

**DEVELOPMENT OF *IN SITU* MONITORING AND DATA-DRIVEN MODELING  
FOR COMPLEX SYSTEMS: CASE STUDY ON SIMULANT MIXTURES OF  
NUCLEAR WASTE**

A Dissertation  
Presented to  
The Academic Faculty

By

Stefani Kocevskaja

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy in the  
School of Chemical and Biomolecular Engineering

Georgia Institute of Technology

May 2022

© Stefani Kocevskaja 2022

**DEVELOPMENT OF *IN SITU* MONITORING AND DATA-DRIVEN MODELING  
FOR COMPLEX SYSTEMS: CASE STUDY ON SIMULANT MIXTURES OF  
NUCLEAR WASTE**

Thesis committee:

Dr. Martha A. Grover, Advisor  
Chemical and Biomolecular Engineering  
*Georgia Institute of Technology*

Dr. Michael A. Filler  
Chemical and Biomolecular Engineering  
*Georgia Institute of Technology*

Dr. Ronald W. Rousseau, Advisor  
Chemical and Biomolecular Engineering  
*Georgia Institute of Technology*

Dr. Anna S. Erickson  
Department of Mechanical Engineering  
*Georgia Institute of Technology*

Dr. Blair K. Brettmann  
Chemical and Biomolecular Engineering  
*Georgia Institute of Technology*

Date approved: April 1, 2022

For my late grandparents Milanka and Stoimen

## ACKNOWLEDGMENTS

First and foremost, I would like to thank my parents, Ljiljana and Branko Kocevski for leaving everything behind to provide a better life for myself and my brother. Thank you to my brother for teaching me perseverance and always being there for me.

Next, I would like to acknowledge my PhD advisors, Martha Grover and Ronald Rousseau. Thank you for continuing to challenge me since the start, you have developed my scientific and leadership skills immensely! Martha, thank you for all the questions that made me think more deeply about my research. Thank you for your leadership, both in the group and in the department. You've been a great mentor and teacher. Ron, thank you for sharing your vast knowledge in chemical engineering, especially crystallization and polymorphism. Thank you for your scientific curiosity, which inspired me to keep going during hard times. Thank you for your honest feedback, and for teaching me the importance of writing a good abstract and conclusions paragraph. Finally, I would like to extend my gratitude to you both for taking me on and allowing me to collaborate with so many excellent researchers. Over the past five years, you gave me an opportunity to learn and grow and I will be forever grateful for your patience and encouragement.

To my committee members, Blair Brettmann, Michael Filler and Anna Erickson, thank you for bringing in different perspectives and ideas. Your feedback has helped me better define my research work and I appreciate your involvement in my growth.

To our collaborators, Mike Stone and Dan Lambert, thank you for always being available to answer my questions on processes at Hanford and SRNL. Your involvement has been instrumental during my PhD journey: I appreciate all the discussions, presentations and email exchanges.

Next, to everyone that shared an office with me, thank you for being amazing coworkers, I will forever miss my time in ES&T 2306. Matt McDonald and Youngjo Kim, you were there from the start and made me feel welcome. I always looked up to you both.

John Maggioni - our deep conversations over tea are deeply missed. You've been a great influence during my PhD, thank you for introducing me to Python and LaTeX and enabling me to learn many new research and coding skills. Steven Crouse and Rupanjali Prasad, I've loved welcoming you on the nuclear waste project. You have made the past year a fun experience and I look forward to taking more trips together. I'm confident you will make great contributions with your research work. To the rest of the Grover group members, old and new, including Chiamaka Obianyor (thank you for being so inspiring and encouraging), Aaron Liu (my birthday twin), Marcellus de Moraes, Hossein Salami, and many others, thank you for your friendship and support over the years. To my undergraduates, Anugraha Babuji and Tyra Jones, thank you for your dedication and inquisitive minds, I wish you the best!

I would also like to thank my friends, both near and far, for spreading your joy. I really appreciate all of you! Haley White, Maggie Manspeaker, Annette Mierzwa, Jeff Guerrero, Dara Kim, Ashley Kim, thank you making me laugh.

And finally, thank you Matt Rivera, for all the love, jokes, adventures, concerts and homemade desserts, I've loved every moment!! Thank you for always being there for me, to share both the good and the bad moments. I'm so glad you joined our first-year ChBE friend group at Georgia Tech in the Fall of 2017 - it has made all the difference. Its been such a wonderful adventure to share my PhD journey with you, BooBoo and Larry.

***Financial acknowledgments***

This thesis is based upon work supported under an Integrated University Program Graduate Fellowship. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Department of Energy Office of Nuclear Energy. Support by the U.S. Department of Energy under Cooperative Agreement DE-FC01-06EW07053, entitled “The Consortium for Risk Evaluation with Stakeholder Participation III,” is also gratefully acknowledged.

## TABLE OF CONTENTS

<b>Acknowledgments</b> . . . . .	iv
<b>List of Tables</b> . . . . .	xi
<b>List of Figures</b> . . . . .	xiv
<b>List of Acronyms and Symbols</b> . . . . .	xxi
<b>Summary</b> . . . . .	xxii
<b>Chapter 1: Introduction and Background</b> . . . . .	1
1.1 Legacy Nuclear Waste in the United States . . . . .	2
1.2 Direct Feed Low Activity Waste (DFLAW) Processing . . . . .	3
1.3 Real-Time In-Line Monitoring (RTIM) at Hanford . . . . .	4
<b>Chapter 2: Multivariate Regression Models for <i>In-Situ</i> Monitoring of Simplified Liquid Low-Activity Waste Simulants<sup>1</sup></b> . . . . .	7
2.1 Introduction . . . . .	7
2.2 Methods . . . . .	8
2.2.1 Spectroscopic Methods . . . . .	8
2.2.2 Chemometric Methods . . . . .	9
2.2.3 Experimental Design . . . . .	12

2.3	Results and Discussion . . . . .	15
2.3.1	ATR-FTIR Data - Closer Look . . . . .	17
2.3.2	Raman Data - Closer Look . . . . .	18
2.4	Conclusions . