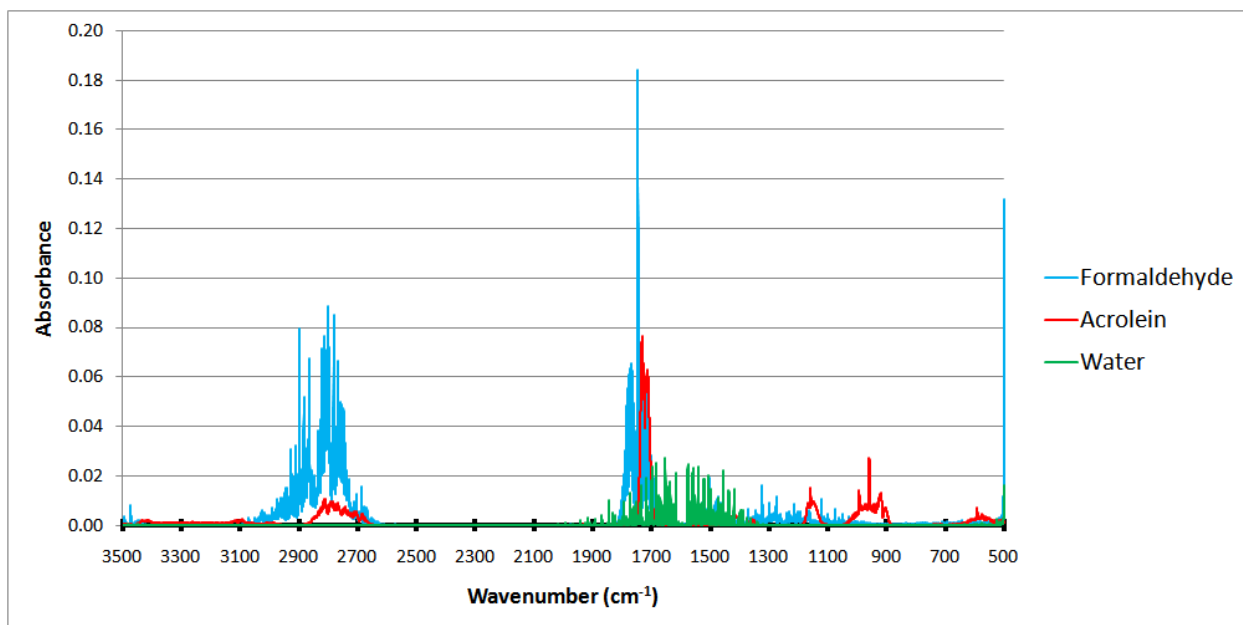


Figure 6-34: Formaldehyde (CH₂O), Acrolein (C₃H₄O), and Water (H₂O) Pure Spectra for Simulated Data Sets
 Illustrating Spectral Overlap Between All Three Components

Table 6-10: CH₂O/C₃H₄O/H₂O Spectra Compositions for Calibration Data Set, [X_C]_(8 x 2655), Ordered from Low to High Concentration of CH₂O

Sample #	Amount of CH ₂ O (x100 ppm)	Amount of C ₃ H ₄ O (x100 ppm)	Amount of H ₂ O (x100 ppm)
1	0.00	2.50	0.00
2	0.50	1.00	50.00
3	1.00	0.50	1.00
4	1.50	2.00	25.00
5	2.00	1.50	40.00
6	3.00	4.75	33.30
7	4.75	3.00	20.00
8	10.00	10.00	5.00

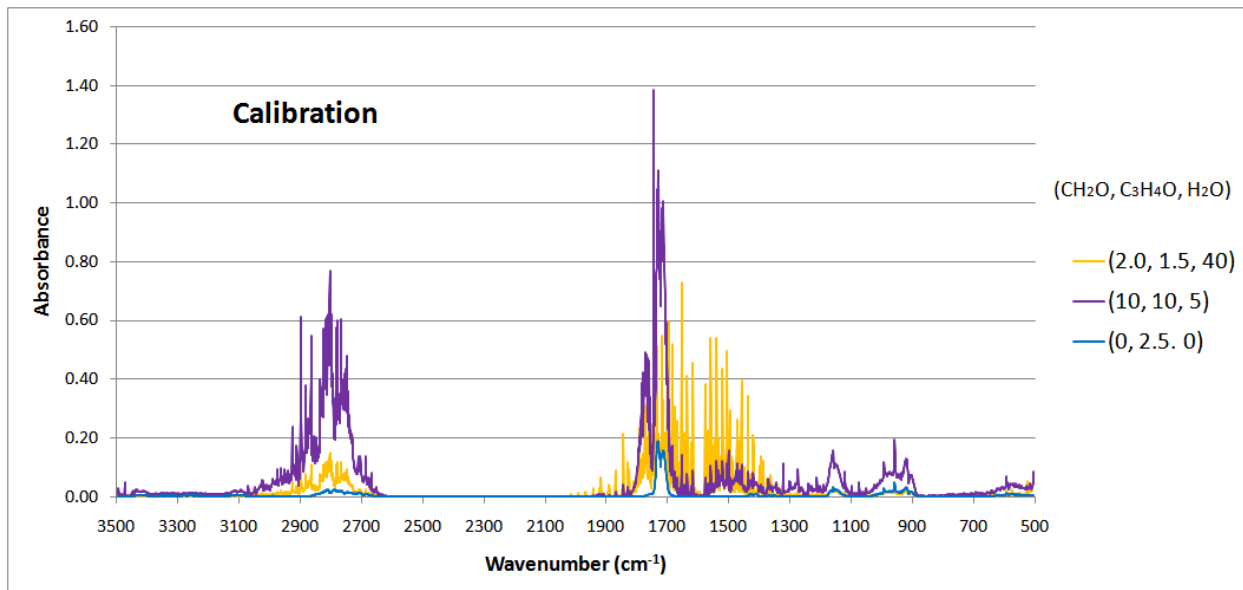


Figure 6-35: CH₂O/C₃H₄O/H₂O Gas Mixtures Calibration Data Set, [X_C]_(8 x 2655); Note – Only 3 of 8 Calibration Spectra Shown

Table 6-11: CH₂O/C₃H₄O/H₂O Spectra Compositions for Prediction Data Set, [X_P]_(3 x 2655), Ordered from Low to High Concentration of CH₂O

Sample #	Amount of CH ₂ O (x100 ppm)	Amount of C ₃ H ₄ O (x100 ppm)	Amount of H ₂ O (x100 ppm)
9	1.33	8.00	30.00
10	2.50	0.00	10.00
11	8.00	1.33	6.67

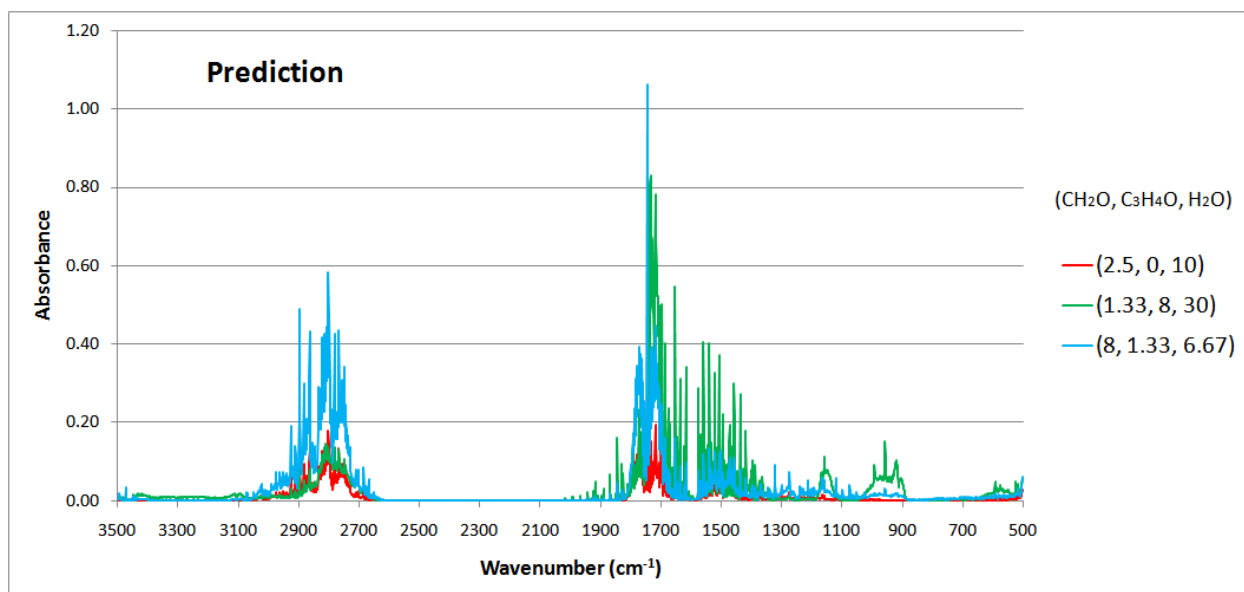


Figure 6-36: CH₂O/C₃H₄O/H₂O Gas Mixtures Prediction Data Set, [X_P]_(3 x 2655)

As shown in the plot of eigenvalues as a function of principal components, the number of significant eigenvalues obtained from PCA is three (Figure 6-37). The first principal component explains 71.9% of the total calibration data set variance, the second principal component explains of 26.9%, and the third principal component explains 1.2% of the total calibration data set variance. With these three principal components, 100.0% of the original data variance can be explained.

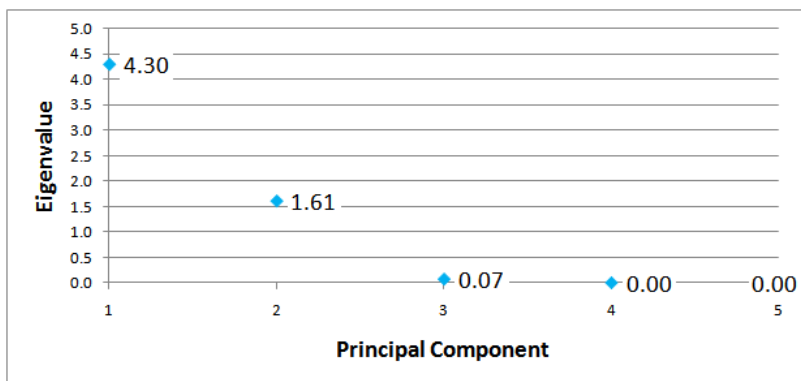


Figure 6-37: SCREE Plot Indicating Eigenvalues for Each Principal Component for CH₂O/C₃H₄O/H₂O Gas Mixtures Data Set; Principal Components 1, 2, and 3 Explain 71.9%, 26.9%, and 1.2%, respectively, of the Total Calibration Data Set Variance

Applying PCR, a plot can be created that shows the calibration data set used to determine the concentrations in the prediction data set as shown in Figures 6-38, 6-39, and 6-40 for CH₂O, H₂O, and C₃H₄O, respectively. With the PCR technique, the RMSE for the calibration and prediction data sets in regards to concentration of CH₂O, H₂O, and C₃H₄O are found to be 0 ppm.

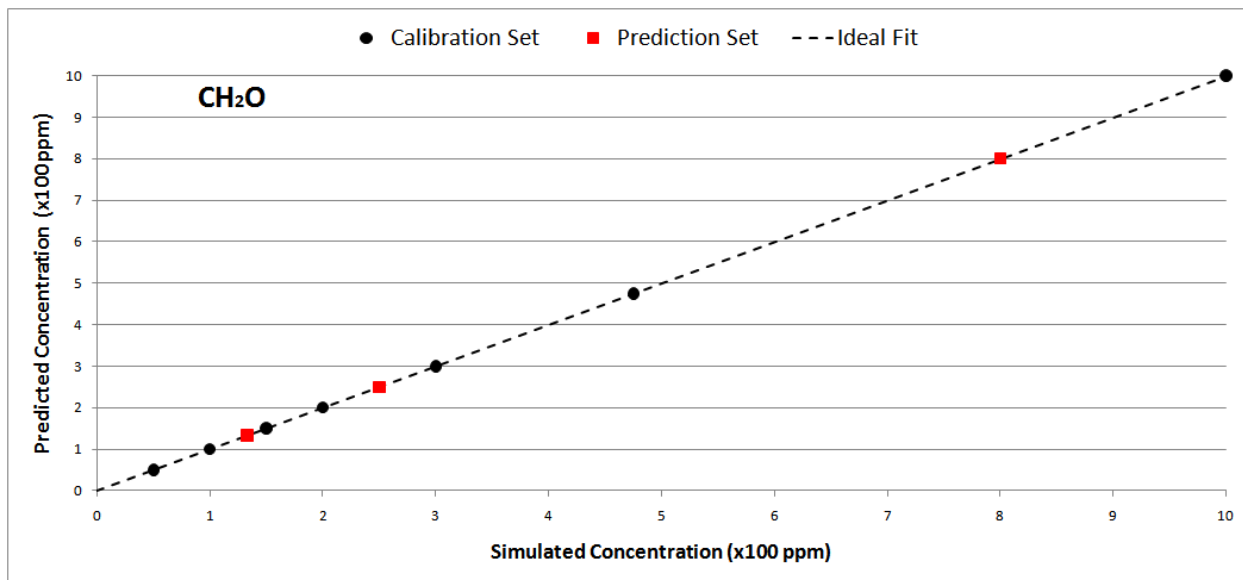


Figure 6-38: Principal Component Regression (PCR) – CH₂O Concentrations in CH₂O/C₃H₄O/H₂O Gas Mixtures;
 RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm

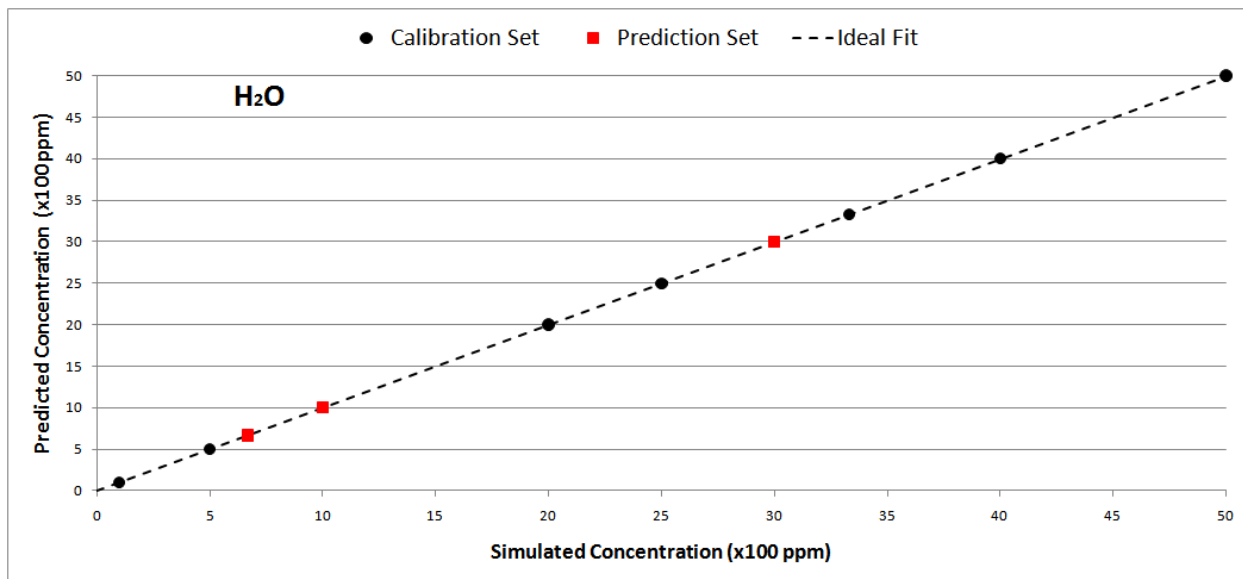


Figure 6-39: Principal Component Regression (PCR) – H₂O Concentrations in CH₂O/C₃H₄O/H₂O Gas Mixtures;
 RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm

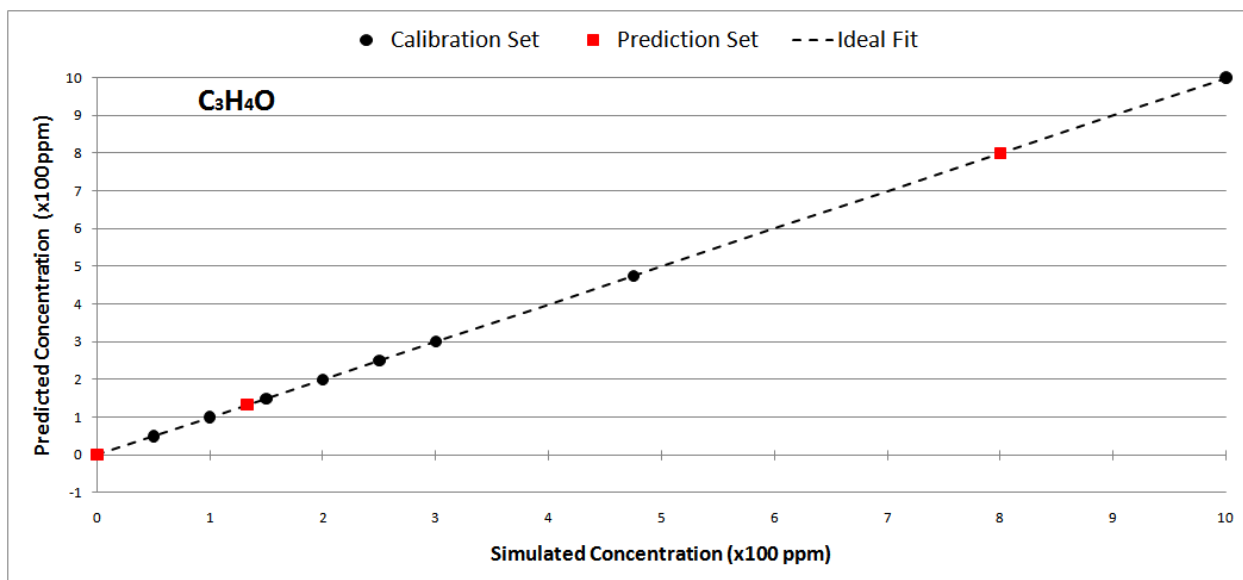


Figure 6-40: Principal Component Regression (PCR) – C₃H₄O Concentrations in CH₂O/C₃H₄O/H₂O Gas Mixtures;
 RMSE Calibration = 0 ppm, RMSE Prediction = 0 ppm

PCA on the entire set of collected spectra isolates the peaks associated with each particular component as shown in Figures 6-41, 6-42, and 6-43. The variables *V-1*, *V-2*, and *V-3* can be easily identifiable as corresponding to the pure spectra of CH₂O, H₂O, and C₃H₄O and, respectively. With the components separated, the contribution of each component to the sample mixtures can be determined for the CH₂O calibration and prediction data sets, for the H₂O calibration and prediction data sets, and for the C₃H₄O calibration and prediction data set.

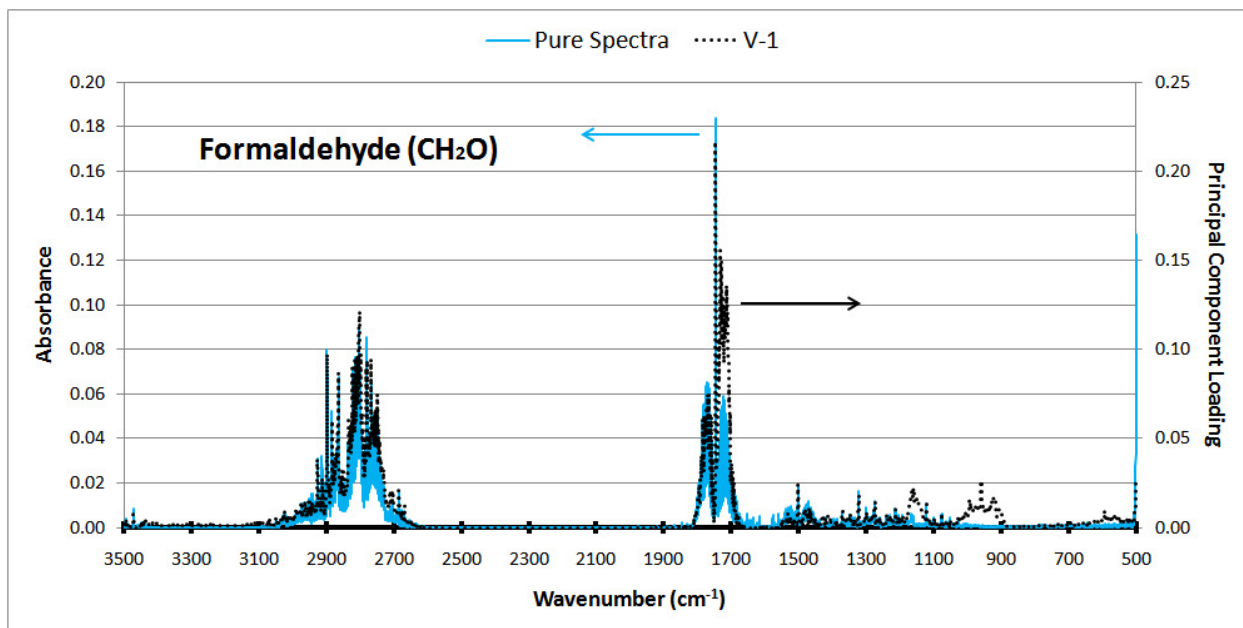


Figure 6-41: PCA Separated CH₂O Spectra in CH₂O/C₃H₄O/H₂O Gas Mixtures

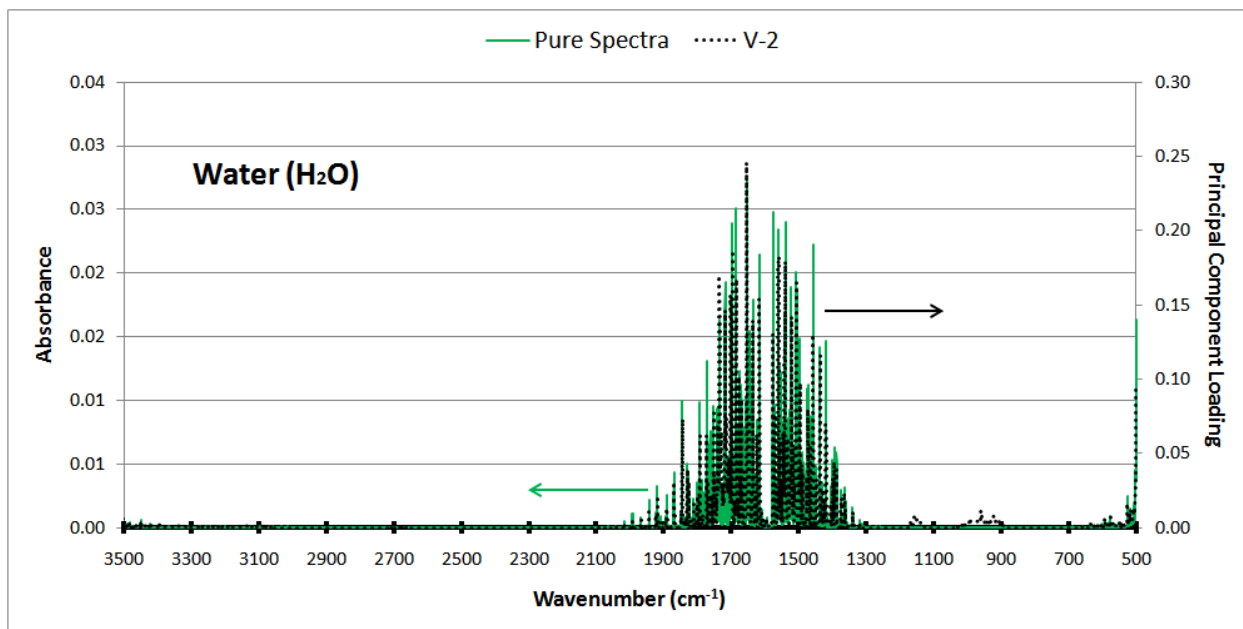


Figure 6-42: PCA Separated H₂O Spectra in CH₂O/C₃H₄O/H₂O Gas Mixtures

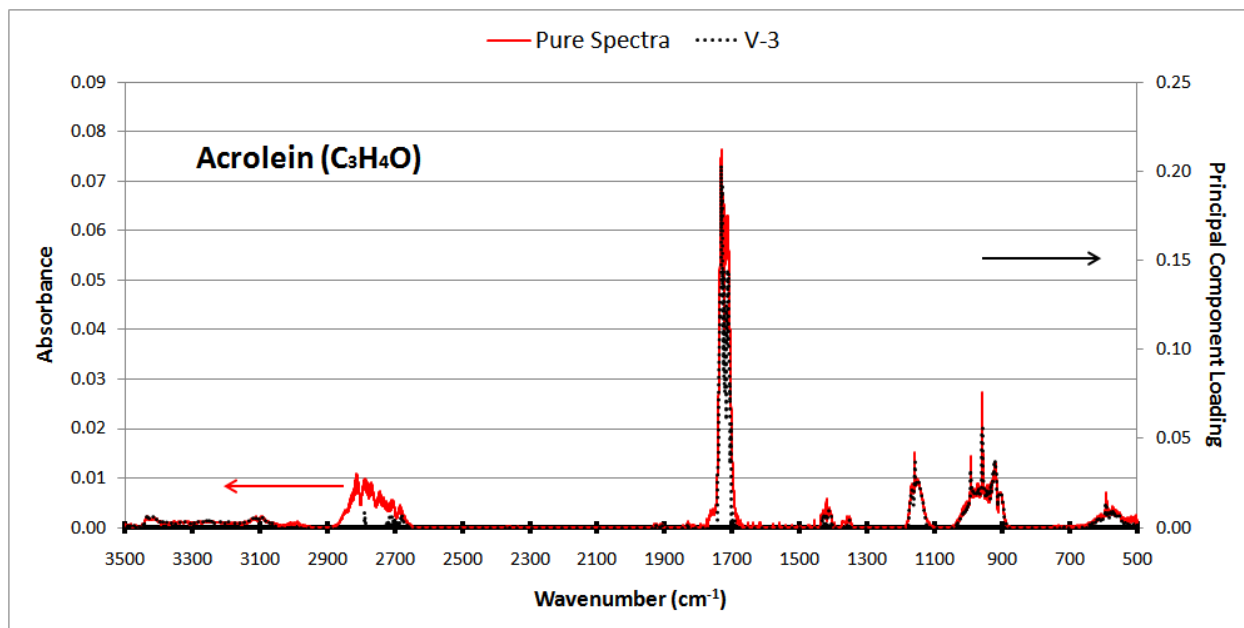


Figure 6-43: PCA Separated C₃H₄O Spectra in CH₂O/C₃H₄O/H₂O Gas Mixtures

The calculated proportionality constants ($k_{\text{CH}_2\text{O}} = 0.08$, $k_{\text{H}_2\text{O}} = 1.25$, $k_{\text{C}_3\text{H}_4\text{O}} = 0.39$) are used to calculate the concentration of the mixtures for CH₂O, H₂O, and C₃H₄O shown in Figures 6-44, 6-45 and 6-46, respectively. With the PCC method, the RMSE for CH₂O, H₂O, and C₃H₄O calibration data sets are found to be 1 ppm, 9 ppm, and 1 ppm, respectively, while the RMSE of the prediction data sets for CH₂O, H₂O, and C₃H₄O are found to be 2 ppm, 6 ppm, and 1 ppm, respectively.

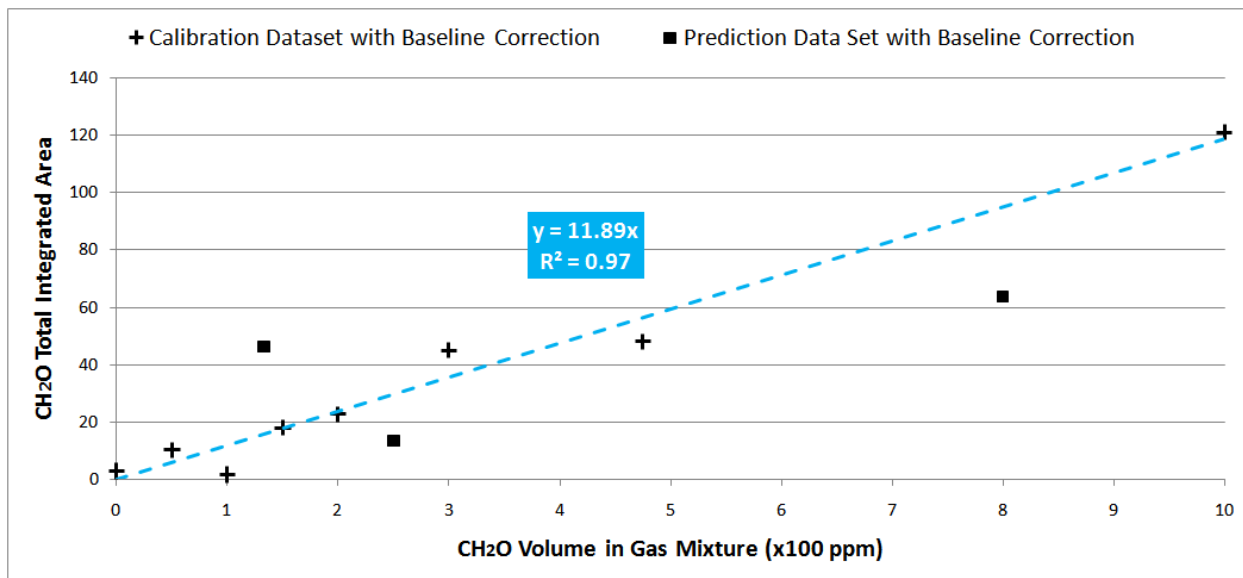


Figure 6-44: PCC CH₂O Calibration and Prediction Data Sets in CH₂O/C₃H₄O/H₂O Gas Mixtures; $k_{\text{CH}_2\text{O}} = 0.08$;
 RMSE Calibration = 1 ppm, RMSE Prediction = 2 ppm

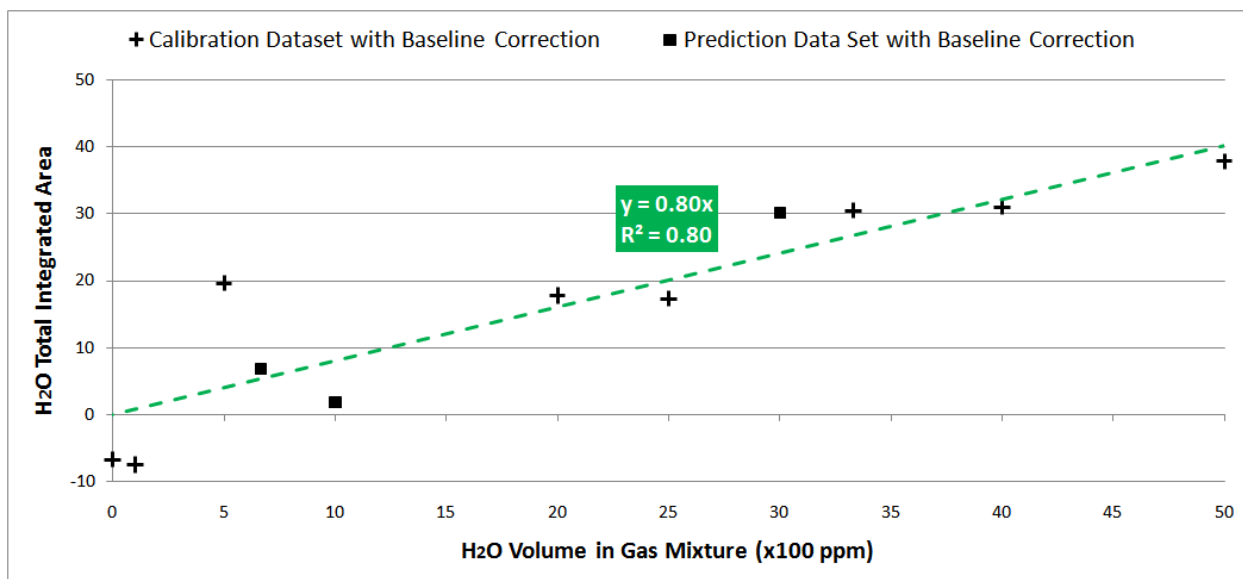


Figure 6-45: PCC H₂O Calibration and Prediction Data Sets in CH₂O/C₃H₄O/H₂O Gas Mixtures; $k_{\text{H}_2\text{O}} = 1.25$;
 RMSE Calibration = 9 ppm, RMSE Prediction = 6 ppm

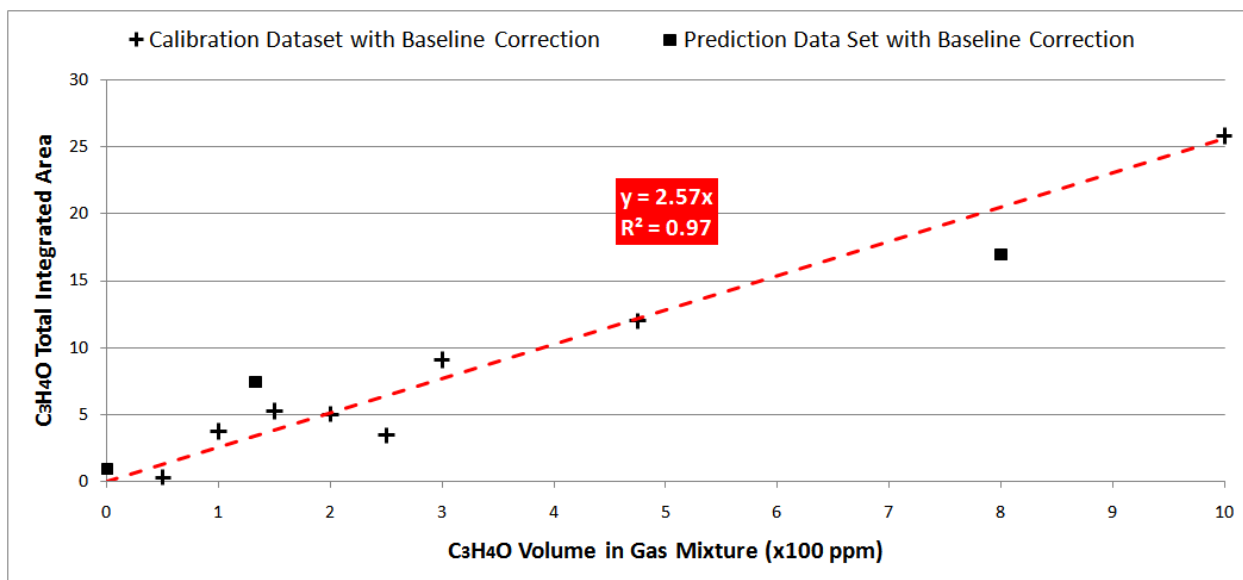


Figure 6-46: PCC C₃H₄O Calibration and Prediction Data Sets in CH₂O/C₃H₄O/H₂O Gas Mixtures; $k_{C_3H_4O} = 0.39$;
 RMSE Calibration = 1 ppm, RMSE Prediction = 1 ppm

As shown in Table 6-12, the comparison of the PCR and PCC techniques indicate that the PCC technique could be a viable alternative quantitative analysis method to the PCR technique. The PCC technique for this example does not perform as well as PCR due to the strong overlap of the two components that make the separation of the three spectra more difficult for PCA. The PCA representation for H₂O still contains small amounts of spectra that are due to CH₂O and C₃H₄O and thus the calculated for the area under the curve for each concentration introduces error within the PCC technique. In addition, similar to the two-component CH₂O/C₃H₄O gas mixtures, the PCA representation for C₃H₄O does not contain the entire spectra for the wavenumber range of 2900-2700 cm⁻¹, where there is a strong overlap of the CH₂O spectra.

Table 6-12: Comparison of Errors Associated with the Principal Component Regression (PCR) and Proportionality Constant Calculation (PCC) Analysis Techniques; CH₂O/C₃H₄O Gas Mixtures

Analysis Technique	RMSE Calibration (ppm)	RMSE Prediction (ppm)
CH ₂ O: PCR	0	0
CH ₂ O: PCC	1	2
H ₂ O: PCR	0	0
H ₂ O: PCC	9	6
C ₃ H ₄ O: PCR	0	0
C ₃ H ₄ O: PCC	1	1

CO/CO₂/H₂O Experimental Data Set

From the experimental FTIR spectral analysis of the CO/CO₂/H₂O gas mixtures, a set of 8 samples ($n = 8$) at 1,001 different wavenumbers ($p = 1,001$) with concentrations given in Table 6-13 were collected as shown in Figure 6-47 to produce a calibration data set $[X_C]_{(8 \times 1001)}$. A set of 4 samples with concentrations given in Table 6-14 were collected as shown in Figure 6-48 to produce a prediction data set $[X_P]_{(4 \times 1001)}$ for PCR. In the case of the PCC, a single data set $[X]_{(12 \times 1001)}$ consisting of the calibration and prediction data sets were combined and PCA was simultaneously performed on all 12 samples to extract out the pure components in each spectra.

Table 6-13: CO/CO₂/H₂O Spectra Compositions for Calibration Data Set, $[X_C]_{(8 \times 1001)}$

Sample #	Amount of CO (ppm)	Amount of CO ₂ (ppm)	Amount of H ₂ O (ppm)
1	91	1	1376
2	113	1	2454
3	96	2	4366
4	90	3	5385
9	71	15	6771
10	25	18	6675
11	34	23	6704
12	8	28	8231

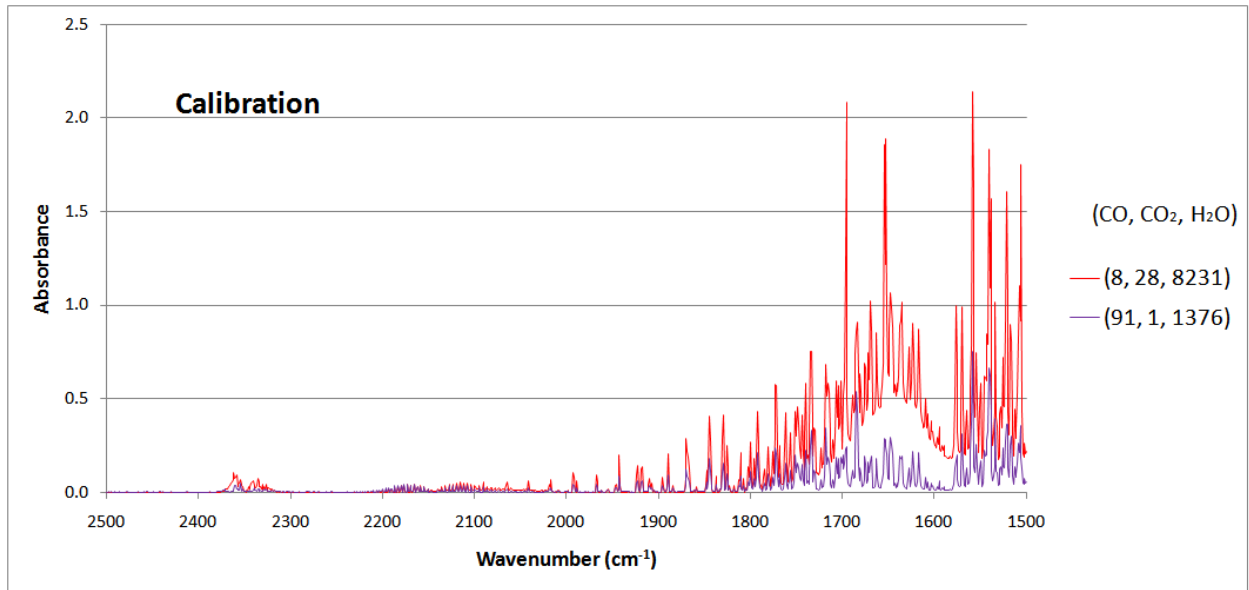


Figure 6-47: CO/CO₂/H₂O Gas Mixtures Calibration Data Set, [X_C]_(8 x 1001); Note – Only 2 of 8 Calibration Spectra Shown

Table 6-14: CO/CO₂/H₂O Spectra Compositions for Prediction Data Set, [X_P]_(4 x 1001)

Sample #	Amount of CO (ppm)	Amount of CO ₂ (ppm)	Amount of H ₂ O (ppm)
5	108	5	5059
6	80	5	5792
7	58	7	5734
8	46	10	6157

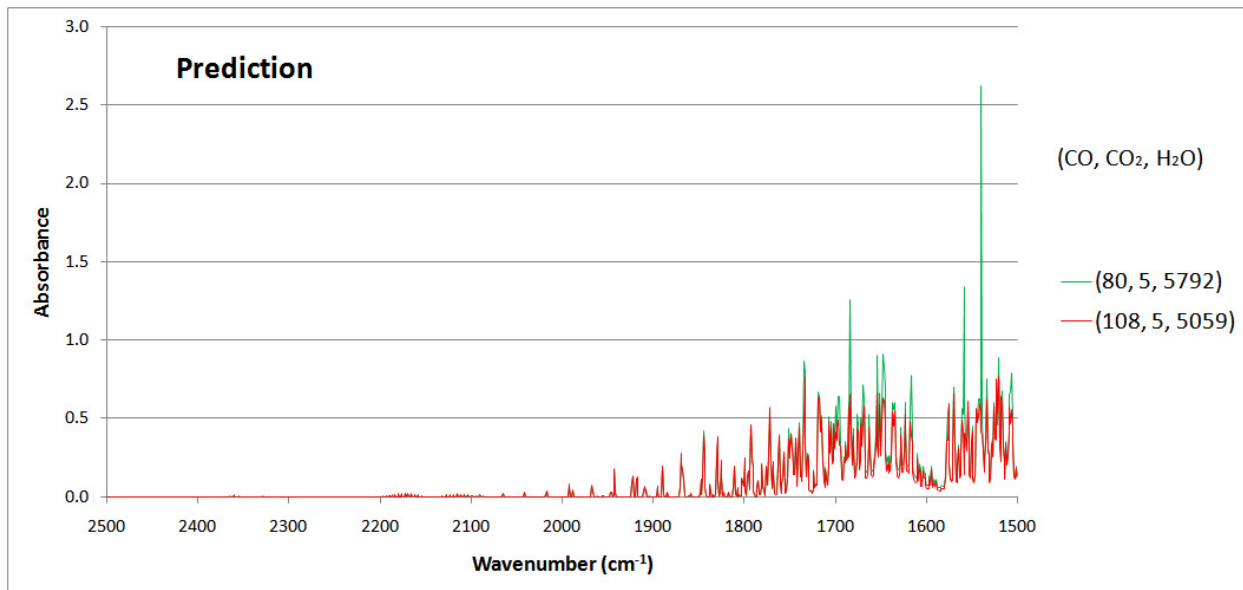


Figure 6-48: CO/CO₂/H₂O Gas Mixtures Prediction Data Set, $[X_P]_{(4 \times 1001)}$; Note – Only 2 of 4 Prediction Spectra Shown

From the plot of eigenvalues as a function of principal components, the number of significant eigenvalues obtained from PCA is determined to be five (Figure 6-49). The first principal component explains 69.7% of the total calibration data set variance, the second principal component explains 12.1%, the third explains 7.7% of the total calibration data set variance, the fourth explains 4.4%, and the fifth explains 3.0%. With these five principal components, 96.9% of the original data variance can be explained.

The analysis indicates that the spectra may contain beneficial data beyond the three principal components expected in the system that explain 89.5% of the original calibration data set variation. This may be an indication that the variance for the IR spectra of individual components in the CO/CO₂/H₂O mixed gas environment is partitioned among multiple principal components instead of just one. By utilizing more principal components, the quantification of the gas concentrations for each component may be improved but there is no evidence to support five distinct components in the experimentally collected IR spectra.

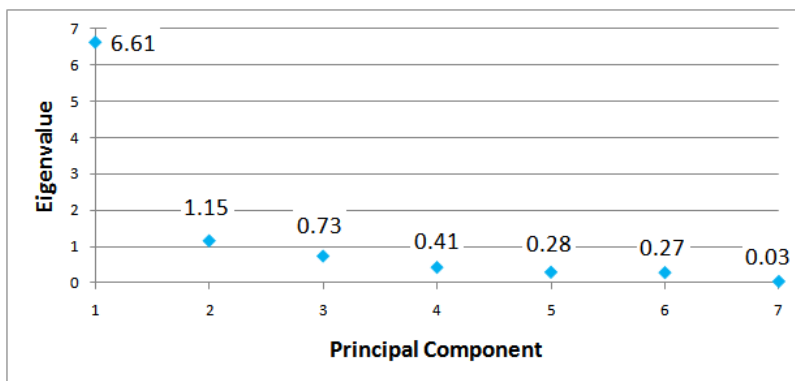


Figure 6-49: SCREE Plot Indicating Eigenvalues for Each Principal Component for CO/CO₂/H₂O Gas Mixtures Data Set; Principal Components 1, 2, and 3 Explain 69.7%, 12.1%, and 7.7% of the Total Calibration Data Set Variance

Applying PCR, a plot can be created that shows the calibration data set used to determine the concentrations in the prediction data set as shown in Figures 6-50, 6-51, and 6-52 for CO, CO₂, and H₂O, respectively. With the PCR technique, the RMSE for calibration for CO is found to be 14 ppm, while the RMSE for prediction for CO is found to be 18 ppm. The RMSEC for CO₂ is found to be 2 ppm, while the RMSEP for CO₂ is found to be 4 ppm. The RMSEC for H₂O is found to be 157 ppm, while the RMSEP for H₂O is found to be 519 ppm.

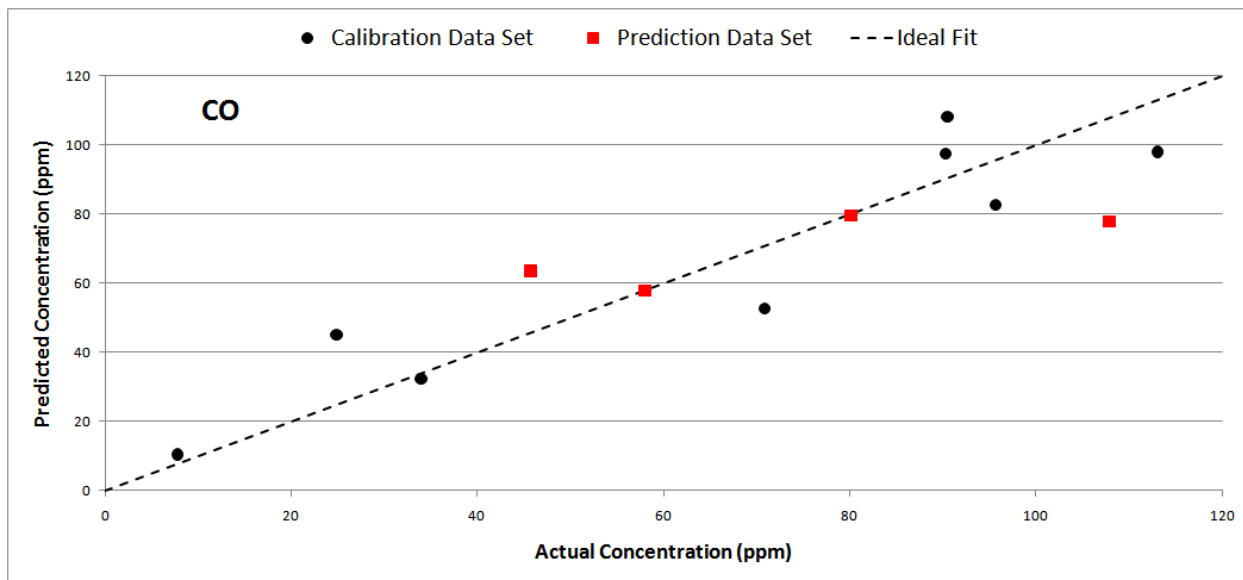


Figure 6-50: Principal Component Regression (PCR) – CO Concentrations in CO/CO₂/H₂O Gas Mixtures; RMSE Calibration = 14 ppm, RMSE Prediction = 18 ppm

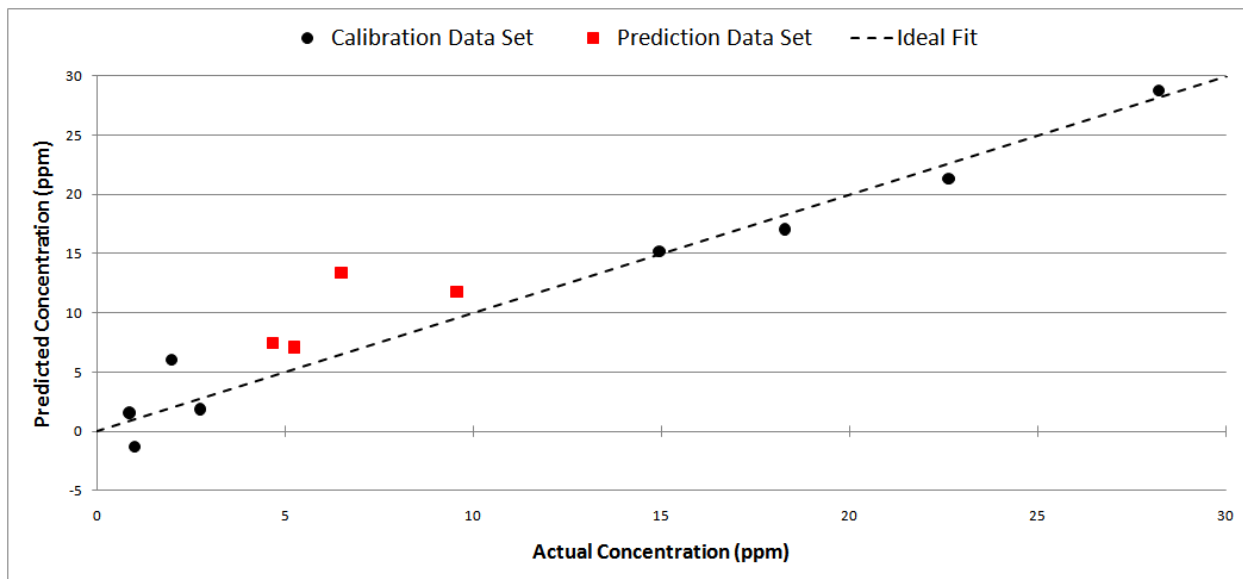


Figure 6-51: Principal Component Regression (PCR) – CO₂ Concentrations in CO/CO₂/H₂O Gas Mixtures; RMSE Calibration = 2 ppm, RMSE Prediction = 4 ppm

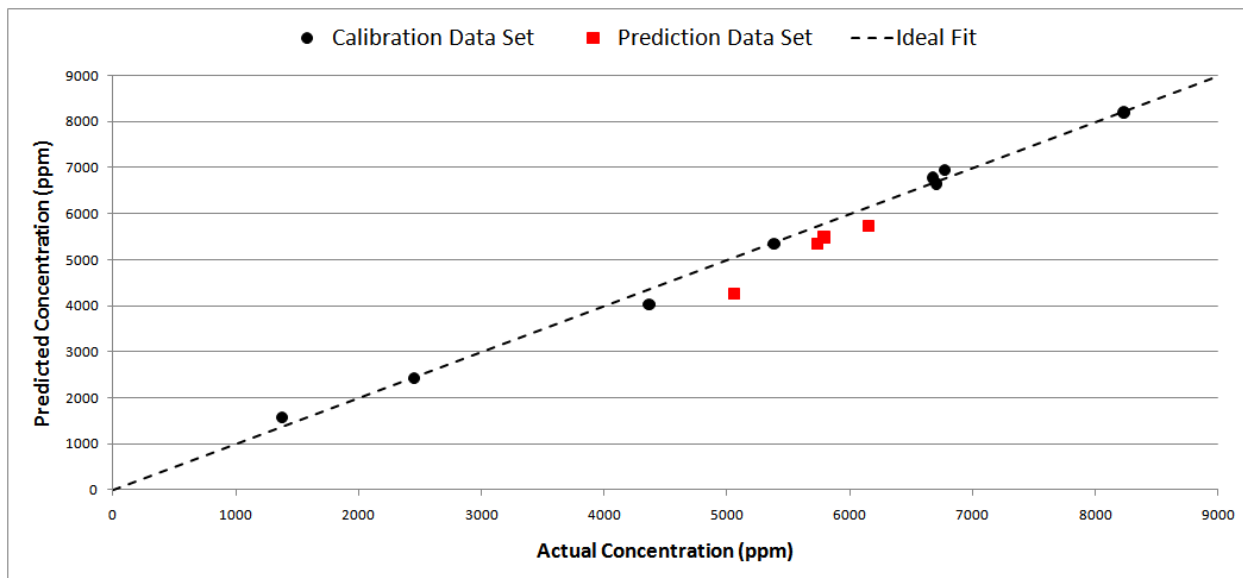


Figure 6-52: Principal Component Regression (PCR) – H₂O Concentrations in CO/CO₂/H₂O Gas Mixtures; RMSE Calibration = 157 ppm, RMSE Prediction = 519 ppm

As shown in Figures 6-53 including more than three principal components does not improve the RMSE for calibration and prediction significantly. In the case of RMSE for prediction, the inclusion of more than three principal components actually makes the results worse, most likely due to the inclusion of repetitive data that causes numerical instability in the regression model. This indicates that even though five principal components are necessary to represent at least 95% of the original calibration data set variance, that no beneficial and quantitative analysis is provided by including more than the three known components within the CO/CO₂/H₂O mixed gas system. In an unknown system, the determined number of components necessary for the analysis would be five and thus the actual system would not perform as well in terms of quantifying the CO, CO₂, and H₂O concentrations.

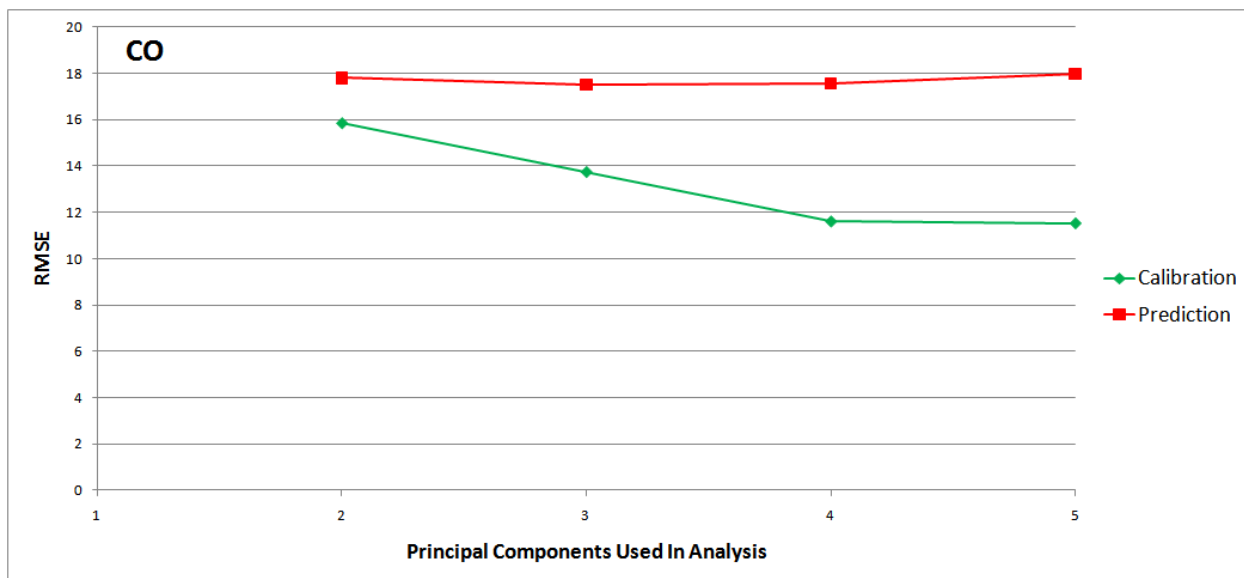


Figure 6-53: CO RMSE Calibration and RMSE Prediction as a Function of the Number of Principal Components Used to Represent the Original Data Set in CO/CO₂/H₂O Gas Mixtures

PCA on the entire set of collected experimental data set spectra was not capable of isolating the peaks associated with each particular component. This failure was due to the strong absorbance of the H₂O in comparison to the CO and CO₂ gas species. The strong contrast between the components is highlighted in the dynamic range values found in the previous PCR regression coefficient analysis, where CO had values of -4.6 to 6.4 (dynamic range of 11.0), CO₂ had values of -1.9 to 1.4 (dynamic range of 3.4), and H₂O had values of -10.4 to 204.7 (dynamic range of 215.1). With this large contrast between the three species, PCA can only identify that there are three main components in the mixed gas environment but cannot quantify the amounts of each. This data set highlights a shortcoming of the PCC analysis technique in dealing with multi-component analysis that contains both significant IR absorbing species and relatively low IR absorbing species. The PCR technique though was able quantitatively determine with

relatively low RMSE values the amounts of CO, CO₂, and H₂O in the three component gas mixture as shown in Table 6-15.

Table 6-15: Errors Associated with the Principal Component Regression (PCR) Analysis Technique; CO/CO₂/H₂O Experimental Gas Mixtures

Analysis Technique	RMSE Calibration (ppm)	RMSE Prediction (ppm)
CO: PCR	14	18
CO ₂ : PCR	2	4
H ₂ O: PCR	157	519

In an attempt to overcome the shortcoming of the PCC analysis technique in dealing with multi-component analysis that contains both significant IR absorbing species and relatively low IR absorbing species, PCA was conducted on the experimentally collected IR spectra from 2500 to 2075 cm⁻¹, which removes the IR spectra of the H₂O component. This truncated the analysis to a set of 8 samples ($n = 8$) at 501 different wavenumbers ($p = 425$) with concentrations given in Table 6-13 for the calibration data set $[X_C]_{(8 \times 425)}$. A set of 4 samples with their concentrations given in Table 6-14 were used for the prediction data set $[X_P]_{(4 \times 425)}$ for PCR. In the case of the PCC, a single data set $[X]_{(12 \times 425)}$ consisting of the calibration and prediction data sets were combined and PCA was simultaneously performed on all 12 samples.

From the plot of eigenvalues as a function of principal components, the number of significant eigenvalues obtained from PCA is determined to be four (Figure 6-54). The first principal component explains 73.4% of the total calibration data set variance, the second principal component explains 12.6%, the third explains 6.2% of the total calibration data set variance, and the fourth explains 4.5%. With these four principal components, 96.7% of the original data variance can be explained.

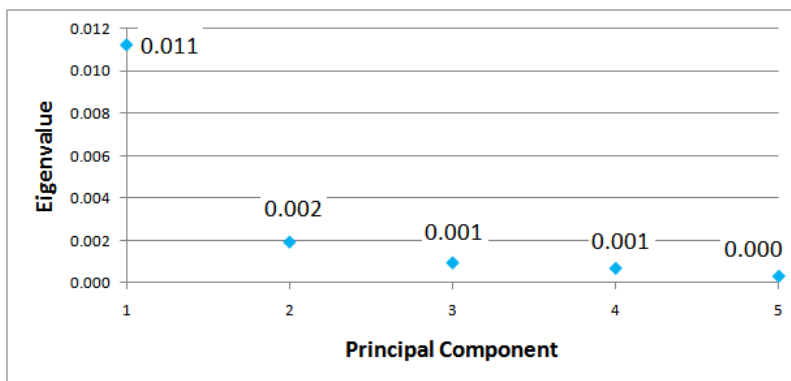


Figure 6-54: SCREE Plot Indicating Eigenvalues for Each Principal Component for CO/CO₂/H₂O Gas Mixtures Data Set; Principal Components 1, 2, and 3 Explain 73.4%, 12.6%, and 6.2% of the Total Calibration Data Set Variance

As shown in Figures 6-55 including more than two principal components does not improve the RMSE for calibration and prediction significantly. In the case of RMSE for prediction, the inclusion of more than the expected number of principal components actually makes the results worse as more principal components are used in the PCR analysis. This indicates that even though five principal components are necessary to represent at least 95% of the original calibration data set variance, that no beneficial and quantitative analysis is provided by including more than the two known components within the CO/CO₂ mixed gas system. In addition, the RMSE for calibration and prediction are significantly worse when PCA is performed on the three-component system where the spectral data is truncated to include only contributions from two components.

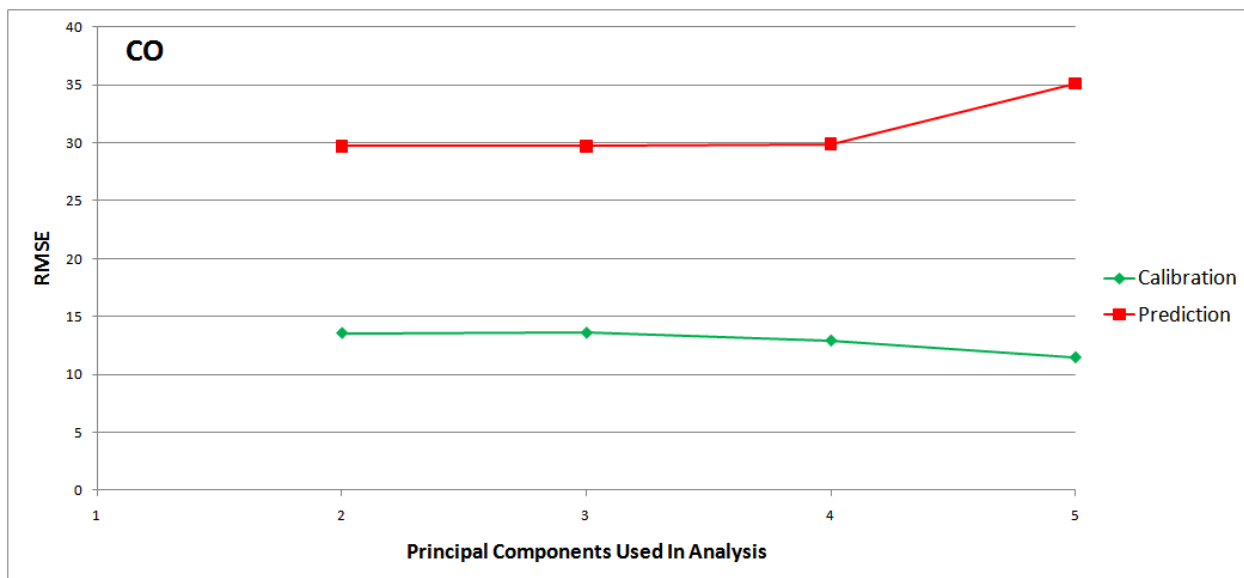


Figure 6-55: CO RMSE Calibration and RMSE Prediction as a Function of the Number of Principal Components Used to Represent the Original Data Set in CO/CO₂/H₂O Gas Mixtures with the IR Spectral Data Truncated to Include Only Contributions from CO and CO₂ (2500 cm⁻¹ to 2075 cm⁻¹) for PCA

PCA on the entire set of collected experimental truncated data set spectra was not capable of isolating the peaks associated with each particular component, as was the case with PCA on the entire set of collected experimental data. This failure indicates that due to the strong absorbance of the H₂O in comparison to the CO and CO₂ gas species within each FTIR scan, the data analysis for wavenumbers containing information for only CO and CO₂ is still adversely affected. Because of this, to determine quantitatively the amount of CO and CO₂ gas species within the system, FTIR scans avoiding the H₂O absorbance range are necessary as was shown in the CO/CO₂ Experimental Data Set where the wavenumber range was 2500-2000 cm⁻¹. This method of analysis covering the wavenumber range of 2500-1500 cm⁻¹ though is still beneficial as it identifies that only three components are present at significant amounts within the spectra as was discussed in relation to Figure 6-62.

Chapter 7: PCA Application to FTIR Spectroscopy Data of Potential Bleed Air Contaminants within the Aircraft Cabin

Discussion of Analyzed Data Sets

Data set 1 was comprised of experimentally obtained IR spectra from four different engine oils used in commercial aircraft engines. The oils analyzed were BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560. These engine oils samples were heated in a thermogravimetric analysis (TGA) chamber (Thermo-Microbalance TG 209 *FI Iris*®) with a programmed heat treatment of 20° C per minute ramp rate until evaporation of the oil was completed. Coupled with the TGA were a FTIR (Bruker FTIR Tensor® 27) and a mass spectrometry (MS) analyzer (QMS 403 C *Aëolos*®) via heated transfer lines. The FTIR scans were recorded at the temperature of greatest mass loss, which for the BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 306 °C (583 °F), 301 °C (574 °F), 306 °C, and 326 °C (619 °F), respectively. The IR spectra for data set 1 was obtained using the FTIR spectrometer over a wavenumber scan range from 4000 cm⁻¹ to 600 cm⁻¹ with a spectral resolution of 2 cm⁻¹. PCA was performed on data set 1 without PCR or PCC to determine the number of components most likely present in the evolved gas species from the heated engine oil samples. This data was then compared to the MS data for verification and the MS data was used to determine the identity of the components.

Data set 2 was comprised of a simulated calibration data set with 10 spectra containing various concentrations of methanol (CH₄O), CH₂O, and CO₂. These gas components were

identified in the MS analysis of data set 1 in addition to PCA identifying three significant components present in data set 1. The prediction data set for data set 2 consisted of the experimental engine oil samples from data set 1. The purpose of this was to determine not only the component present but to quantify the concentrations of each of the identified gases through PCR. With the concentrations of the individual gases determined from PCR for the engine oil samples, predicted spectra were calculated and compared to the original FTIR scan data over the wavenumber range of 4000 cm^{-1} to 600 cm^{-1} at a resolution of 2 cm^{-1} . The RMSE between the predicted spectra and actual spectra of the engine oils after completion of PCA with PCR was computed based on absorbance values at each wavenumber. The experimental data used in data sets 1 and 2 were collected by Netzsch Instruments, Inc. (Burlington, MA, USA).

Data set 3 consisted of time-evolved room temperature ($23\text{-}24\text{ }^{\circ}\text{C}$ / $74\text{-}75\text{ }^{\circ}\text{F}$) gas analysis of heated BP Turbo Oil 2380. The 1mL sample of engine oil was heated in a cylindrical furnace with a programmed heat treatment of 10° C per minute ramp rate until the hold temperature was reached, which in this case was an oil temperature of approximately $275\text{ }^{\circ}\text{C}$ ($527\text{ }^{\circ}\text{F}$). The engine oil heating system was operated by Amanda Neer (M.S. Graduate Student, Materials Engineering, Auburn University). The specific details of the system can be found in her Master Thesis. The gas was monitored with a temperature sensor and allowed to cool to room temperature as it was transferred to the Spectrum GX FTIR with variable pathlength long path gas cell discussed in Chapter 2 with the use of a vacuum pump. The FTIR scans were recorded at time intervals every 5 minutes while the engine oil was held at $275\text{ }^{\circ}\text{C}$ for 1 hour before the heating furnace was shut off and the sample allowed to return to room temperature. The IR spectra for data set 3 were obtained over a wavenumber scan range from 4000 cm^{-1} to 600 cm^{-1} with a spectral resolution of 2 cm^{-1} . The pathlength of the gas cell was 2.24 m.

PCA Results and Discussions

Data Set 1 – Engine Oil Samples at Temperatures of Greatest Mass Loss

From the experimental FTIR spectral analysis of the engine oils at temperature of greatest mass loss, a set of 4 samples ($n = 4$) at 1763 different wavenumbers ($p = 1763$) were collected as shown in Figure 7-1 for PCA. As shown in Figure 7-2, the number of significant eigenvalues obtained from PCA is three. The first principal component explains 76.1% of the total calibration data set variance, the second principal component explains of 21.6% of the total calibration data set variance, and the third principal component explains of 2.3% of the total calibration data set variance. With these three principal components, 100.0% of the original data variance can be explained.

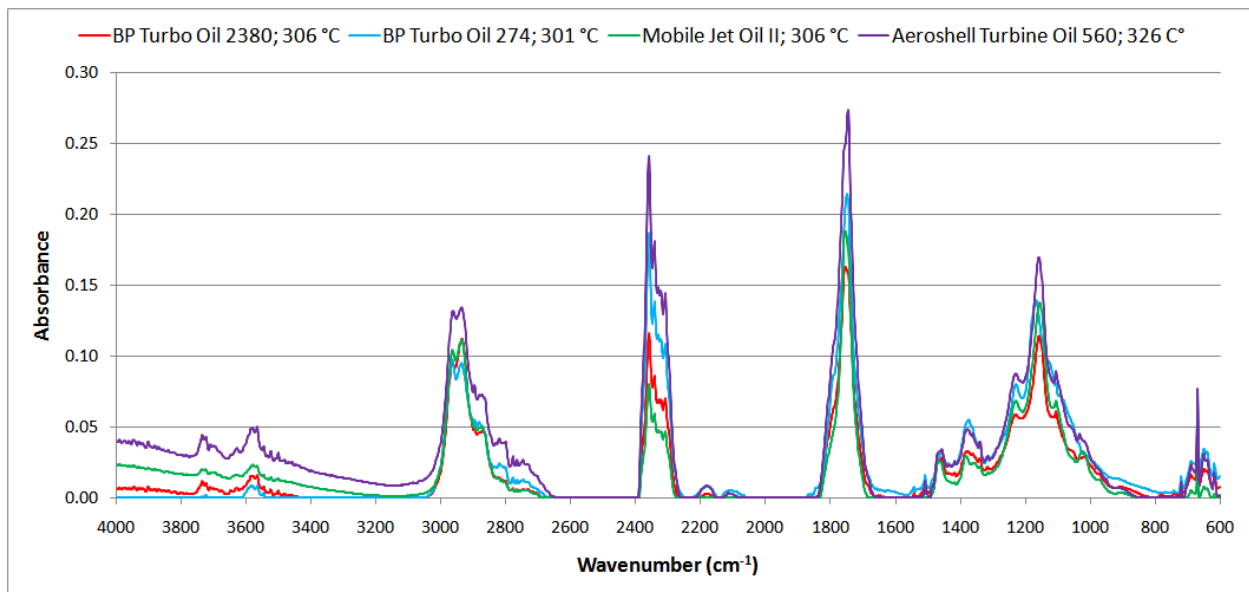


Figure 7-1: FTIR Scans of Engine Oil Samples at Temperatures of Greatest Mass Loss used for PCA; BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 at 306 °C (583 °F), 301 °C (574 °F), 306 °C, and 326 °C (619 °F), respectively

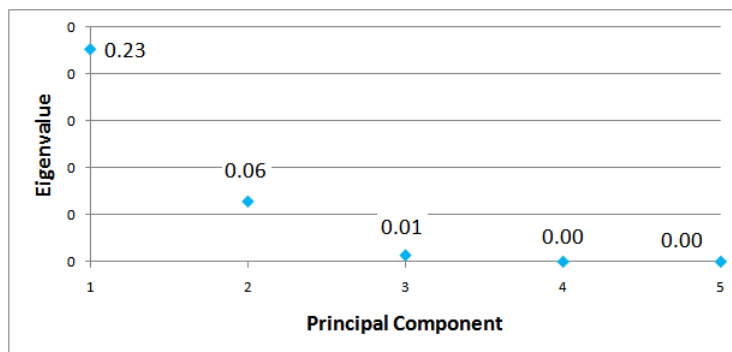


Figure 7-2: SCREE Plot Indicating Eigenvalues for Each Principal Component for Engine Oil Data Set 1; Principal Components 1, 2, and 3 Explain 76.1%, 21.6%, and 2.3% of the Total Engine Oil Data Set 1 Variance

Figure 7-3 highlights the MS data for the BP Turbo Oil 274 at 301 °C (574 °F), and Mobile Jet Oil II at 306 °C (583 °F). This MS data corresponds to characteristic MS data for CH₄O, CH₂O, and CO₂, whose database spectra are shown in Figure 7-4. As evident based on relative intensities of the two oils analyzed in Figure 7-3, the amount of CH₂O is significantly higher than that of CO₂. In addition, in regards to Figure 7-3, the peak that corresponds to CH₄O at atomic mass unit (amu) 32 is not included because it is such a high intensity that the relative intensities for the other components would become indistinguishable. This indicates that a majority of the evolved gas at the temperature of highest mass loss for the engine oils is due to CH₄O. The analysis could have been possible with the MS data, but a comparison of the FTIR plot would have to have been made to numerous alcohol and aldehyde based molecules to determine the exact gas species present.

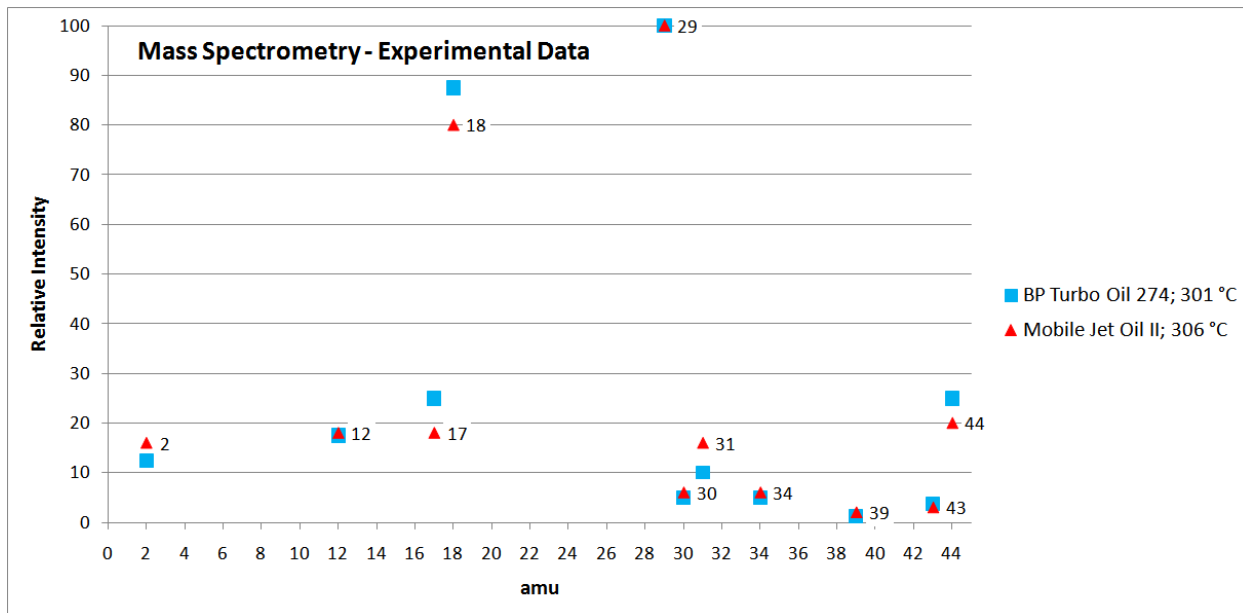


Figure 7-3: Mass Spectrometry (MS) Data of Engine Oil Samples at Temperatures of Greatest Mass Loss; BP Turbo Oil 274, and Mobile Jet Oil II at 301 °C (574 °F), and 306 °C (578 °F), respectively

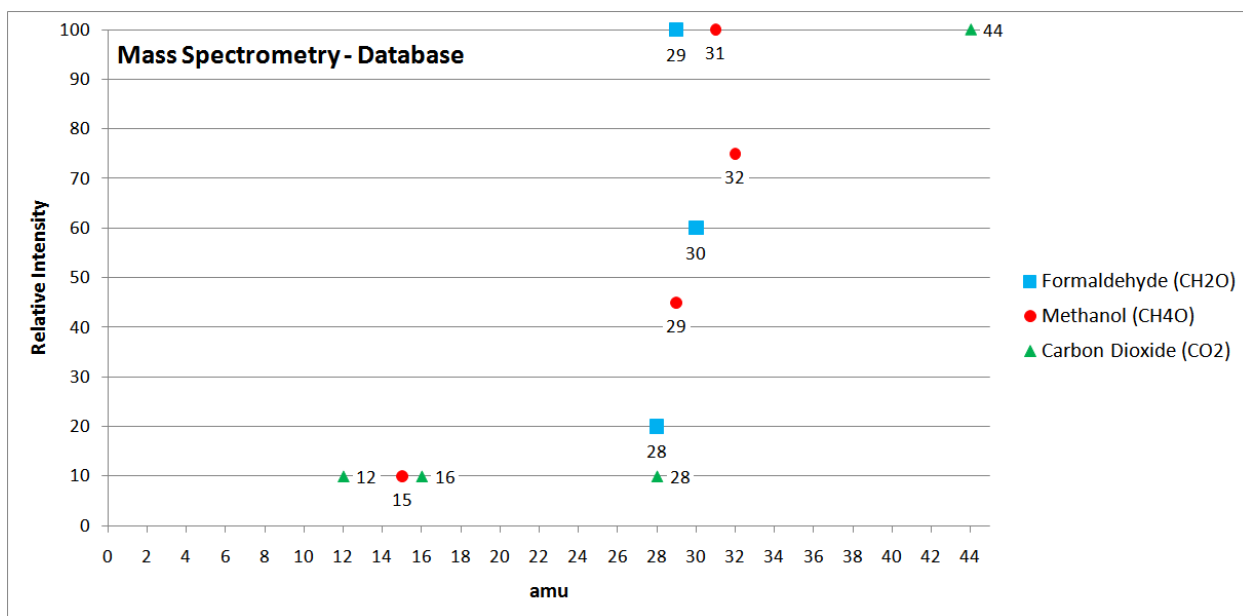


Figure 7-4: Mass Spectrometry (MS) Data Database Files for Formaldehyde (CH₂O), Methanol (CH₄O), and Carbon Dioxide (CO₂)

Data Set 2 – Simulated CH₄O/CH₂O/CO₂ Gas Mixtures & Engine Oil Samples at Temperatures of Greatest Mass Loss

From the simulated FTIR spectral analysis of the pure CH₄O/CH₂O/CO₂ gas mixtures shown in Figure 7-5, a set of 10 samples ($n = 10$) at 1,763 different wavenumbers ($p = 1,763$) with concentrations given in Table 7-1 were calculated as shown in Figure 7-6 to produce a calibration data set $[X_C]_{(10 \times 1763)}$. The experimental data used in data set 1 (Figure 7-1) was used as the prediction data set $[X_P]_{(4 \times 1763)}$ for PCR.

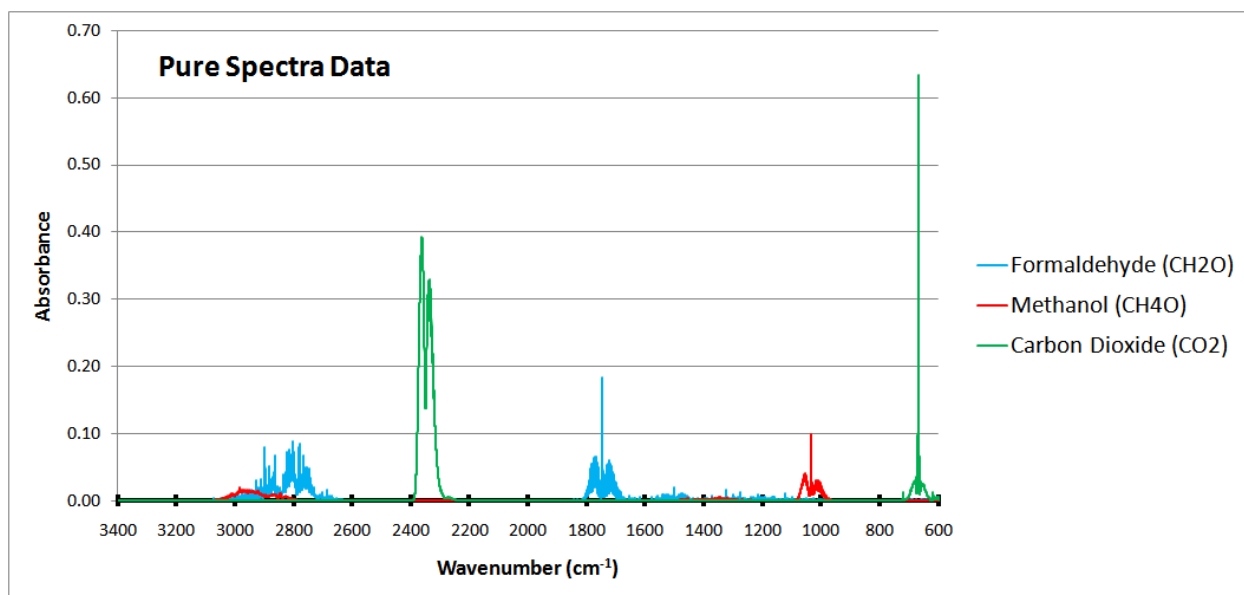


Figure 7-5: Formaldehyde (CH₂O), Methanol (CH₄O), and Carbon Dioxide (CO₂) Pure Spectra for Simulated Data Set Illustrating Spectral Overlap Between CH₂O and CH₄O

Table 7-1: Calibration Data Set 2 Composed of Simulated Spectra of Various Concentrations of Methanol (CH₄O), Formaldehyde (CH₂O), and Carbon Dioxide (CO₂)

Sample	CH ₄ O Concentration (x100 ppm)	CH ₂ O Concentration (x100 ppm)	CO ₂ Concentration (x100 ppm)
1	0.0	3.0	2.0
2	0.5	2.0	4.0
3	1.0	0.0	3.5
4	2.0	6.0	5.5
5	3.5	7.0	0.0
6	5.0	1.5	0.5
7	6.5	3.5	4.5
8	8.0	4.5	2.5
9	9.0	0.5	3.0
10	10.0	5.0	1.0

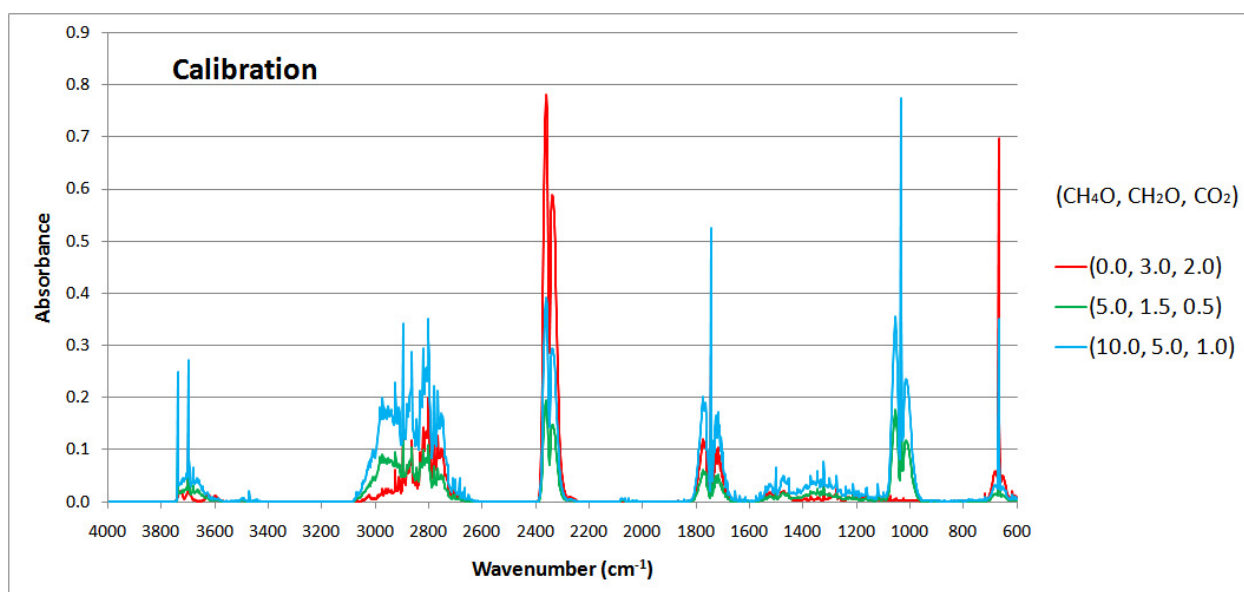


Figure 7-6: CH₄O/CH₂O/CO₂ Gas Mixtures Calibration Data Set, [X_C]_(10 x 1763); Note – Only 3 of 10 Calibration Spectra Shown

As shown in Figure 7-7, the number of significant eigenvalues obtained from PCA is three. The first principal component explains 86.8% of the total calibration data set variance, the second principal component explains of 8.6%, and the third principal component explains 4.6% of the total calibration data set variance. With these three principal components, 100.0% of the original data variance can be explained.

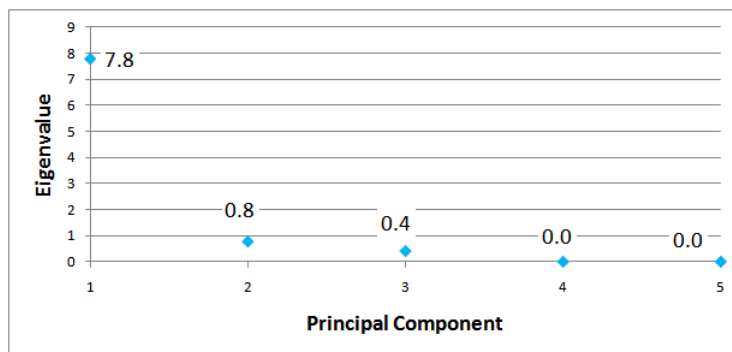


Figure 7-7: SCREE Plot Indicating Eigenvalues for Each Principal Component for CH₄O/CH₂O/CO₂ Gas Mixtures Data Set; Principal Components 1, 2, and 3 Explain 86.8%, 95.4%, and 4.6%, respectively, of the Total Calibration Data Set Variance

Applying PCR, a plot can be created that shows the calibration data set used to determine the concentrations in the prediction data sets as shown in Figures 7-8, 7-9, and 7-10 for CH₄O, CH₂O, and CO₂, respectively. The RMSE for the calibration data set in regards to concentrations of CH₄O, CH₂O, and CO₂ are found to be 0 ppm (simulated data sets). The predicted CH₄O concentrations for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 206, 221, 214, and 292 ppm, respectively. The predicted CH₂O concentrations for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 145, 190, 142, and 263 ppm, respectively. The predicted CO₂ concentrations for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 30, 49, 20, and 63 ppm, respectively.

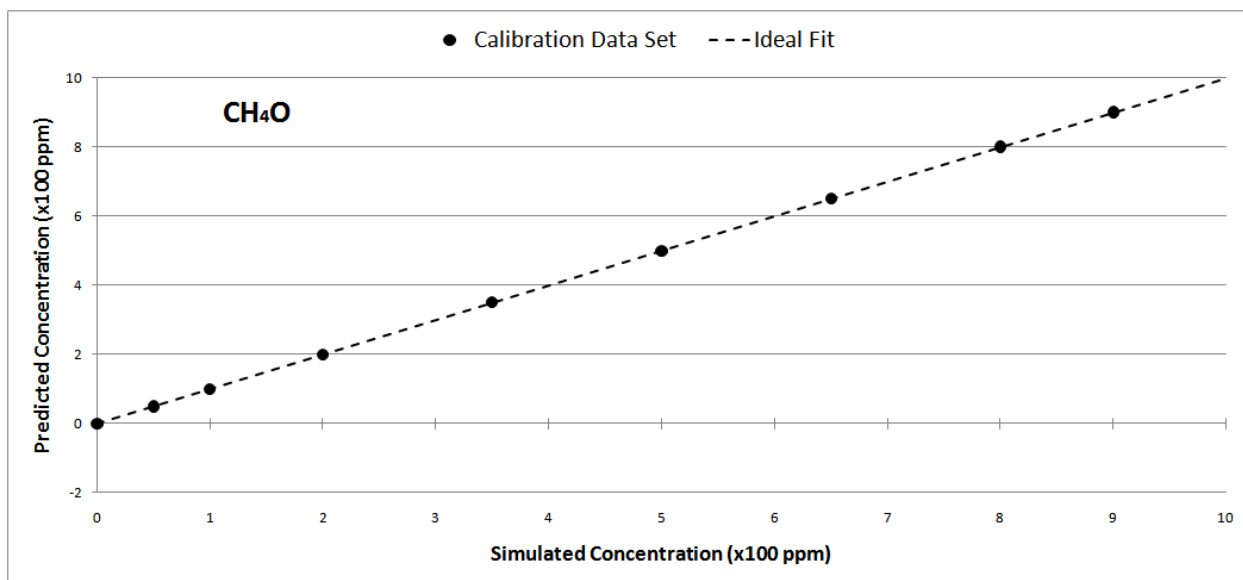


Figure 7-8: Principal Component Regression (PCR) – CH₄O Concentrations in CH₄O/CH₂O/CO₂ Gas Mixtures; RMSE Calibration = 0 ppm; Predicted concentrations for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 206, 221, 214, and 292 ppm, respectively

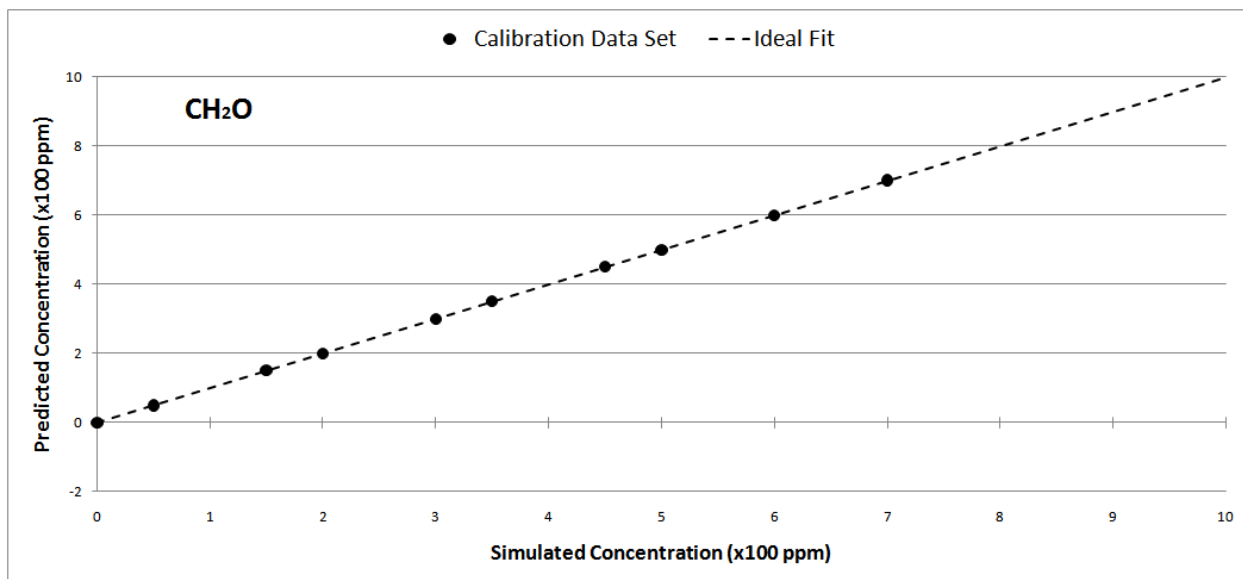


Figure 7-9: Principal Component Regression (PCR) – CH₂O Concentrations in CH₄O/CH₂O/CO₂ Gas Mixtures; RMSE Calibration = 0 ppm; Predicted concentrations for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 145, 190, 142, and 263 ppm, respectively

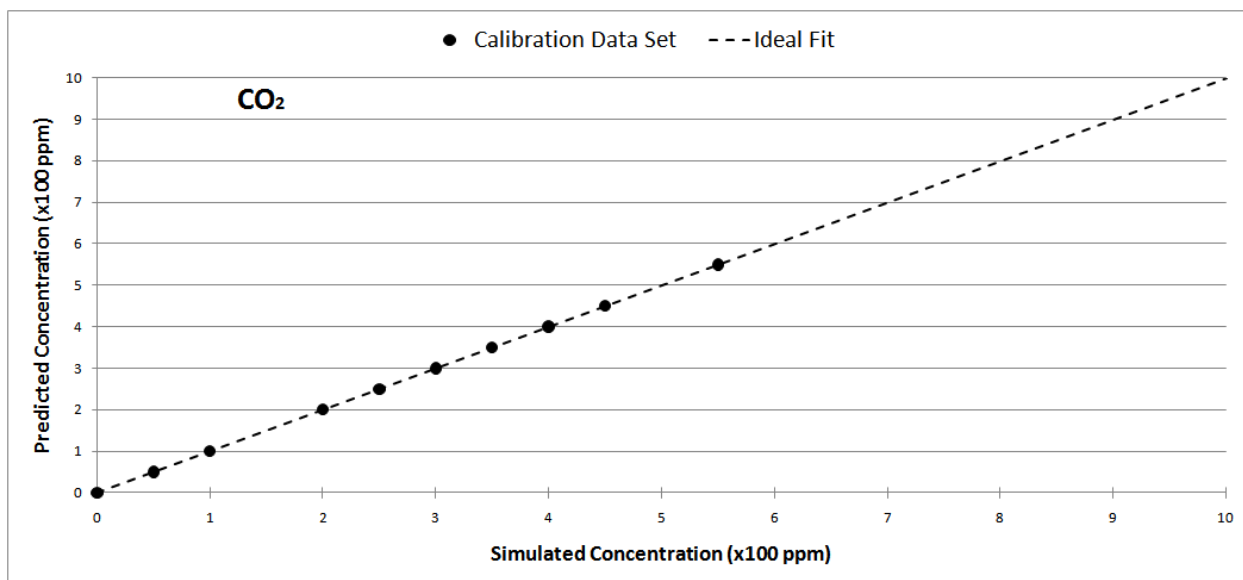


Figure 7-10: Principal Component Regression (PCR) – CO₂ Concentrations in CH₄O/CH₂O/CO₂ Gas Mixtures; RMSE Calibration = 0 ppm; Predicted concentrations for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560 were 30, 49, 20, and 63 ppm, respectively

After using PCR to determine concentrations of CH₄O, CH₂O, and CO₂ for the four engine oil samples (Table 7-2), a reconstructed prediction spectrum can be calculated and compared to the experimentally obtained FTIR data as shown in Figures 7-11 thru 7-14 for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560, respectively. The RMSE between the predicted and actual spectra for each of the engine oils are shown in Table 7-3. The calculation of RMSE with this method is necessary because the actual amounts of the gas components are unknown.

Table 7-2: PCR Calculated Concentrations of CH₄O, CH₂O, and CO₂ for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560

Sample	CH ₄ O Concentration (ppm)	CH ₂ O Concentration (ppm)	CO ₂ Concentration (ppm)
BP Turbo Oil 2380	206	145	30
BP Turbo Oil 274	221	190	49
Mobile Jet Oil II	214	142	20
Aeroshell Turbine Oil 560	292	263	63

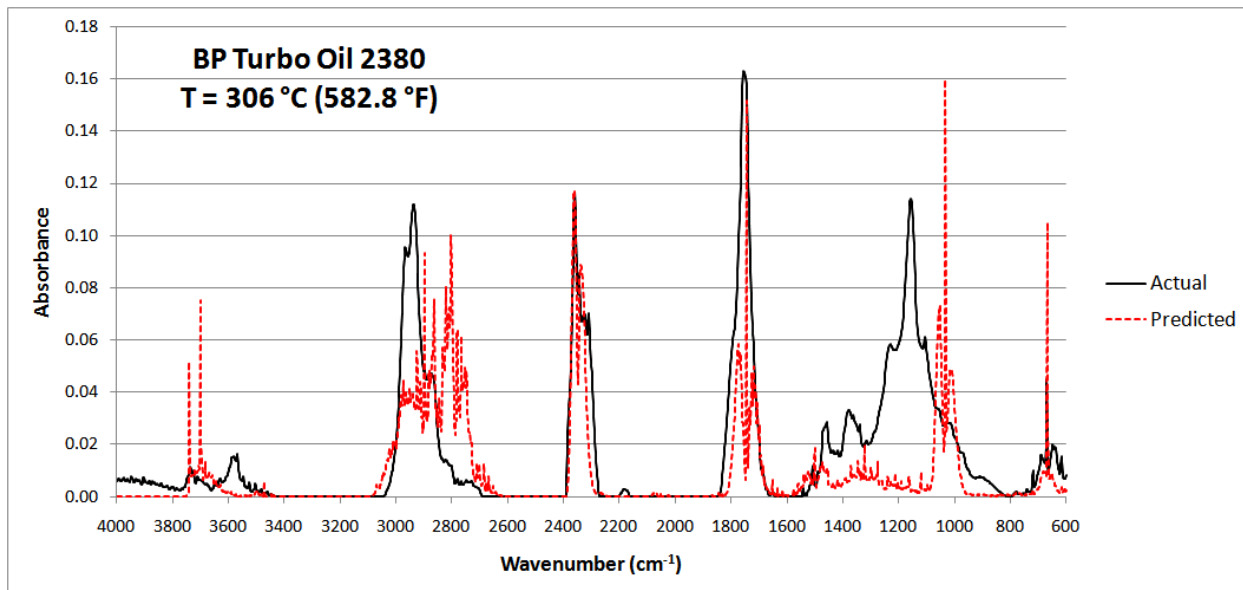


Figure 7-11: Predicted Spectra for BP Turbo Oil 2380 based on PCR Calculated Concentrations of CH₄O, CH₂O, and CO₂ Gas Mixtures; RMSE = 2.4%

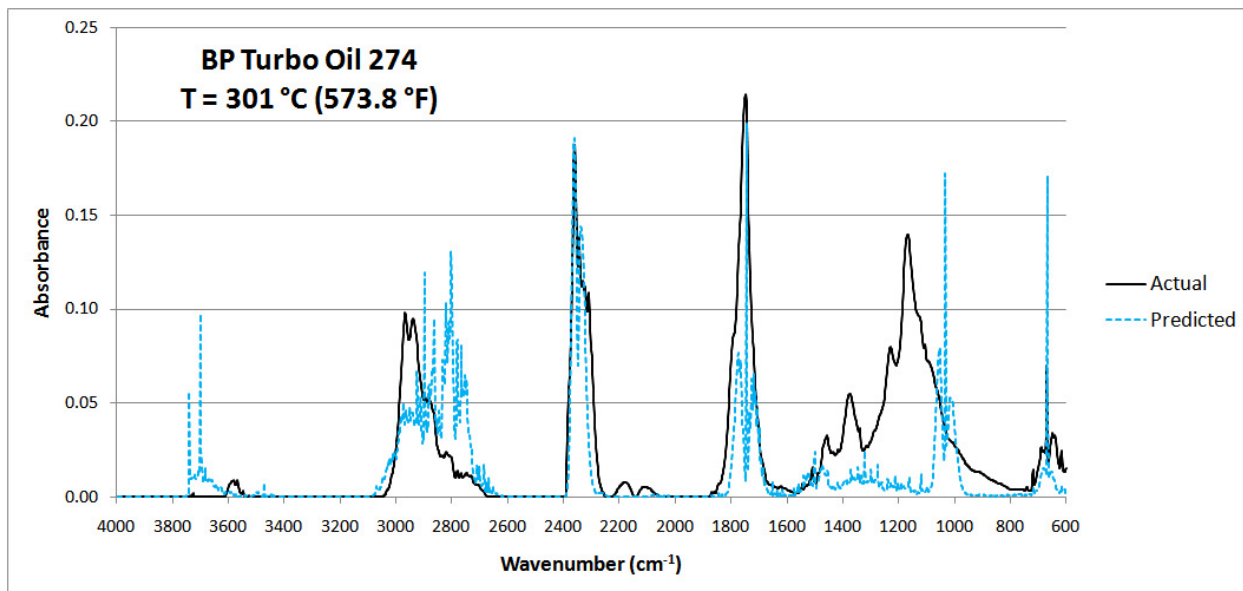


Figure 7-12: Predicted Spectra for BP Turbo Oil 274 based on PCR Calculated Concentrations of CH₄O, CH₂O, and CO₂ Gas Mixtures; RMSE = 3.1%

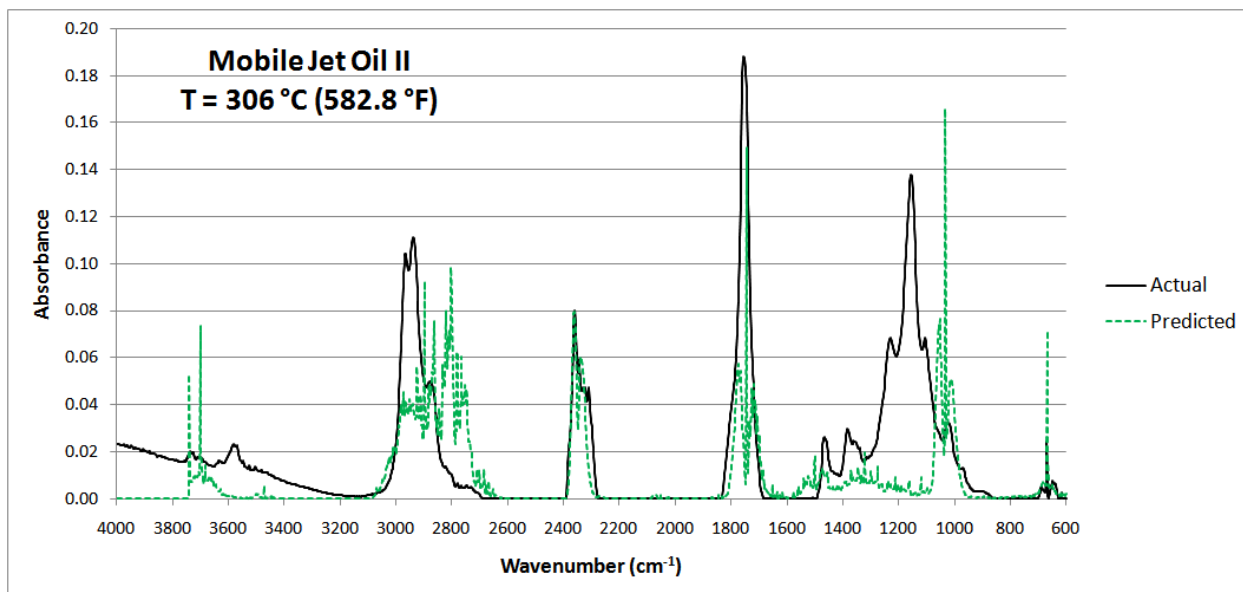


Figure 7-13: Predicted Spectra for Mobile Jet Oil II based on PCR Calculated Concentrations of CH_4O , CH_2O , and CO_2 Gas Mixtures; RMSE = 2.7%

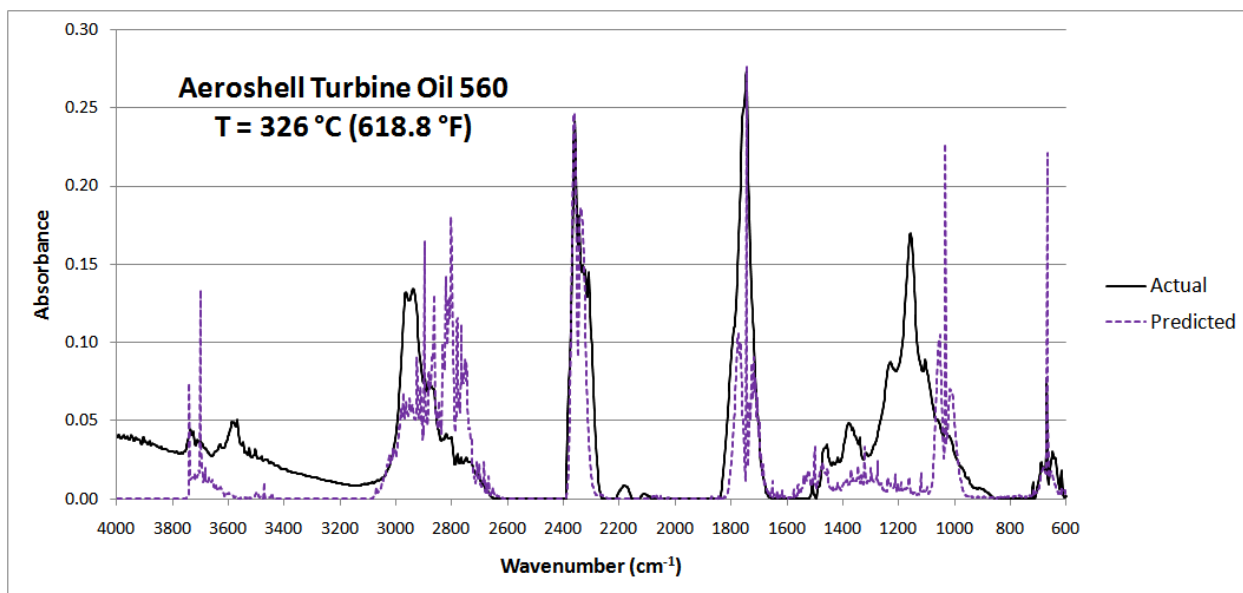


Figure 7-14: Predicted Spectra for Aeroshell Turbine Oil 560 based on PCR Calculated Concentrations of CH_4O , CH_2O , and CO_2 Gas Mixtures; RMSE = 3.9%

Table 7-3: RMSE for Predicted Spectra based on PCR Calculated Concentrations of CH₄O, CH₂O, and CO₂ for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560

Sample	RMSE (%)
BP Turbo Oil 2380	2.4
BP Turbo Oil 274	3.1
Mobile Jet Oil II	2.7
Aeroshell Turbine Oil 560	3.9

From the predicted spectra, it was noted that the peaks associated with methanol did not line up properly at the expected wavenumbers, particularly the wavenumbers associated with the C-O stretching ($1150\text{ cm}^{-1} - 1050\text{ cm}^{-1}$) and the O-H stretching ($3700\text{ cm}^{-1} - 3500\text{ cm}^{-1}$). This indicates that there are disturbances in the hydrogen bonding present within CH₄O due to the clustering of CH₄O molecules. This is similar to observed and theoretical predicted shifts reported in literature, where the peaks associated with the C-O stretching shift towards higher wavenumbers (blue shifts, i.e. shift to lower frequency) and the peaks associated with the O-H stretching shift towards lower wavenumbers (red shifts, i.e. shift to higher frequency) [51-54]. The blue shift in the C-O stretching peak is due to a shortening of the bond, while the red shift in the O-H stretching is due to the elongation of the bond, which occur due to the mutual interactions within the hydroxyl group via the hydrogen bonded CH₄O clusters [55]. This mutual interaction within the hydroxyl group is illustrated in Figure 7-15. Wavenumber shifts are not observed in the CH₂O or CO₂ spectrum components and are not to be expected since these molecules do not have hydrogen bonding present.

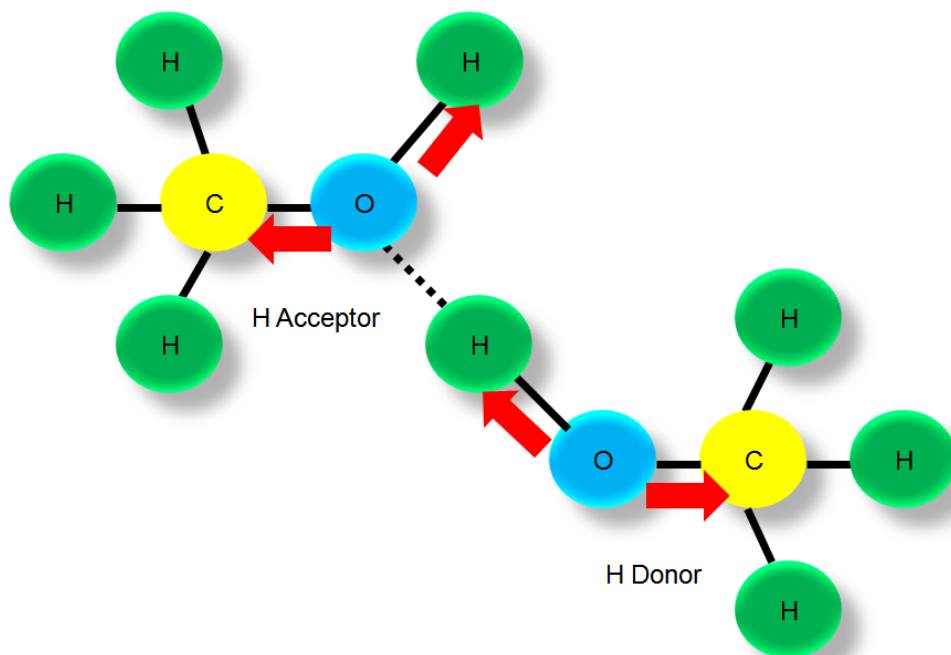


Figure 7-15: Illustration of Mutual Interactions within the Hydroxyl Group via Hydrogen Bonded Methanol (CH_4O) Clusters that Leads to O-H Bond Lengthening and C-O Bond Shortening Explaining the Observed Red and Blue Wavenumber Shifts, Respectively, within the CH_4O IR Spectra Component of the Engine Oil at High Temperature

An additional method to describe the observed shifts in characteristic wavenumber for the C-O and O-H stretching modes is to recognize that molecular vibrations can be treated utilizing Newtonian mechanics. In this case, each vibration or stretching mode corresponds to a spring with a spring or force constant, k , as defined in Equation 7.1 (Hooke's Law), which can then be related to effective mass of the molecule, μ (defined in Equation 7.2), and the acceleration using Newton's 2nd Law of Motion (Equation 7.3). In Equation 7.2, m_1 and m_2 correspond to the mass of the bonded atoms.

$$F = -kx \qquad \text{(Hooke's Law)} \qquad (7.1)$$

$$\mu = \frac{m_1 + m_2}{m_1 m_2} \quad (7.2)$$

$$\mu \frac{d^2 x}{dt^2} = -kx \quad (7.3)$$

Rearranging Equation 7.3 yields the homogenous linear differential equation of 2nd order shown in Equation 7.4, which has a standard general solution in which the force constant can be related to the vibrational mode characteristic wavenumber, ν (cm⁻¹), through Equation 7.5, where c is the speed of light in cm/s².

$$\mu \frac{d^2 x}{dt^2} + kx = 0 \quad (7.4)$$

$$\nu = \frac{1}{2\pi c} \sqrt{k\mu} \quad (7.5)$$

With the observed shifts, a new calibration data set for the PCA can be created that has the CH₄O modified as shown in Figure 7-16, with a C-O stretch blue shifted by approximately 121 cm⁻¹ and the O-H stretch red shifted by approximately 121 cm⁻¹. Using Equation 7.5, the room temperature C-O stretch with a peak located at ~1033 cm⁻¹ corresponds to $k_{C-O} = 4.3$ N/cm, while the shifted peak located at ~1154 cm⁻¹ corresponds to $k_{C-O} = 5.4$ N/cm. Again using Equation 7.5, the room temperature O-H stretch with a peak located at ~3682 cm⁻¹ corresponds to $k_{O-H} = 7.6$ N/cm, while the shifted peak located at ~3561 cm⁻¹ corresponds to $k_{O-H} = 7.1$ N/cm. These force constant values and their respective changes are consistent with those reported in literature of hydrogen bonding associated with methanol [51-55]. The actual determination of wavenumber shifts as a function of temperature can only be estimated through inspection at this

time but as detailed in the future work section found in Chapter 9, more experiments could be undertaken to better quantify these shifts.

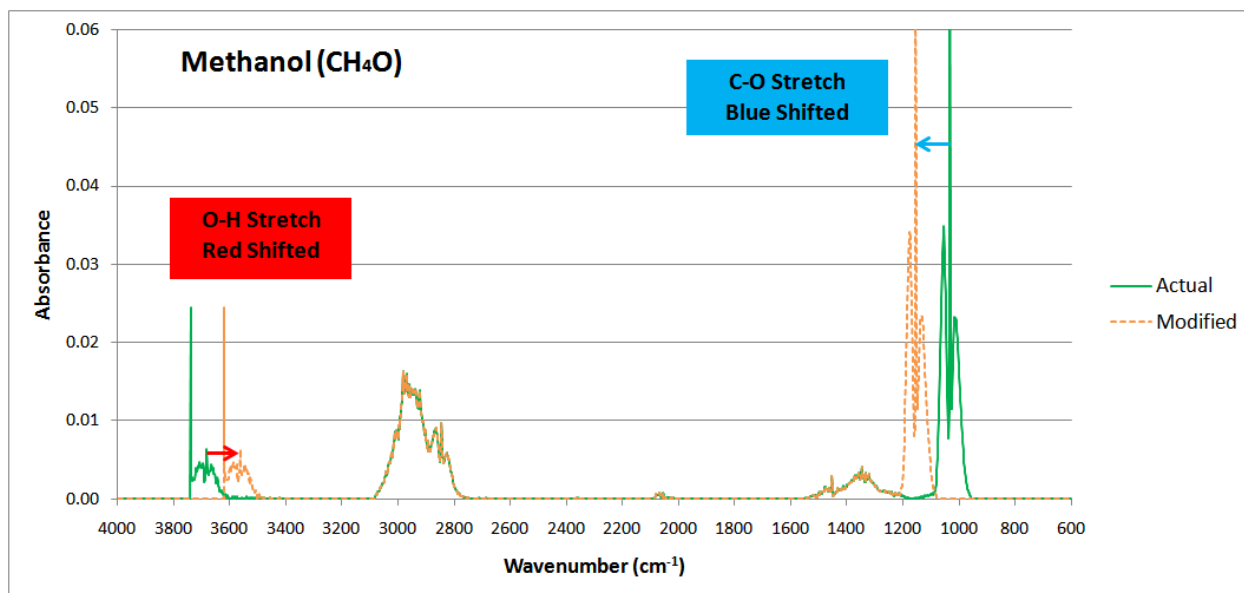


Figure 7-16: Comparison of Actual Methanol (CH_4O) Spectra to Modified CH_4O Spectra with the O-H Stretching Bands Red Shifted and the C-O Stretching Bands Blue Shifted Bands Due to High Temperature Disturbance of Hydrogen Bonds

From the simulated FTIR spectral analysis of the pure CH_4O (modified)/ $\text{CH}_2\text{O}/\text{CO}_2$ gas mixtures, a set of 10 samples ($n = 10$) at 1,763 different wavenumbers ($p = 1,763$) with the same concentrations given in Table 7-1 were calculated as shown in Figure 7-17 to produce a calibration data set $[\text{X}_\text{C}]_{(10 \times 1763)}$. The experimental data used in data set 1 (Figure 7-1) was used as the prediction data set $[\text{X}_\text{P}]_{(4 \times 1763)}$ for PCR.

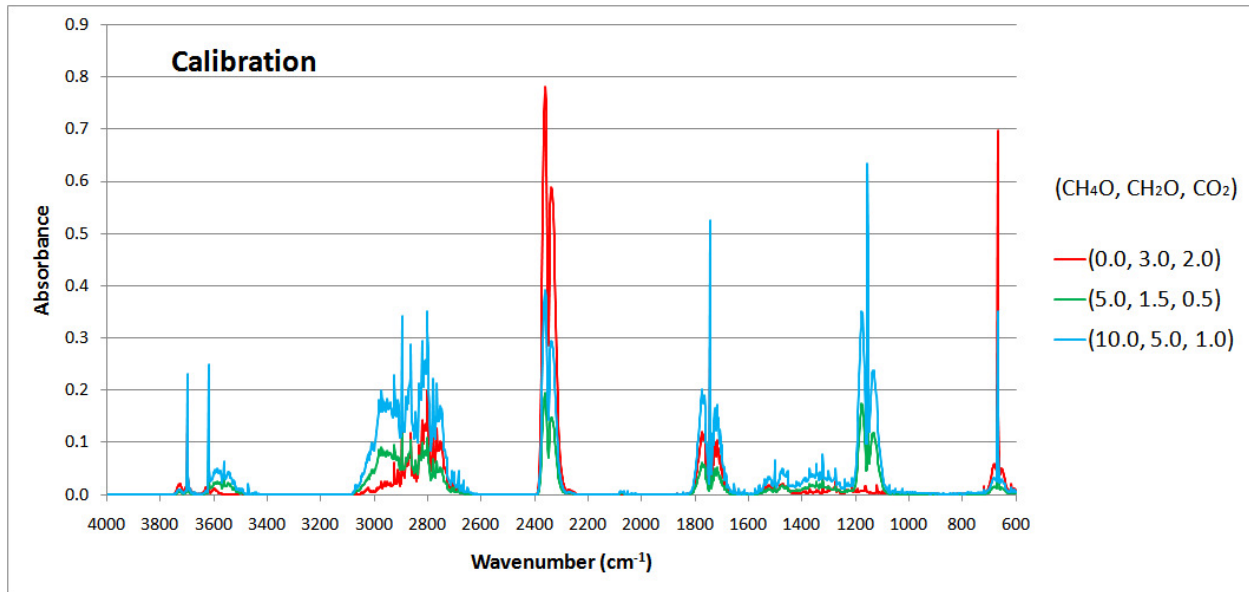


Figure 7-17: CH₄O (Modified)/CH₂O/CO₂ Gas Mixtures Calibration Data Set, [X_C]_(10 x 1763); Note – Only 3 of 10 Calibration Spectra Shown

After using PCA with PCR to determine concentrations of CH₄O, CH₂O, and CO₂ with the modified methanol calibration spectra for the four engine oil samples (Table 7-4), a reconstructed prediction spectrum can be calculated. This predicted spectrum is then compared to the experimentally obtained FTIR data as shown in Figures 7-18 thru 7-21 for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560, respectively. The RMSE between the predicted and actual spectra for the standard CH₄O and modified CH₄O spectra for each of the engine oils are shown in Table 7-5.

Table 7-4: PCR Calculated Concentrations of CH₄O (Modified), CH₂O, and CO₂ for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560

Sample	CH ₄ O Concentration (ppm)	CH ₂ O Concentration (ppm)	CO ₂ Concentration (ppm)
BP Turbo Oil 2380	350	127	30
BP Turbo Oil 274	417	166	49
Mobile Jet Oil II	407	118	20
Aeroshell Turbine Oil 560	511	235	63

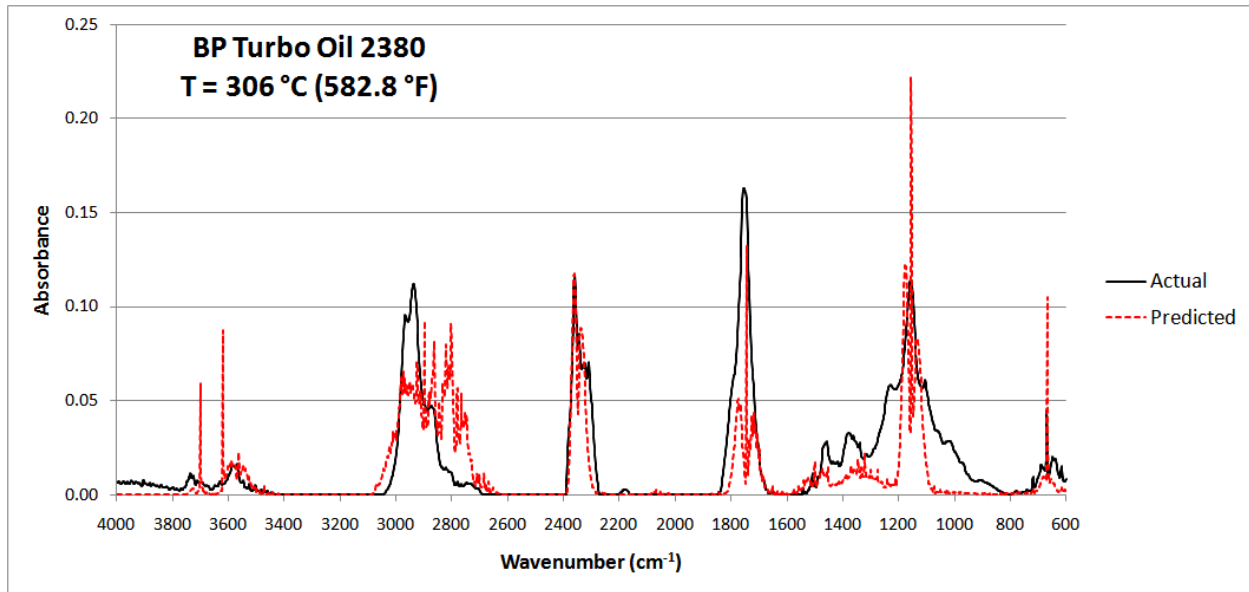


Figure 7-18: Modified Predicted Spectra for BP Turbo Oil 2380 based on PCR Calculated Concentrations of CH_4O , CH_2O , and CO_2 Gas Mixtures; RMSE = 2.1%

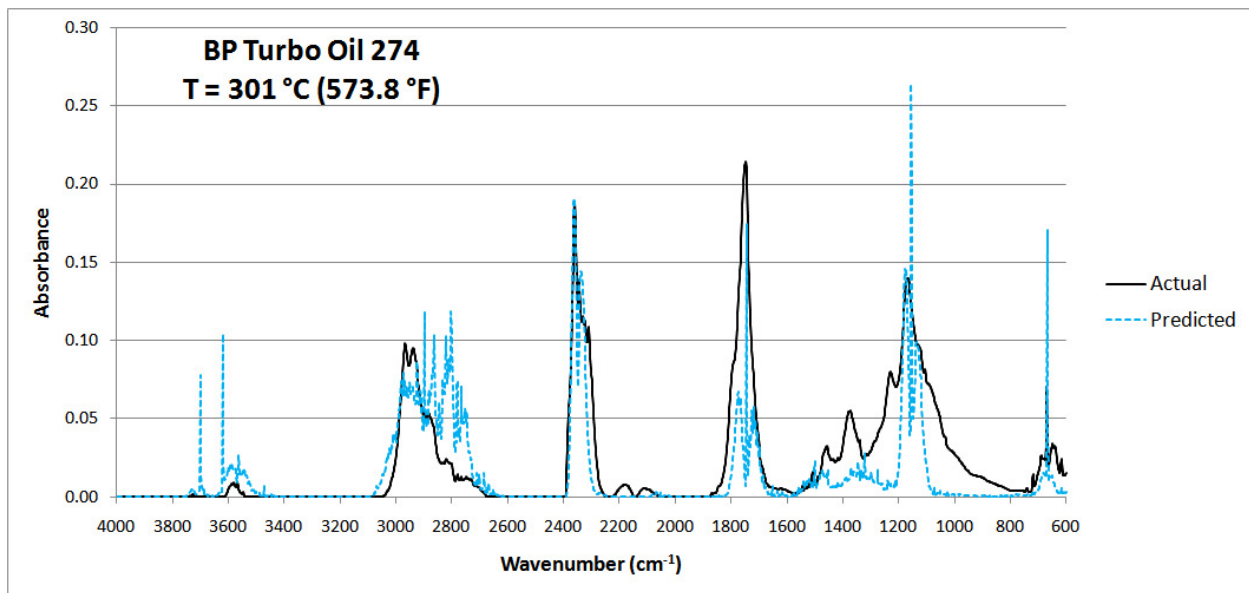


Figure 7-19: Modified Predicted Spectra for BP Turbo Oil 274 based on PCR Calculated Concentrations of CH_4O , CH_2O , and CO_2 Gas Mixtures; RMSE = 2.7%

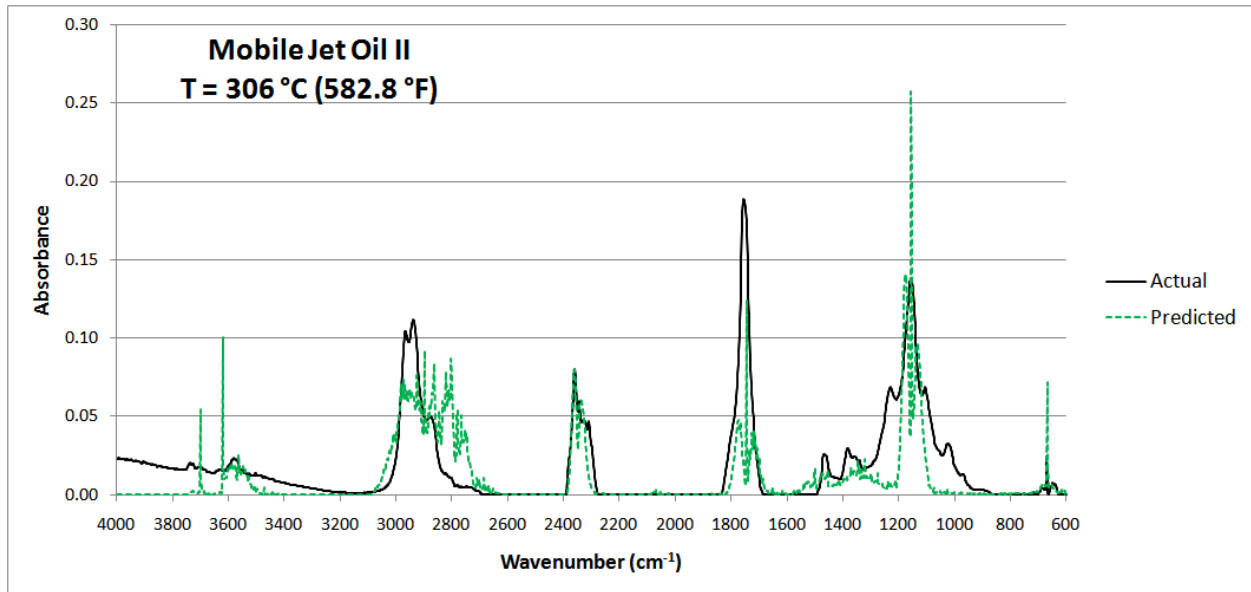


Figure 7-20: Modified Predicted Spectra for Mobile Jet Oil II based on PCR Calculated Concentrations of CH_4O , CH_2O , and CO_2 Gas Mixtures; RMSE = 2.2%

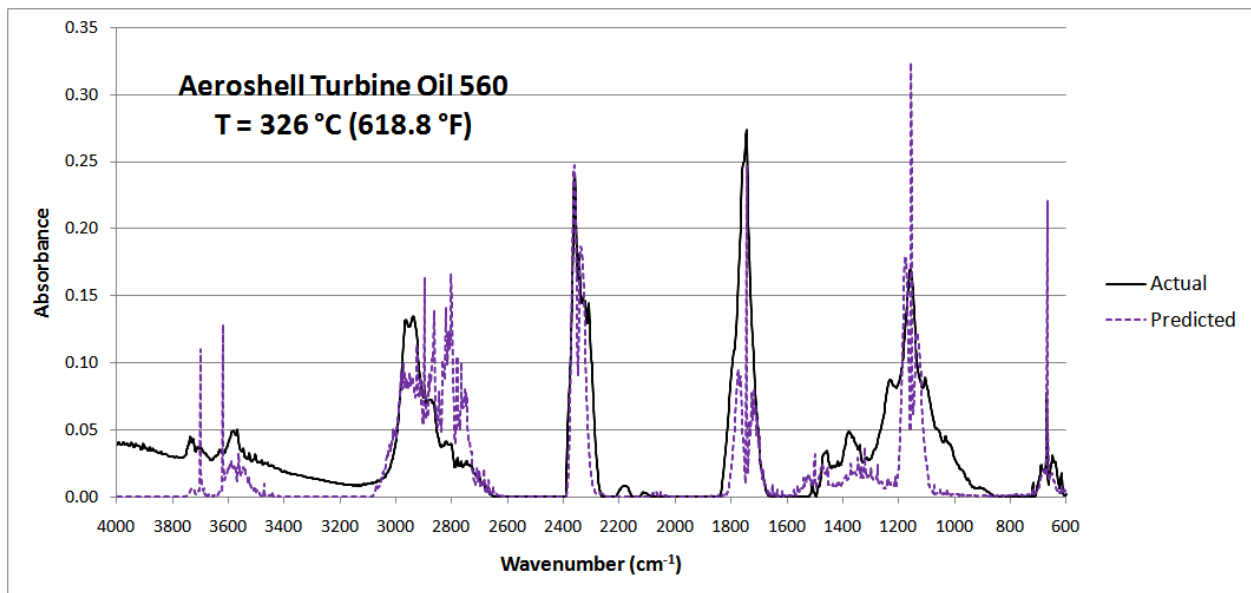


Figure 7-21: Modified Predicted Spectra for Aeroshell Turbine Oil 560 based on PCR Calculated Concentrations of CH_4O , CH_2O , and CO_2 Gas Mixtures; RMSE = 3.4%

Table 7-5: Comparison of RMSE using the Standard CH₄O Spectra and the Modified CH₄O Spectra for Predicted Spectra based on PCR Calculated Concentrations

Sample	RMSE (%) Std. CH ₄ O	RMSE (%) Mod. CH ₄ O
BP Turbo Oil 2380	2.4	2.1
BP Turbo Oil 274	3.1	2.7
Mobile Jet Oil II	2.7	2.2
Aeroshell Turbine Oil 560	3.9	3.4

In an attempt to reduce further the RMSE values of the predicted spectra, PCA along with PCR was conducted on truncated simulated and experimentally collected IR spectra data from 3200 to 1600 cm⁻¹. This removes the IR spectra due to the shifted C-O and O-H stretching bands of the CH₄O component. Within the 3200 to 1600 cm⁻¹ wavenumber range, IR spectra for all three components are still present. The truncated analysis consisted of a set of 10 samples ($n = 10$) at 830 different wavenumbers ($p = 830$) with the same concentrations given in Table 7-1, shown in Figure 7-22, to produce a calibration data set $[X_C]_{(10 \times 830)}$. The experimental data used in data set 1 (Figure 7-1) was truncated and used as the prediction data set $[X_P]_{(4 \times 830)}$ for PCR.

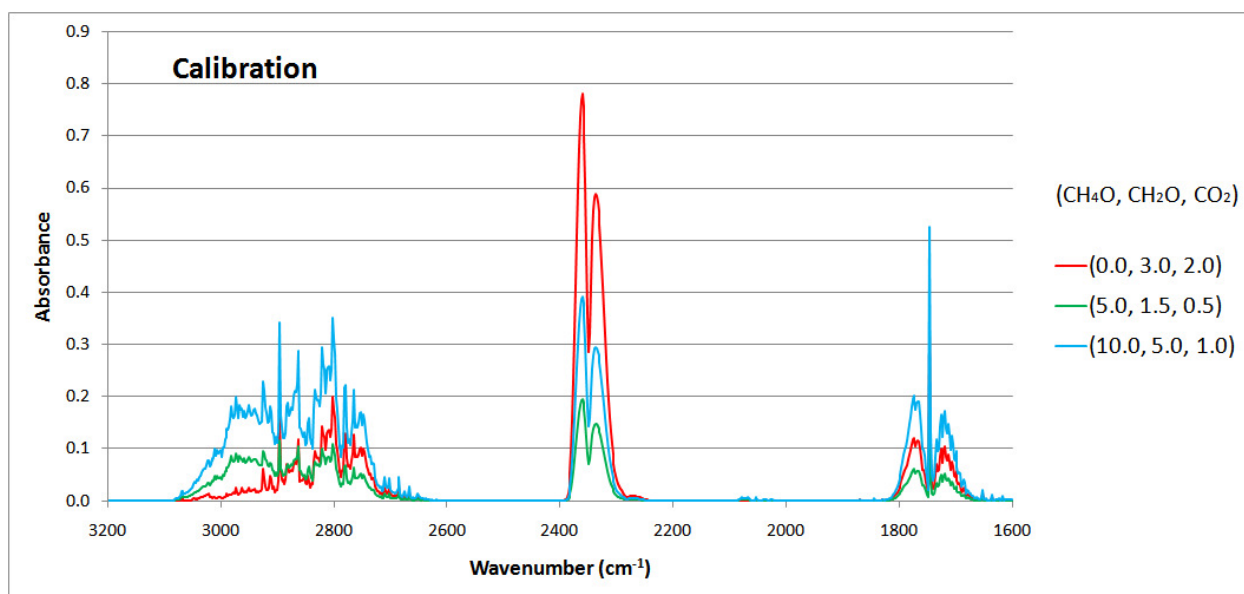


Figure 7-22: Truncated (3200 – 1600 cm⁻¹) CH₄O/CH₂O/CO₂ Gas Mixtures Calibration Data Set, $[X_C]_{(10 \times 830)}$; Note
– Only 3 of 10 Calibration Spectra Shown

After using PCA with PCR to determine concentrations of CH₄O, CH₂O, and CO₂ with the truncated calibration and prediction data sets, shown in Table 7-6, a reconstructed prediction spectrum is calculated. This predicted spectrum is then compared to the experimentally obtained FTIR data as shown in Figures 7-23 thru 7-26 for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560, respectively.

Table 7-6: Truncated (3200 – 1600 cm⁻¹) PCR Calculated Concentrations of CH₄O, CH₂O, and CO₂ for BP Turbo Oil 2380, BP Turbo Oil 274, Mobile Jet Oil II, and Aeroshell Turbine Oil 560

Sample	CH ₄ O Concentration (ppm)	CH ₂ O Concentration (ppm)	CO ₂ Concentration (ppm)
BP Turbo Oil 2380	437	111	31
BP Turbo Oil 274	378	159	50
Mobile Jet Oil II	468	106	21
Aeroshell Turbine Oil 560	567	221	65

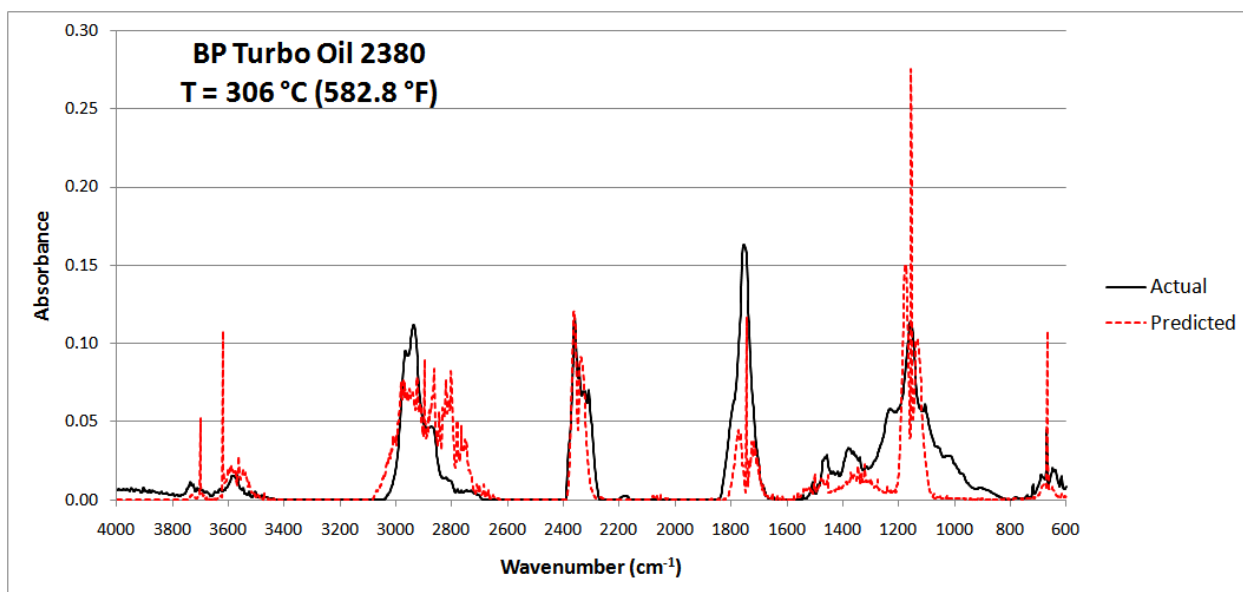


Figure 7-23: Modified Predicted Spectra for BP Turbo Oil 2380 based on Truncated (3200 – 1600 cm⁻¹) PCR Calculated Concentrations of CH₄O, CH₂O, and CO₂ Gas Mixtures; RMSE = 2.1%

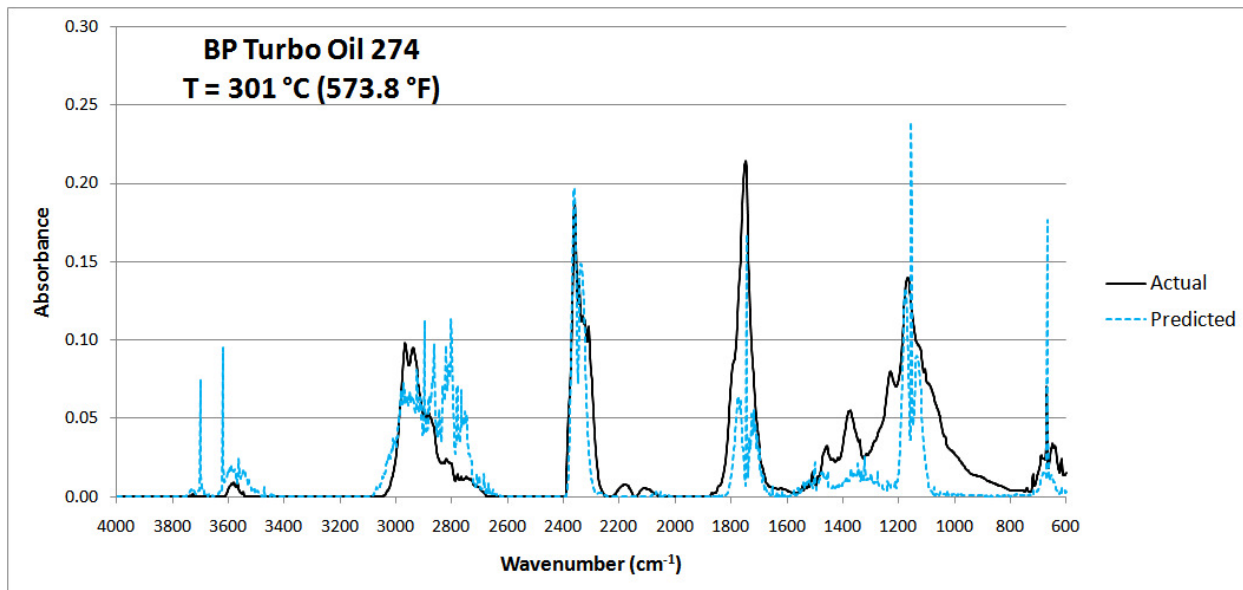


Figure 7-24: Modified Predicted Spectra for BP Turbo Oil 274 based on Truncated (3200 – 1600 cm^{-1}) PCR
 Calculated Concentrations of CH_4O , CH_2O , and CO_2 Gas Mixtures; RMSE = 2.7%

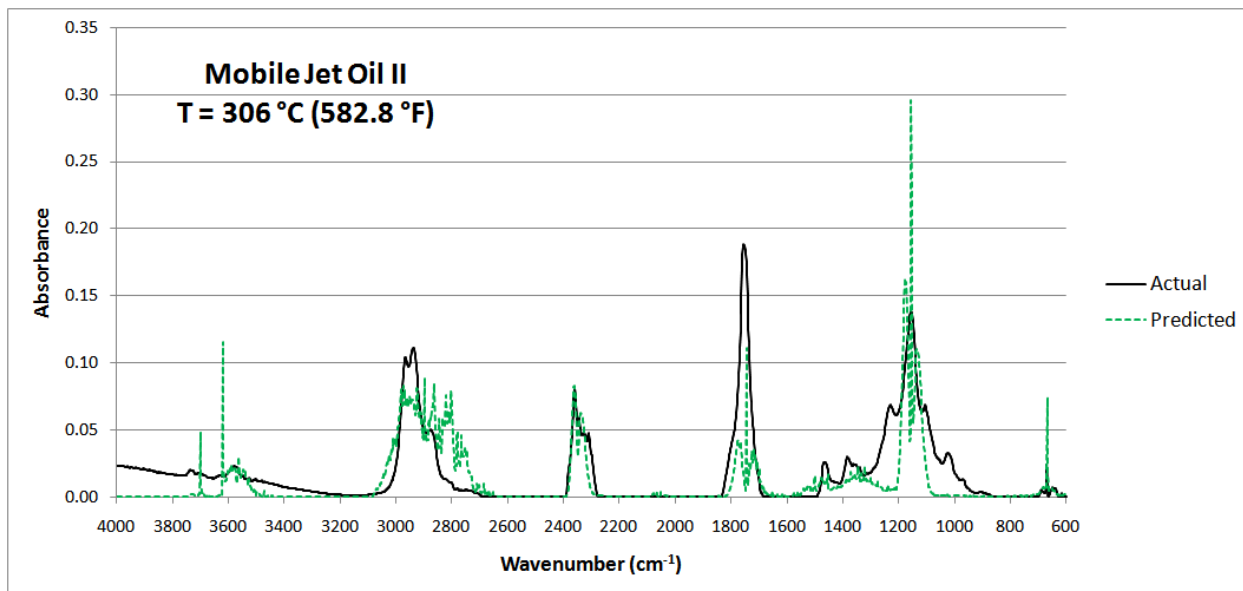


Figure 7-25: Modified Predicted Spectra for Mobile Jet Oil II based on Truncated (3200 – 1600 cm^{-1}) PCR
 Calculated Concentrations of CH_4O , CH_2O , and CO_2 Gas Mixtures; RMSE = 2.3%

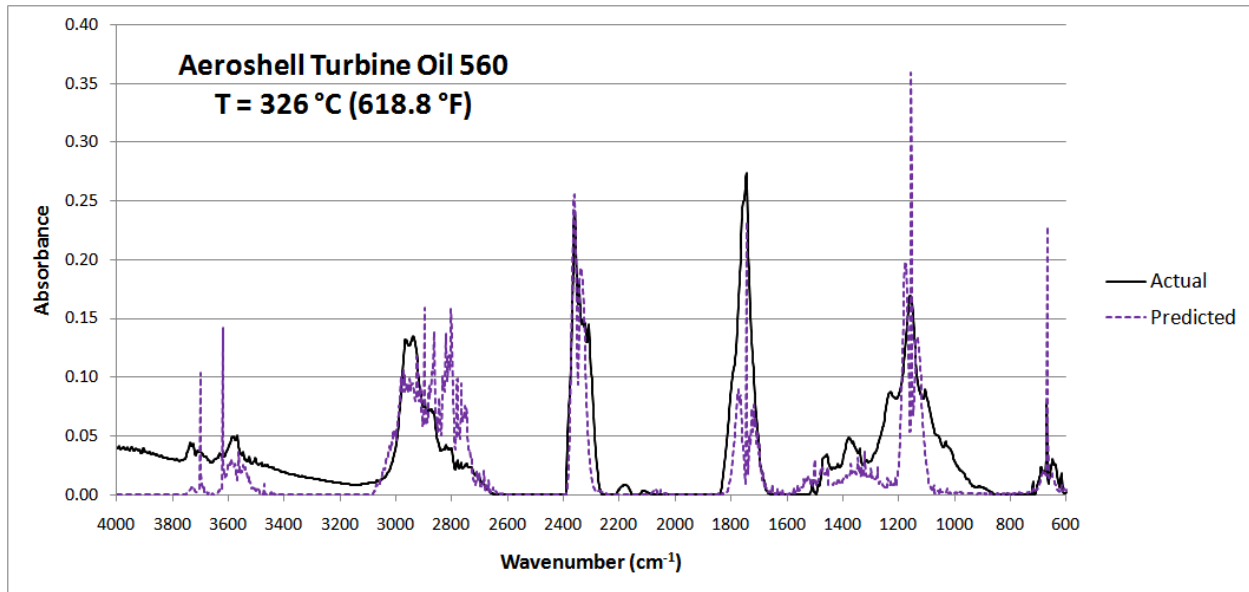


Figure 7-26: Modified Predicted Spectra for Aeroshell Turbine Oil 560 based on Truncated (3200 – 1600 cm^{-1}) PCR Calculated Concentrations of CH_4O , CH_2O , and CO_2 Gas Mixtures; RMSE = 3.4%

A comparison of the RMSE between the predicted and actual spectra using each analysis method for each of the engine oils are shown in Table 7-7. A summary of the calculated concentrations of each of the three components is shown in Tables 7-8 thru 7-11 for each engine oil sample and each method of analysis.

Table 7-7: Comparison of RMSE using the Standard CH_4O Spectra, the Modified CH_4O Spectra, and the Truncated Spectra for Predicted Spectra based on PCR Calculated Concentrations

Sample	RMSE (%) Std. CH_4O	RMSE (%) Mod. CH_4O	RMSE (%) Truncated
BP Turbo Oil 2380	2.4	2.1	2.1
BP Turbo Oil 274	3.1	2.7	2.7
Mobile Jet Oil II	2.7	2.2	2.3
Aeroshell Turbine Oil 560	3.9	3.4	3.4

Table 7-8: PCR Calculated Concentrations of CH_4O , CH_2O , and CO_2 for BP Turbo Oil 2380

Sample	CH_4O Concentration (ppm)	CH_2O Concentration (ppm)	CO_2 Concentration (ppm)
Standard CH_4O Spectra	206	145	30
Modified CH_4O Spectra	350	127	30
Truncated Spectra	437	111	31

Table 7-9: PCR Calculated Concentrations of CH₄O, CH₂O, and CO₂ for BP Turbo Oil 274

Sample	CH ₄ O Concentration (ppm)	CH ₂ O Concentration (ppm)	CO ₂ Concentration (ppm)
Standard CH ₄ O Spectra	221	190	49
Modified CH ₄ O Spectra	417	166	49
Truncated Spectra	378	159	50

Table 7-10: PCR Calculated Concentrations of CH₄O, CH₂O, and CO₂ for Mobile Jet Oil II

Sample	CH ₄ O Concentration (ppm)	CH ₂ O Concentration (ppm)	CO ₂ Concentration (ppm)
Standard CH ₄ O Spectra	214	142	20
Modified CH ₄ O Spectra	407	118	20
Truncated Spectra	468	106	21

Table 7-11: PCR Calculated Concentrations of CH₄O, CH₂O, and CO₂ for Aeroshell Turbine Oil 560

Sample	CH ₄ O Concentration (ppm)	CH ₂ O Concentration (ppm)	CO ₂ Concentration (ppm)
Standard CH ₄ O Spectra	292	263	63
Modified CH ₄ O Spectra	511	235	63
Truncated Spectra	567	221	65

Performing PCA with the modified CH₄O spectra in the calibration data set improves the RMSE for the predicted engine oil samples, in addition to calculating a higher concentration of CH₄O in each of the engine oil samples. The higher concentration is due to a better alignment of the peaks that shift in the CH₄O component. With more of the spectra of the engine oil samples being attributed to the CH₄O component, the calculated concentrations for the CH₂O component slightly decrease. When the analysis is performed with a truncated calibration data set, the CH₄O concentrations are slightly increased and the CH₂O concentrations are further reduced. The concentration values calculated with the truncated method have the highest probability of best representing the actual values since they do not analyze portions of the spectra that contain peak-shifted areas. An additional source of error for concentration calculations in the prediction data sets can be attributed to the broadening of the peaks due to the increase in the gas temperature. The calculated amount of CO₂ for each of the three cases, standard CH₄O,

modified CH_4O , and truncated spectrum, is unaffected because the CO_2 IR spectra do not overlap with either CH_4O or the CH_2O spectra. This analysis indicates that PCA on simulated calibration data sets is capable of identifying and quantifying gas species within experimental and unknown prediction data sets.

Data Set 3 – Simulated $\text{CH}_4\text{O}/\text{CH}_2\text{O}/\text{CO}_2/\text{CO}/\text{H}_2\text{O}$ Gas Mixtures & BP Turbo Oil 2380 Engine Oil Time-Evolved Samples

From the simulated FTIR spectral analysis of the pure $\text{CH}_4\text{O}/\text{CH}_2\text{O}/\text{CO}_2/\text{CO}/\text{H}_2\text{O}$ gas mixtures shown in Figure 7-27, a set of 10 samples ($n = 10$) at 1,763 different wavenumbers ($p = 1,763$) with concentrations given in Table 7-12 were calculated to produce a calibration data set $[\text{X}_\text{C}]_{(10 \times 1763)}$. The experimental data consisting of time evolved IR spectra of the heated BP 2380 engine oil (Figure 7-28) was used as the prediction data set $[\text{X}_\text{P}]_{(20 \times 1763)}$ for PCR.

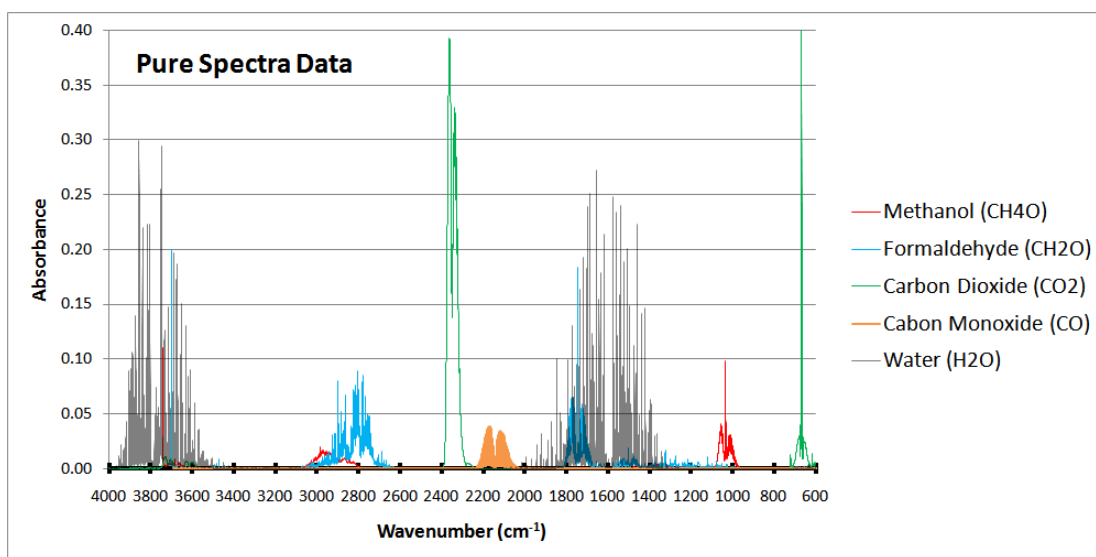


Figure 7-27: Methanol (CH_4O), Formaldehyde (CH_2O), Carbon Dioxide (CO_2), Carbon Monoxide (CO), and Water (H_2O) Pure Spectra for Simulated Data Set Illustrating Spectral Overlap Between CH_4O , CH_2O , and H_2O

Table 7-12: Calibration Data Set 2 Composed of Simulated Spectra of Various Concentrations of Methanol (CH₄O), Formaldehyde (CH₂O), Carbon Dioxide (CO₂), Carbon Monoxide (CO), and Water (H₂O)

Sample	CH ₄ O Concentration (x100 ppm)	CH ₂ O Concentration (x100 ppm)	CO ₂ Concentration (x100 ppm)	CO Concentration (x100 ppm)	H ₂ O Concentration (x1000 ppm)
1	0.0	3.0	2.0	4.5	5.0
2	0.5	2.0	4.0	6.0	6.5
3	1.0	0.0	3.5	0.5	1.0
4	2.0	6.0	5.5	0.0	3.5
5	3.5	7.0	0.0	1.5	2.0
6	5.0	1.5	0.5	0.6	3.0
7	6.5	3.5	4.5	2.0	0.0
8	8.0	4.5	2.5	3.0	0.5
9	9.0	0.5	3.0	0.2	4.5
10	10.0	5.0	1.0	5.0	6.0

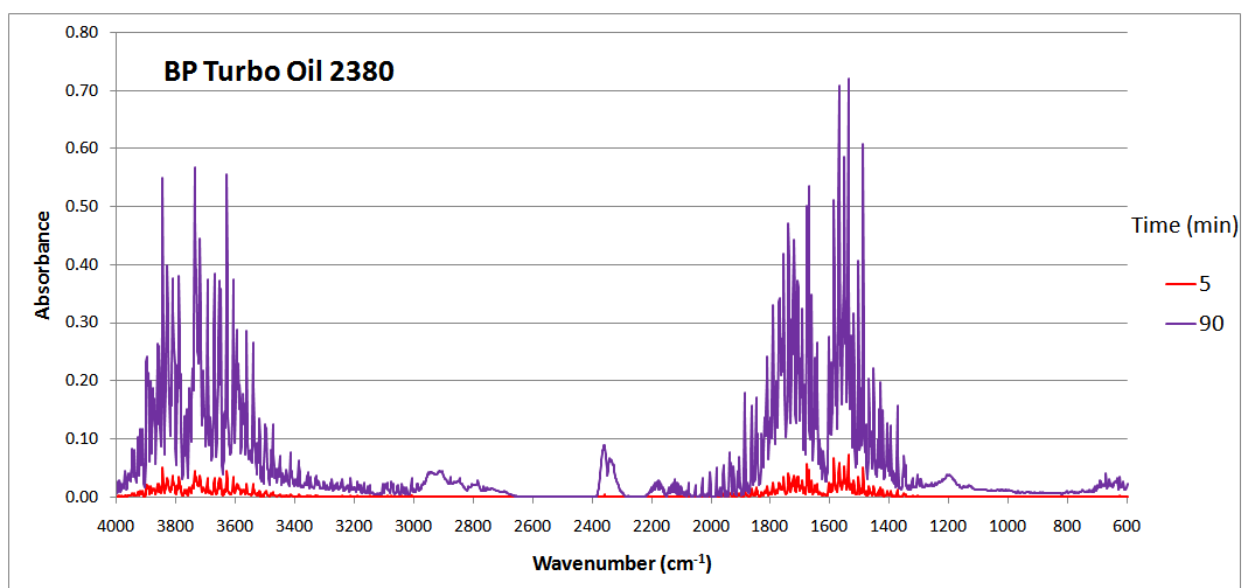


Figure 7-28: BP Turbo Oil 2390 Time Evolved Spectra Used for Prediction Data Set, [X_P]_(20 x 1763); Note – Only 2 of 20 Prediction Spectra Shown, Time 5 minutes and 90 minutes

Applying PCR, a plot can be created that shows the calibration data set used to determine the concentrations in the prediction data sets. The RMSE for the calibration data set in regards to concentrations of CH₄O, CH₂O, CO₂, CO, and H₂O are found to be 0 ppm (simulated data sets).

Figure 7-29 shows the predicted CH₄O, CH₂O, and CO concentrations for the time evolved BP Turbo Oil 2380, while Figures 7-30 and 7-31 shown the predicted CO₂ and H₂O time evolved

concentrations, respectively. Noted on the figures are the times at which the heater reached the set point (Time = 25 min.) and the time at which the heater was turned off (Time = 90 min.).

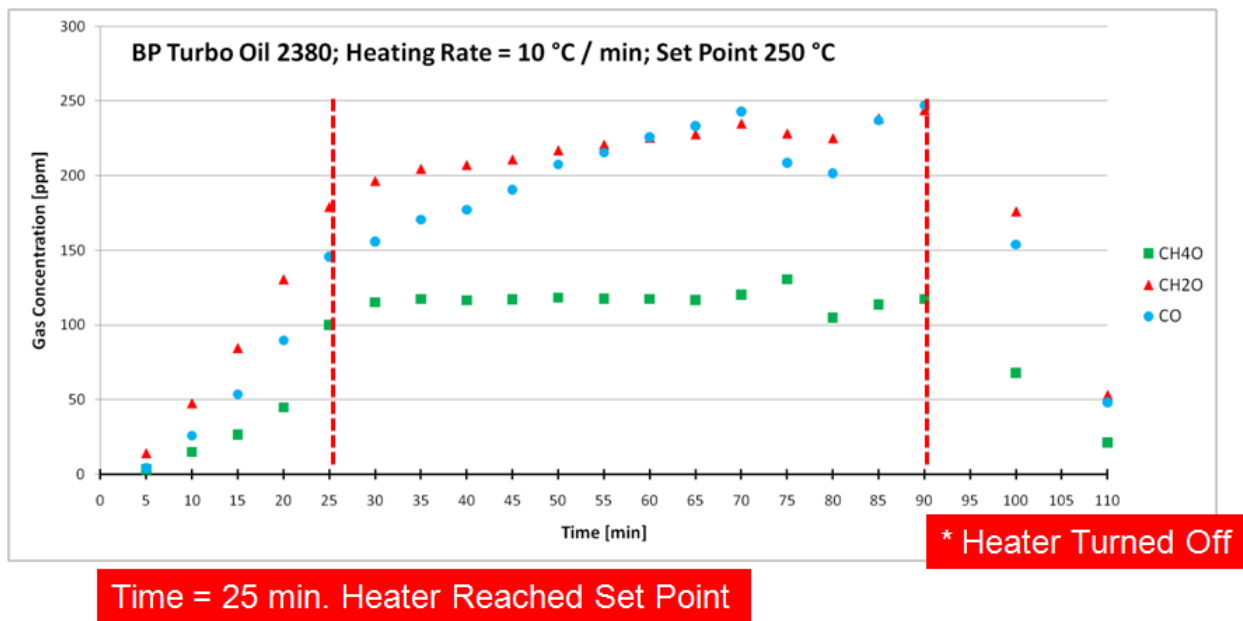


Figure 7-29: Principal Component Regression (PCR) Calculated Gas Concentrations for CH₄O, CH₂O, and CO of BP Turbo Oil 2380 Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off

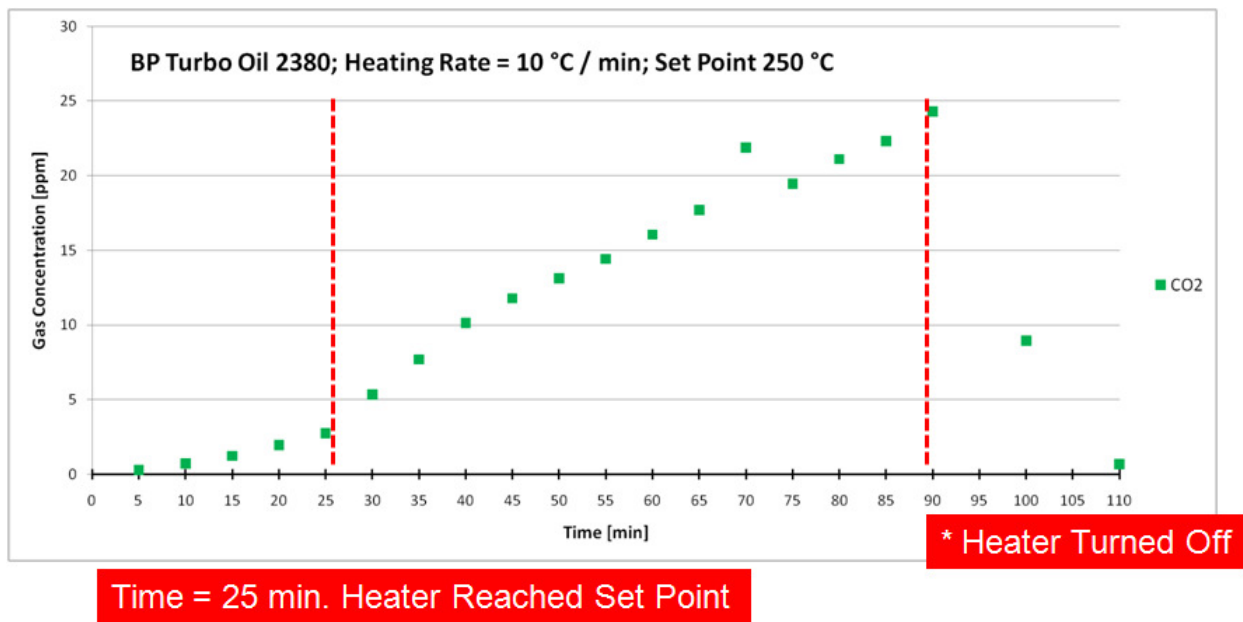


Figure 7-30: Principal Component Regression (PCR) Calculated Gas Concentrations for CO₂ of BP Turbo Oil 2380
Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off

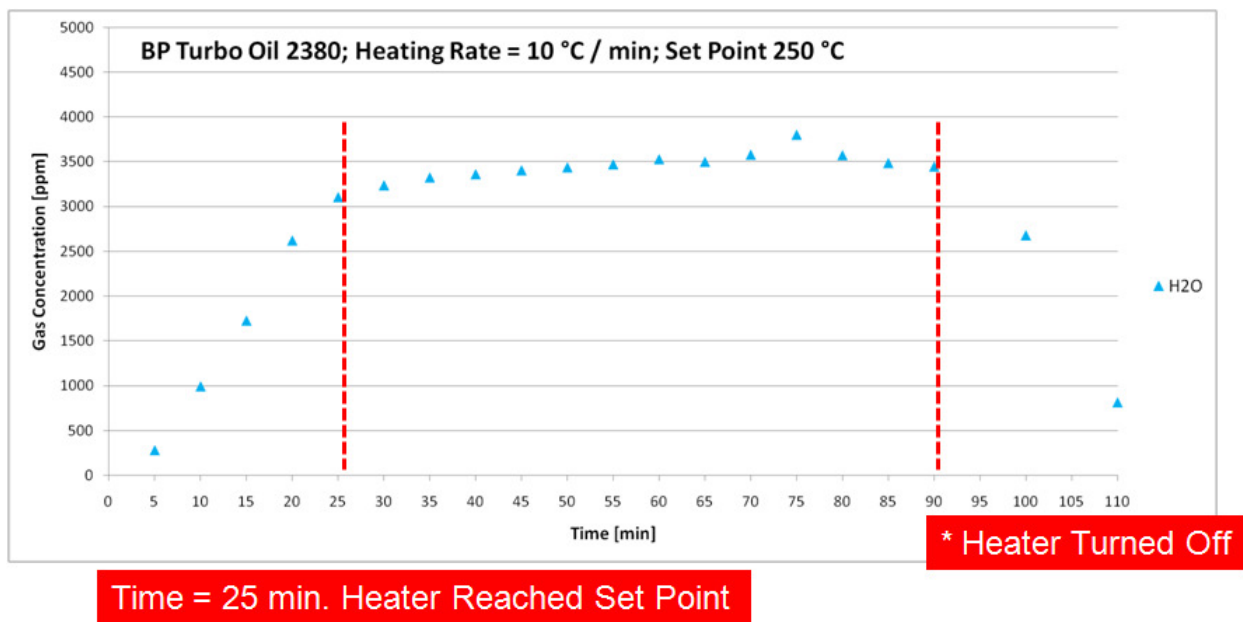


Figure 7-31: Principal Component Regression (PCR) Calculated Gas Concentrations for H₂O of BP Turbo Oil 2380
Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off

After using PCR to determine concentrations for CH₄O, CH₂O, CO₂, CO, and H₂O of the BP Turbo Oil 2380 time evolved samples, reconstructed prediction spectra are calculated. These reconstructed spectra are then compared to the experimentally obtained FTIR data as shown in Figures 7-32 thru 7-35 for BP Turbo Oil 2380 at Time = 10 min., 30 min., 60 min., and 90 min. The RMSE between the predicted and actual spectra for each of the time-evolved spectra are calculated to be 2.7%, 8.4%, 9.2%, and 9.0% for 10 min., 30 min., 60min., and 90 min., respectively. The average RMSE between the predicted and actual spectra for all 20 of the time-evolved spectra is found to be 7.5%.

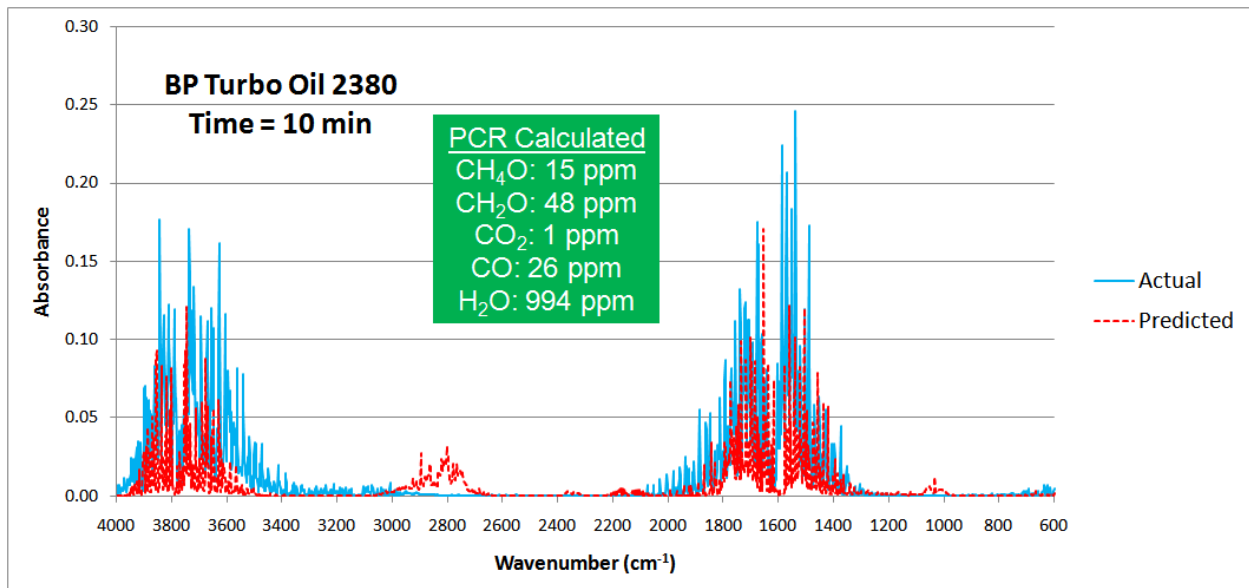


Figure 7-32: Predicted Spectra for BP Turbo Oil 2380 at Time = 10 min. based on PCR Calculated Concentrations of CH₄O, CH₂O, CO₂, CO, and H₂O Gas Mixtures; RMSE = 2.7%

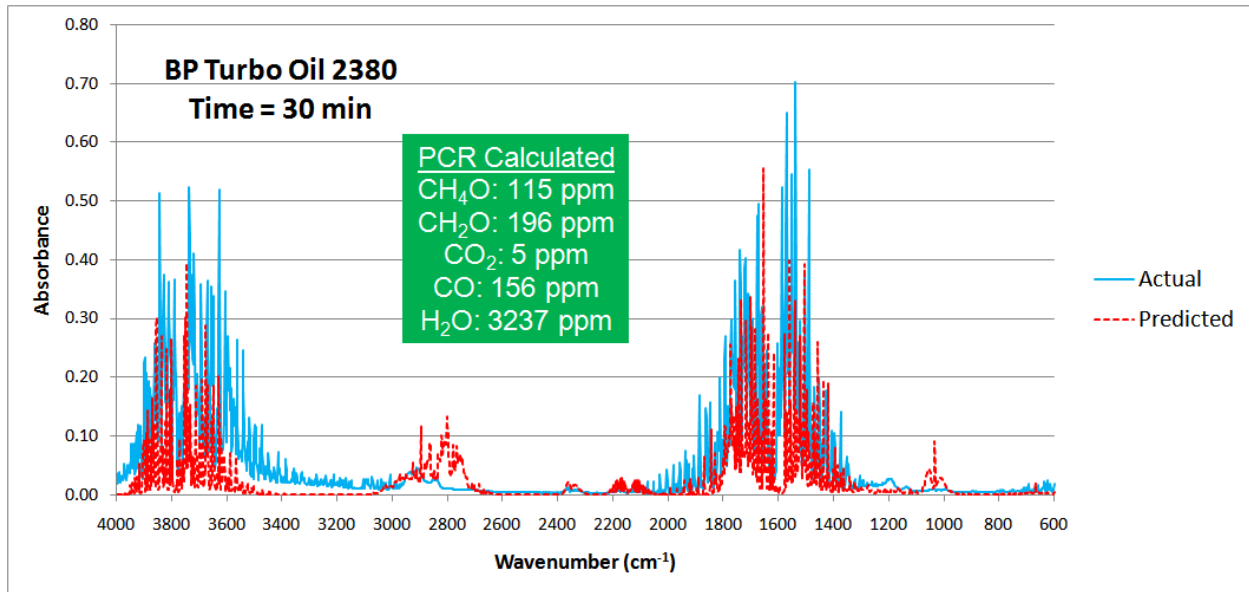


Figure 7-33: Predicted Spectra for BP Turbo Oil 2380 at Time = 30 min. based on PCR Calculated Concentrations of CH₄O, CH₂O, CO₂, CO, and H₂O Gas Mixtures; RMSE = 8.4%

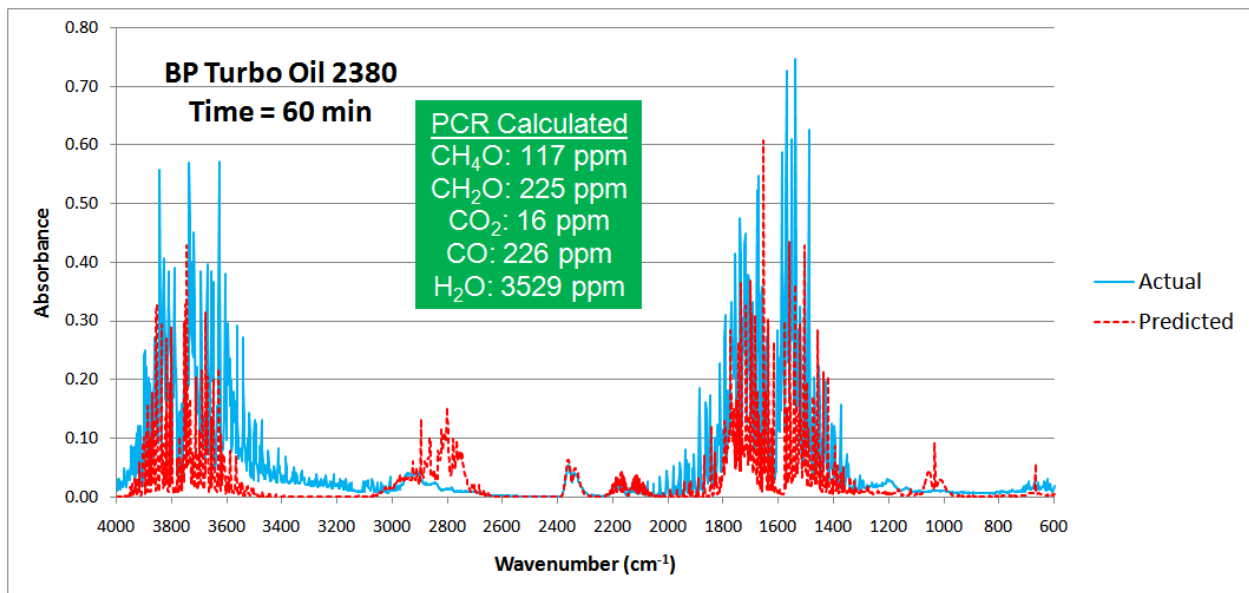


Figure 7-34: Predicted Spectra for BP Turbo Oil 2380 at Time = 60 min. based on PCR Calculated Concentrations of CH₄O, CH₂O, CO₂, CO, and H₂O Gas Mixtures; RMSE = 9.2%

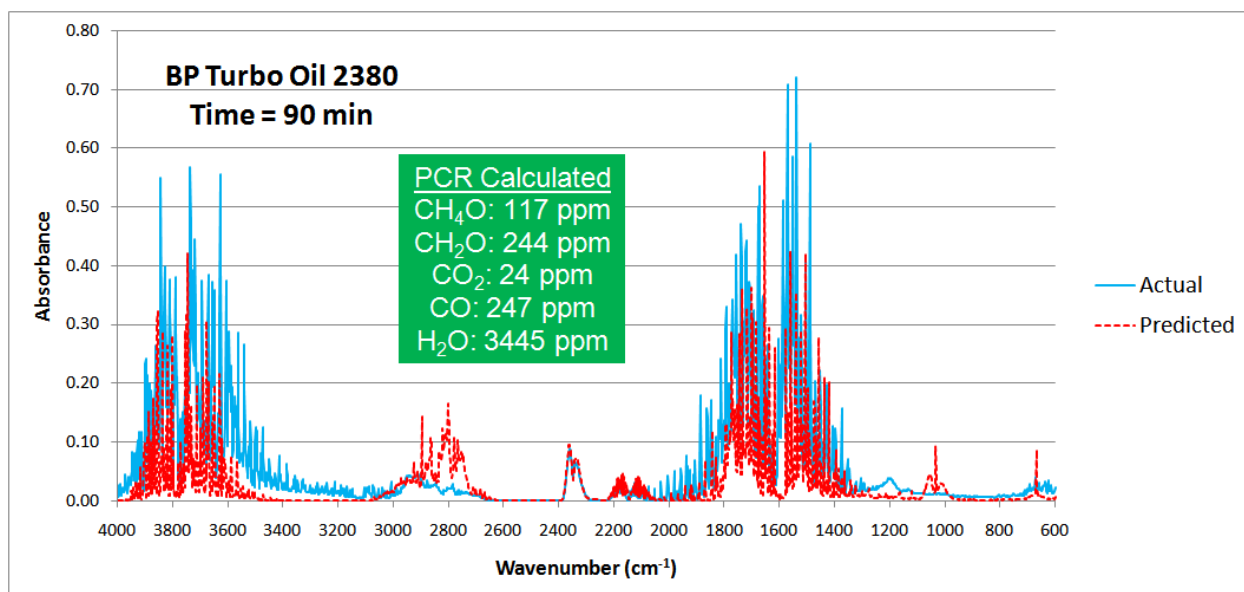


Figure 7-35: Predicted Spectra for BP Turbo Oil 2380 at Time = 90 min. based on PCR Calculated Concentrations of CH₄O, CH₂O, CO₂, CO, and H₂O Gas Mixtures; RMSE = 9.0%

In an attempt to reduce the RMSE values of the predicted spectra, PCA along with PCR was conducted on truncated simulated and experimentally collected IR spectra data from 4000 to 2000 cm⁻¹. This removes the region in IR spectra that has considerable overlap between the CH₂O and H₂O components. Within the 4000 to 2000 cm⁻¹ wavenumber range, characteristics of IR spectra for all five components are still present. The truncated analysis consisted of a set of 10 samples ($n = 10$) at 1037 different wavenumbers ($p = 1037$) with the same concentrations given in Table 7-12, to produce a calibration data set $[X_C]_{(10 \times 1037)}$. The BP 2390 time-evolved spectra data was truncated and used as the prediction data set $[X_P]_{(20 \times 1037)}$ for PCR. Figure 7-36 shows the predicted CH₄O, CH₂O, and CO concentrations from the truncated analysis for the time evolved BP Turbo Oil 2380, while Figure 7-37 shows the predicted H₂O time evolved concentrations from the truncated analysis, respectively. The predicted CO₂ time evolved concentrations from the truncated analysis does not change significantly from Figure 7-30.

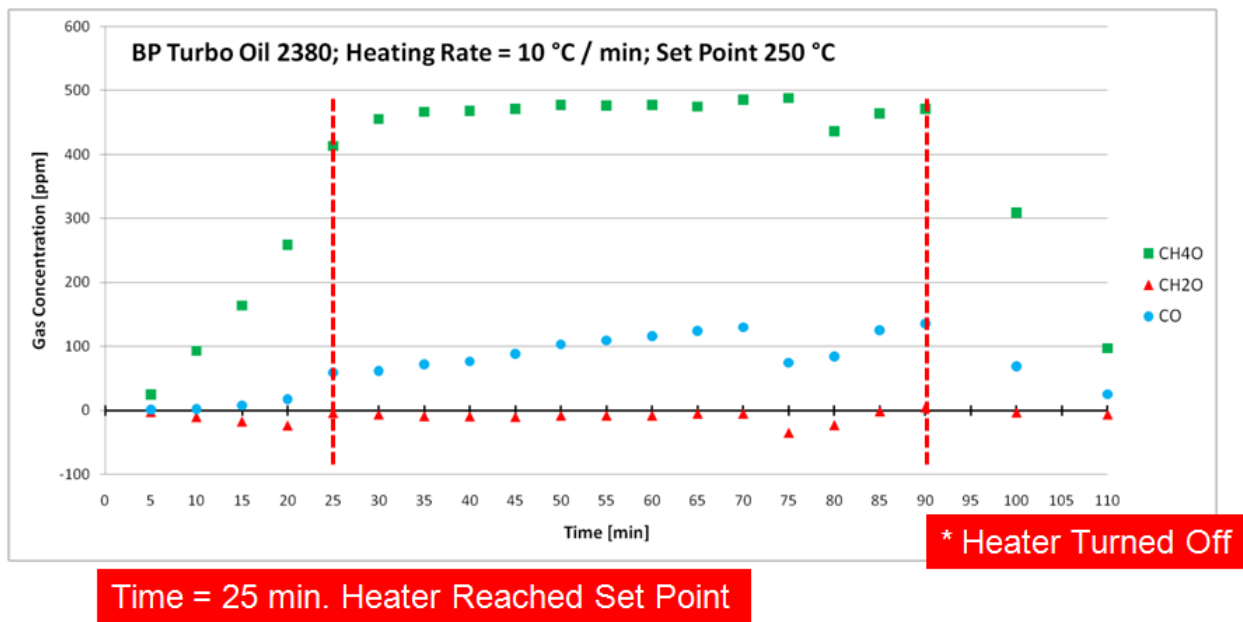


Figure 7-36: Principal Component Regression (PCR) Calculated Gas Concentrations for CH₄O, CH₂O, and CO of BP Turbo Oil 2380 Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off

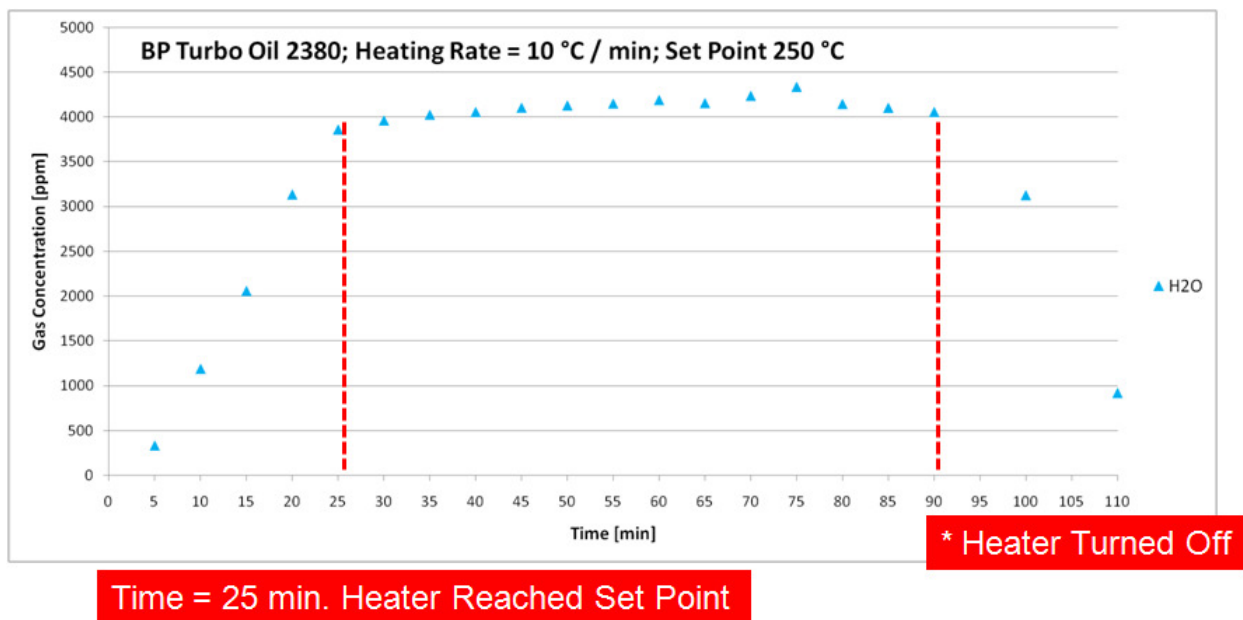


Figure 7-37: Principal Component Regression (PCR) Calculated Gas Concentrations for H₂O of BP Turbo Oil 2380 Time Evolved Spectra; Time = 25 min. Heater Reached Set Point, Time = 90 min. Heater Turned Off

After using PCR to determine concentrations for CH₄O, CH₂O, CO₂, CO, and H₂O of the BP Turbo Oil 2380 time evolved samples, reconstructed prediction spectra are calculated. These reconstructed spectra are then compared to the experimentally obtained FTIR data as shown in Figures 7-38 thru 7-41 for BP Turbo Oil 2380 at Time = 10 min., 30 min., 60 min., and 90 min. The RMSE between the predicted and actual spectra for each of the time-evolved spectra are calculated to be 2.2%, 7.2%, 7.8%, and 7.5% for 10 min., 30 min., 60min., and 90 min., respectively. The average RMSE between the predicted and actual spectra for all 20 of the time-evolved spectra is found to be 6.4%. A comparison of the RMSE between the predicted and actual spectra using full and truncated spectral analysis method for each of the time-evolved samples are shown in Table 7-13.

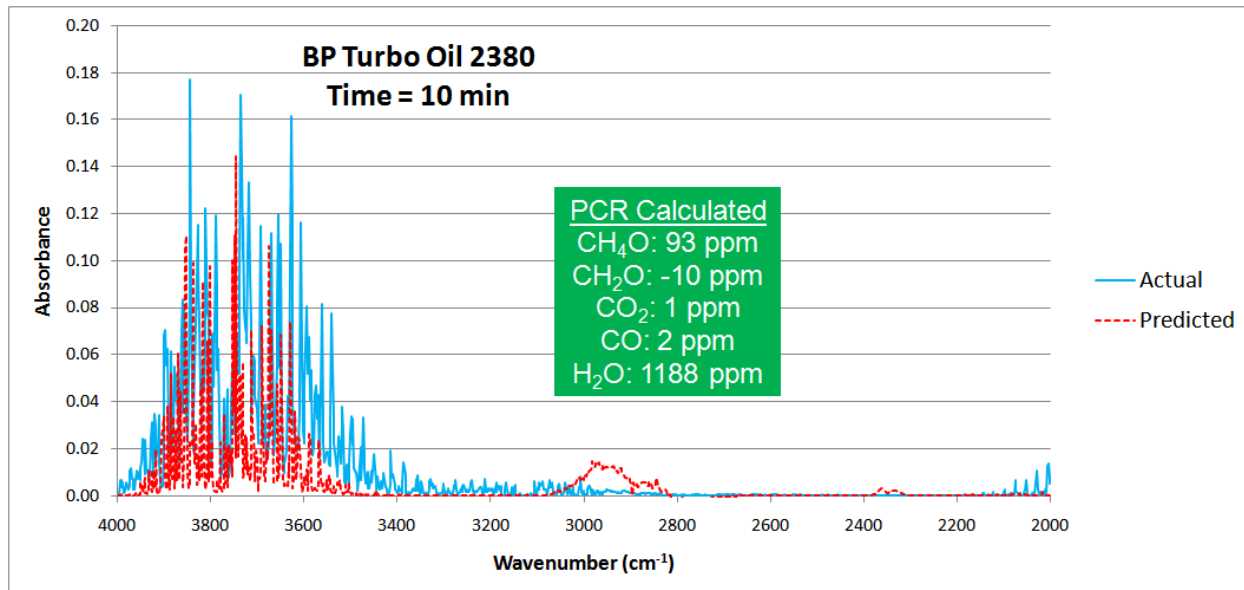


Figure 7-38: Predicted Spectra for BP Turbo Oil 2380 at Time = 10 min. based on Truncated (4000-2000 cm⁻¹) PCR Calculated Concentrations of CH₄O, CH₂O, CO₂, CO, and H₂O Gas Mixtures; RMSE = 2.2%

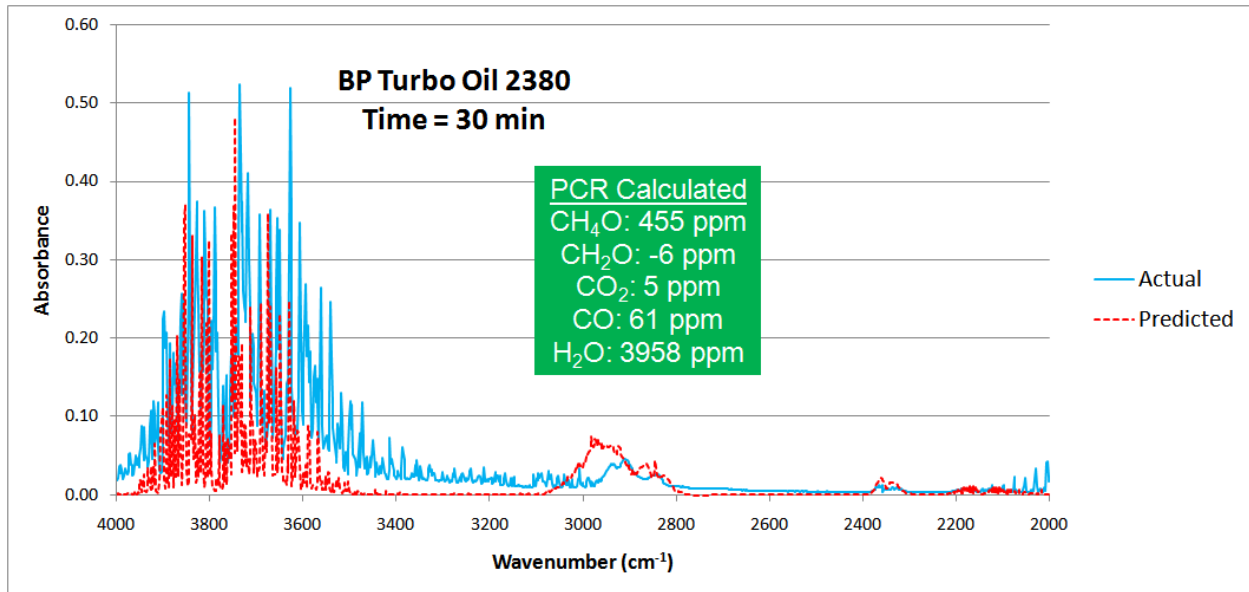


Figure 7-39: Predicted Spectra for BP Turbo Oil 2380 at Time = 30 min. based on Truncated (4000-2000 cm⁻¹) PCR
 Calculated Concentrations of CH₄O, CH₂O, CO₂, CO, and H₂O Gas Mixtures; RMSE = 7.2%

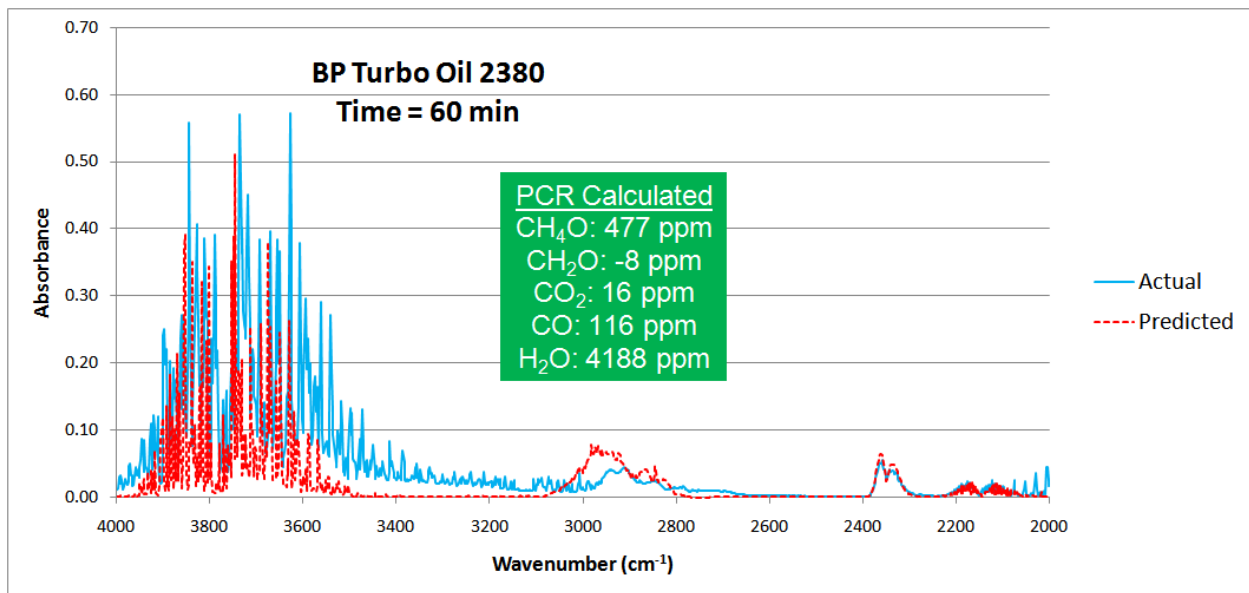


Figure 7-40: Predicted Spectra for BP Turbo Oil 2380 at Time = 60 min. based on Truncated (4000-2000 cm⁻¹) PCR
 Calculated Concentrations of CH₄O, CH₂O, CO₂, CO, and H₂O Gas Mixtures; RMSE = 7.8%

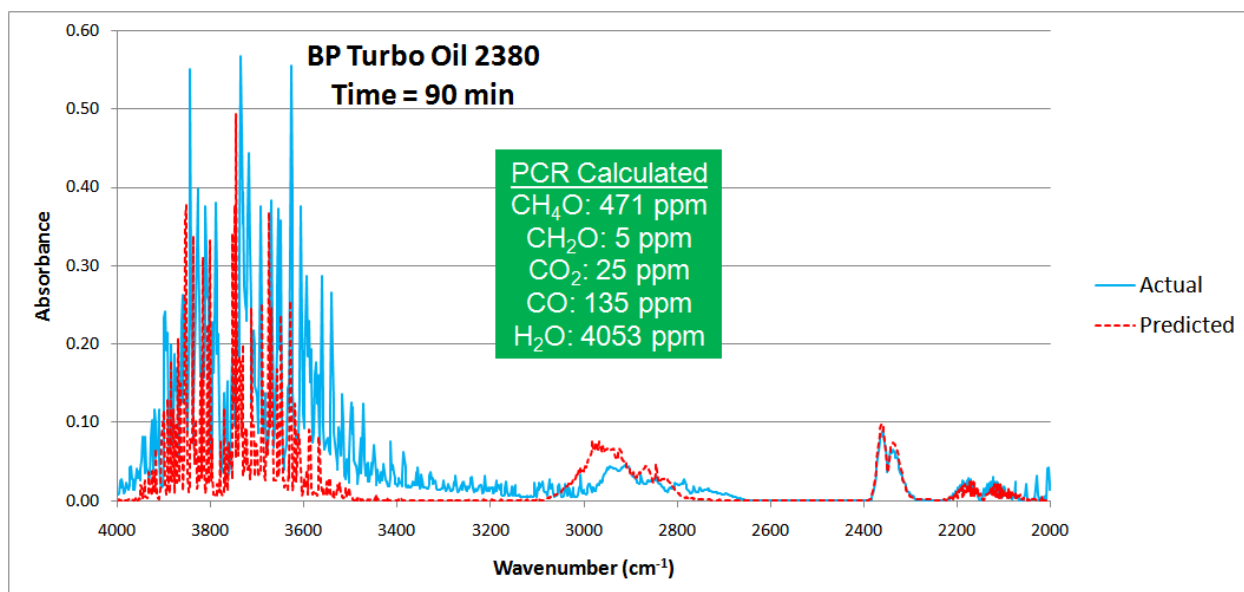


Figure 7-41: Predicted Spectra for BP Turbo Oil 2380 at Time = 90 min. based on Truncated (4000-2000 cm^{-1}) PCR Calculated Concentrations of CH_4O , CH_2O , CO_2 , CO , and H_2O Gas Mixtures; RMSE = 7.5%

Table 7-13: Comparison of RMSE using the Full and the Truncated Spectra, for the PCR Calculated Concentrations

Sample	RMSE (%)	RMSE (%)
	Full Spectra	Truncated
Time = 10 min.	2.7	2.2
Time = 30 min.	8.4	7.2
Time = 60 min.	9.2	7.8
Time = 90 min.	9.0	7.5
Average over 20 Samples	7.5	6.4

Performing PCA with the truncated spectra in the calibration data set improves the RMSE (reduced from an average of 7.5% to 6.4%) for the predicted time-evolved BP 2380 Turbo Oil samples, in addition to calculating a higher concentration of CH_4O in each of the engine oil samples. The higher concentration is due to a more accurate calculation concentration of the CH_2O component when the significant overlap region between H_2O and CH_2O is not analyzed. With more of the spectra of the engine oil samples being attributed to the H_2O component, the calculated concentrations for the CH_2O component significantly decrease. As was the case with the data set 2, the concentration values calculated with the truncated method

have the highest probability of best representing the actual values since they do not analyze portions of the spectra that contain areas of significant overlap with a H₂O, which is a strong IR absorbing gas. In addition to improving the calculated concentration of CH₂O, removal of the wavenumbers below 2000 cm⁻¹ better quantifies the CO concentration in the time-evolved gas mixture. This is due to the relatively small but critical overlap between the CO and H₂O IR spectra. The calculated amount of CO₂ for each of the two cases, full and truncated spectrum, is unaffected because the CO₂ IR spectra do not overlap with either CH₄O, CH₂O, CO or the H₂O spectra. This analysis indicates that PCA on simulated calibration data sets is capable of identifying and quantifying gas species within experimental and unknown prediction data sets.

Chapter 8: Conclusions

The specific purpose of this research was to utilize a mathematical technique called principal component analysis (PCA) in conjunction with principal component regression (PCR) and proportionality constant calculations (PCC) to simplify complex, multi-component infrared (IR) spectra data sets into a reduced data set used for determination of the concentrations of the individual components. The application of this analytical numerical technique to IR spectrum analysis could play an important role in improving performance of commercial sensors that airlines and aircraft manufacturers could potentially use in an aircraft cabin environment for multi-gas component monitoring.

PCA along with PCR and PCC was successfully applied to the monitoring of H_2O_2 concentration in an aqueous solution in which analysis was performed on variable volumes of solutions in both the calibration and prediction data set. The analysis was then applied to both simulated and experimental two and three component gas systems that could be potential environmental air contaminants within the aircraft cabin. These analyzed systems consisted of mixtures of $\text{CH}_2\text{O}/\text{C}_3\text{H}_4\text{O}$, CO/CO_2 , $\text{CH}_2\text{O}/\text{C}_3\text{H}_4\text{O}/\text{H}_2\text{O}$, and $\text{CO}/\text{CO}_2/\text{H}_2\text{O}$ gas spectra. After the PCA application to two and three component systems, the technique was further expanded to include the monitoring of potential bleed air contaminants from engine oil combustion, in which a simulation data set was utilized to predict gas components and concentrations in unknown engine oil samples at high temperatures as well as time-evolved gas from heating of engine oil.

Chapter 9: Future Work

Based on the success of PCA application to the H₂O₂ solutions, it is recommended that the engine oils, as received, in liquid form be analyzed. This analysis should indicate the major components of the engine oil and these can then be compared to the gas species that evolved when the liquid engine oil is heated up to the point of combustion. Another segment of this future work should be the development of a model that accurately simulates the observed shifts in methanol at higher temperatures. Within this research, the shift was estimated for the particular temperature that the engine oils experienced the greatest mass loss. With further experiments monitoring a much wider temperature profile, the movement in the characteristic IR spectra of methanol should be readily observed.

Further experiments should be performed where analysis of the gas phase products released at various temperatures are monitored with the FTIR at or near room temperature. The experimental setup best simulate the monitoring of the bleed air contaminants as would be measured in the aircraft cabin. Additional holding temperatures of the heating furnace used in the time-evolved engine oil study should be investigated to determine if other gas components are released and if the concentrations of the gas species significantly change as a function of temperature. In addition, other aircraft engine oils should be investigated to determine characteristics of each.

As discussed in Chapter 2, engine oils are known to contain a toxic chemical called tricresyl phosphate (TCP) and to understand more thoroughly its IR characteristics within the

aircraft cabin environment, experiments similar to those performed with the engine oil should be carried out. TCP in the liquid phase, diluted with methanol, would be an ideal solution to monitor in the gas phase with FTIR to determine characteristics temperatures that the TCP solution could be released within an aircraft cabin.

References

- [1] Airline News Resource (October 2010). U.S. Business Travel Spending, Trips to Increase This Year Despite Slowdown in Economic Growth. *The Business Travel Quarterly Outlook – United States*, Retrieved November 1, 2010 from <http://www.airlinenewsresource.com/article49669.html>
- [2] Federal Aviation Administration (March 2010). Forecast Highlights 2010-2030. *FAA Aerospace Forecast Fiscal Years 2010-2030*, Retrieved November 3, 2010 from http://www.faa.gov/data_research/aviation/aerospace_forecasts/2010-2030
- [3] McAllister, B. (February 2004). Al Qaeda and the Innovative Firm: Demythologizing the Network. *Studies in Conflict & Terrorism*, 27, 297-319.
- [4] Kean, T.H., Hamilton, L. H., Ben-Veniste, R., Kerrey, B., Fielding, F.F., Lehman, J. F., Gorelick, J. S., Roemer, T. J., Gorton, S., and Thompson, J. R. (2004). The 9/11 Commission Report. Retrieved November 1, 2010 from <http://govinfo.library.unt.edu/911/report/911Report.pdf>
- [5] Hwang, G. M., DiCarlo, A., Teig, L. J., Lin, G., and Harkin, M. (May 2009). Detecting infectious and biological contaminants aboard aircraft – Is it feasible? *IEEE Conference on Technologies for Homeland Security*, 477-484.
- [6] Gupta, J. K., Lin, C. –H., and Chen, Q. (June 2010). Transport of expiratory droplets in an aircraft cabin. *International Journal of Indoor Environment and Health*, Retrieved November 2, 2010 from <http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0668.2010.00676.x/pdf>
- [7] Federal Aviation Administration (March 2010). Review of 2009. *FAA Aerospace Forecast Fiscal Years 2010-2030*, Retrieved November 3, 2010 from http://www.faa.gov/data_research/aviation/aerospace_forecasts/2010-2030
- [8] Netten, C. V., and Leung, V. (March/April 2001). Hydraulic Fluids and Jet Engine Oil: Pyrolysis and Aircraft Air Quality. *Environmental Health*, 56(2), 181-186.
- [9] Spengler, J. D., and Wilson, D. G. (April 2003). Air quality in aircraft. *Proceedings Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, 217, 323-335.

- [10] Winder, C., and Balouet, J. (June 2002). The Toxicity of Commercial Jet Oils. *Environmental Research*, 89(2), 146-164.
- [11] Hunt, E., Reid, D., Space, D., and Tilton, F. (n.d.). The Commercial Airliner Environmental Control System: Engineering Aspects of Cabin Air Quality. Retrieved November 24, 2010 from <http://www.boeing.com/commercial/cabinair/ecs.pdf>
- [12] Harrison, R., Murawski, J., McNeely, E., Guerriero, J., and Milton, D. (April 2009). Exposure to Aircraft Bleed Air Contaminants Among Airline Workers: A Guide for Health Care Providers. Retrieved November 2, 2010 from <http://www.ohrca.org/Medicalprotocol031909.pdf>
- [13] Prorok, B., Gale, W. F., Gale, H. S., Simonian, A., Kim, D. -J., Hong, J. -W., Cheng, Z. -Y., Callender, C. M., Sofyan, N., Loo, N. M., Kim, Y. E., Low, P., and Reifenberger, R. R. (November 2006). *Proceedings of 4th International Aviation Security Technology Symposium*. Retrieved November 2, 2010 from http://acer.eng.auburn.edu/partners/conf_papers/prorok_paper_4th_int_aviation_security_sym_nov_2006.pdf
- [14] Sinnett, M. (2007). 787 No Bleed Systems: Saving Fuel and Enhancing Operational Efficiencies. *Aero Quarterly*. Retrieved November 8, 2010 from http://www.boeing.com/commercial/aeromagazine/articles/qtr_4_07/AERO_Q407_article2.pdf
- [15] Shlens, J. (April 2009). A Tutorial on Principal Component Analysis. Retrieved November 8, 2010 from <http://www.snl.salk.edu/~shlens/pca.pdf>
- [16] Hunt, E., and Space, D. (n.d.). The Airplane Cabin Environment: Issues Pertaining to Flight Attendant Comfort. The Boeing Company. Retrieved December 6, 2010 from <http://www.boeing.com/commercial/cabinair/ventilation.pdf>
- [17] Tamas, G., Weschler, C., Bako-Biro, Z., Wyon, D., and Strom-Tejsen, P. (October 2006). Factors affecting ozone removal rates in a simulated aircraft cabin environment. *Atmospheric Environment*, 40(32), 6122-6133.
- [18] Strom-Tejsen, P., Weschler, C., Wargocki, P., Myskow, D., and Zarzycka, J. (June 2007). The influence of ozone on self-evaluation of symptoms in a simulated aircraft cabin. *Journal of Exposure Science and Environmental Epidemiology*, 18(3), 272-281.
- [19] Thibeault, C. (January 1997). Special Committee Report: Cabin Air Quality. *Aviation, Space, and Environmental Medicine*, 68(1), 80-82.
- [20] Zhang, T., Chen, Q., and Lin, C. (2007). Optimal sensor placement for airborne containment detection in an aircraft cabin. *HVAC&R Research*, 13(5), 683-696.

- [21] Chou, S., Overfelt, R., Gale, W., Gale, H., Shannon, C., Buschle-Diller, G., and Watson, J. (August 2009). Effects of Hydrogen Peroxide on Common Aviation Textiles. *Civil Aerospace Medical Institute*, Report No. DOT/FAA/AM-09/16.
- [22] Gale, W., Sofyan, N., Gale, H., Sk, M., Chou, S., Fergus, J., and Shannon, C. (January 2009). Effect of vapour phase hydrogen peroxide, as a decontaminant for civilian aviation applications, on microstructure, tensile properties and corrosion resistance of 2024 and 7075 age hardenable aluminum alloys and 304 austenitic stainless steel. *Material Science and Technology*, 25(1), 76-84.
- [23] Rickloff, J. (November 2008). Factors Influencing Hydrogen Peroxide Gas Sterilant Efficacy. Advanced Barrier Concepts, Inc. 2008, 1-4. Retrieved December 7, 2010 from http://isolationinfo.com/docs/Sterilant%20Efficacy%20Factors%20JRR.pdf?org=/tech_val_topics.asp
- [24] Sk, M. H., Overfelt, R. A., Haney, R. L., and Fergus, J.W. (2010). Hydrogen Embrittlement of 4340 Steel due to Condensation during Vaporized Hydrogen Peroxide Treatment. *Materials Science and Engineering A*, doi:10.1016/j.msea.2011.01.100
- [25] Adams, D., Brown, G. P., Fritz, C., and Todd, T. R. (1998). Calibration of a near-infrared (NIR) H₂O₂ vapor monitor. *Pharmaceutical Engineering*, 18(4), 66-85.
- [26] Pandrangi, L., and Morrison, G. (February 2008). Ozone interactions with human hair: Ozone uptake rates and product formation, *Atmospheric Environment*, 42, 5079-5089.
- [27] Board on Environmental Studies, National Research Council. (2002). The Airliner Cabin Environment and the Health of Passengers and Crew. National Academy Press, Washington D.C., USA. Retrieved December 8, 2010 from <http://www.nap.edu/openbook.php?isbn=0309082897>
- [28] Winder, C. (2006). Air monitoring studies for aircraft contamination. *Current Topics in Toxicology*, 3, 33-48.
- [29] Mackerer, C., Barth, M., Krueger, A., Chawala, B., and Roy, T. (1999). Comparison of Neurotoxic Effects and Potential Risks from Oral Administrations or Ingestion of Tricresyl Phosphate and Jet Engine Oil Containing Tricresyl Phosphate. *Journal of Toxicology and Environmental Health, Part A*, 56, 293-328.
- [30] van Netten, C., and Lueng, V. (2000). Comparison of the Constituents of Two Jet Engine Lubricating Oils and Their Volatile Pyrolytic Degradation Products. *Applied Occupational and Environmental Hygiene*, 15(3), 277-283.
- [31] Nakamoto, K. (2009). Infrared and Raman Spectra of Inorganic and Coordination Compounds, Part A: Theory and Applications in Inorganic Chemistry 6th ed. Wiley Publishing, Hoboken, NJ, USA.

- [32] Nyquist, R. (2001) *Interpreting Infrared, Raman, and Nuclear Magnetic Resonance Spectra: Volume 1 – Variables in Data Interpretation of Infrared and Raman Spectra*, Academic Press, San Diego, Ca, USA.
- [33] Grobelnik, B. (March 2006). Seminar II: Infrared Spectroscopy. University of Ljubljana.
- [34] Barrow, G. (1963). *The Structure of Molecules – An Introduction to Molecular Spectroscopy*. W. A. Benjamin, Inc., New York, NY, USA.
- [35] PerkinElmer, Inc. (December 2000). *Spectrum GX User's Guide: Release D*.
- [36] Califano, S. (1976). *Vibrational States*. Wiley-Interscience Publishing, New York, USA.
- [37] Etchegoin, P. G., Meyer, M., Blackie, E., and Le Ru, E. C. (November 2007). Statistics of Single-Molecule Surface Enhanced Raman Scattering Signals: Fluctuation Analysis with Multiple Analyte Technique. *Analytical Chemistry*, 79(21), 8411-8415.
- [38] Cortes, E., Etchegoin, P. G., Le Ru, E. C., Fainstein, A., Vela, M. E., and Salvarezza, R.C. (July 2010). Electrochemical Modulation for Signal Discrimination in Surface Enhanced Raman Scattering (SERS). *Analytical Chemistry*, 82(16), 6919-6925.
- [39] Mc Evoy, K. M., Genet, M. J., and Dupont-Gillain, C. C. (August 2008). Principal Component Analysis: A Versatile Method for Processing and Investigation of XPS Spectra. *Analytical Chemistry*, 80(19), 7226-7238.
- [40] Osten, D. W., and Kowalski, B.R. (May 1984). Multivariate curve resolution in liquid chromatography. *Analytical Chemistry*, 56(6), 991-995.
- [41] Vandeginste, B. G., Derks, W., and Kateman, G. (1985). Multicomponent self-modelling curve resolution in high-performance liquid chromatography by iterative target transformation analysis. *Analytica Chimica Acta*, 173, 253-264.
- [42] Vandeginste, B., Essers, R., Bosman, T., Reijnen, J., and Katemen, G. (May 1985). Three-component curve resolution in liquid chromatography with multiwavelength diode array detection. *Analytical Chemistry*, 57(6), 971-985.
- [43] Windig, W., and Guilment, J. (July 1991). Interactive self-modeling mixture analysis. *Analytical Chemistry*, 63(14), 1425-1432.
- [44] Banas, K., Banas, A., Moser, H. O., Bahou, M., Li, W., Yang, P., Cholewa, M., and Lim, S. K. (March 2010). Multivariate Analysis Techniques in the Forensics Investigation of the Postblast Residues by Means of Fourier Transform-Infrared Spectroscopy. *Analytical Chemistry*, 82(7), 3038-3044.

- [45] Nieuwoudt, H.H., Prior, B. A., Pretorius, I. S., Manley, M., and Bauer, F. F. (2004). Principal Component Analysis Applied to Fourier Transform Infrared Spectroscopy for the Design of Calibration Sets for Glycerol Prediction Models in Wine and for the Detection and Classification of Outlier Samples. *Journal of Agriculture and Food Chemistry*, 52(12), 3726-3735.
- [46] Bu, D., and Brown, C. W. (2000). Self-Modeling Mixture Analysis by Interactive Principal Component Analysis. *Applied Spectroscopy*, 54(8), 1214-1221.
- [47] Knorr, F. J., and Futrell, J. H. (July 1979). Separation of Mass Spectra of Mixtures by Factor Analysis. *Analytical Chemistry*, 51(8), 1236-1241.
- [48] Adams, M. J. (1995). Chemometrics in Analytical Spectroscopy. The Royal Society of Chemistry, Thomas Graham House, Science Park, Cambridge, U.K.
- [49] Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., Kaufman, L. (1988). Chemometrics: a textbook (Data handling in science and technology; no. 2). Elsevier Science Publishing Company Inc. New York, NY, U.S.A.
- [50] Huckaba, C. E., and Keyes, F. G. (April 1948). The Accuracy of Estimation of Hydrogen Peroxide by Potassium Permanganate Titration. *Journal of American Chemical Society*, 70(4), 1640-1644.
- [51] Marco, J., Orza, J. M., and Abboud, J. -L. M. (1994). Fourier Transform infrared study of gas phase H-bonding: absorptivities and formation equilibrium constants of fluoroalcohol complexes. *Vibrational Spectroscopy*, 6, 267-283.
- [52] Bodis, J., Kornatowski, J., and Lercher, J. A. (2006). FT-IR Spectroscopy Study of Interactions Between Methanol and MeAPO-5 Single Crystals. *Seria F Chemica*, 9, 29-38.
- [53] Dixon, J. R., George, W. O., Hossain, M. F., Lewis, R., and Price, J. M. (January 1997). Hydrogen-bonded forms of methanol. *Journal Chemical Society, Faraday Transactions*, 93 (20), 3611-3618.
- [54] Keefe, C. D., Gillis, E. A., and MacDonald, L. (2009). Improper Hydrogen-Bonding CH*Y Interactions in Binary Methanol Systems as Studied by FTIR and Raman Spectroscopy. *Journal of Physical Chemistry A*, 113, 2544-2550.
- [55] Joseph, J., and Jemmis, E. D. (2007). Red-, Blue-, or No-Shift in Hydrogen Bonds: A Unified Explanation. *Journal of American Chemical Society*, 129, 4620-4632.

Appendix A: Principal Component Analysis MATLAB® Source Code

```
% data.txt is the location of the file with absorbance values for each wavenumber and sample
fid=fopen('data.txt');
% [fileData_x fileData_y] = [wavenumbers+1 samples]
% extract data from data.txt file into a fileData matrix
[fileData]=fscanf(fid,'%f %f',[1764 17]);
X=fileData';
fclose(fid);

% data matrix [X]
[n,p]=size(X);
% n-1=num of samples
% p=num of wavenumbers
Xdata=zeros(n,p-1);
y=zeros(n,1);
for i=1:n
    for j=1:p-1
        Xdata(i,j)=X(i,j+1);
    end
end
for i=1:n
    y(i,1)=X(i,1);
end

[n,p]=size(Xdata)
Xdata_mean=mean(Xdata);
Xdata_meanAdj=Xdata-repmat(Xdata_mean,[n 1]);
y_mean=mean(y);
y_meanAdj=y-repmat(y_mean,[n 1]);
eigValues=flipud(eig(Xdata_meanAdj'*Xdata_meanAdj));
cumVar_explained=cumsum(eigValues./sum(eigValues));
num_sig_eigValues=0;

% determine number of significant eigValues until cumulative variance explained >= XX%
i=1;
while cumVar_explained(i)<0.95
    num_sig_eigValues=num_sig_eigValues+1;
    i=i+1;
end
num_sig_eigValues=6
```

```

% [loadings] loadings matrix
% [scores] score matrix
% columns are in order of decreasing component variance
% princomp automatically subtracts of column means (use raw data)
[loadings,scores,latent]=princomp(Xdata);
% reduced loadings & scores based on num_sig_eigValues
reduced_loadings=zeros(p,num_sig_eigValues);
reduced_scores=zeros(n,num_sig_eigValues);
for i=1:p
    for j=1:num_sig_eigValues
        reduced_loadings(i,j)=loadings(i,j);
    end
end
for i=1:n
    for j=1:num_sig_eigValues
        reduced_scores(i,j)=scores(i,j);
    end
end

% ***** PCR VARIABLES *****
% vector of estimates of the regression coefficients
% [b] is product of eigenvectors [loadings] and y-loadings [q]
% y-loadings [q] determined by regression of [y] on [scores]
% [D] is diagonal matrix with each diagonal element equal to 1/tk
% tk is eigenvalue of factor k
% [q] = [D][scores]'[y]
D=zeros(num_sig_eigValues,num_sig_eigValues);
for i=1:num_sig_eigValues
    for j=1:num_sig_eigValues
        if(i==j)
            D(i,j)=1/eigValues(i);
        end
    end
end
q=D*reduced_scores*y;
b=reduced_loadings*q;
a=y_mean-Xdata_mean*b;

% output key program variables to .txt files for further analysis in Excel
dlmwrite('output1.txt',X,'delimiter','\t','precision',4);
dlmwrite('output2.txt',reduced_loadings,'delimiter','\t','precision',4);
dlmwrite('output4.txt',latent,'delimiter','\t','precision',4);
dlmwrite('output5.txt',q,'delimiter','\t','precision',4);
dlmwrite('output6.txt',b,'delimiter','\t','precision',4);
dlmwrite('output7.txt',a,'delimiter','\t','precision',4);

```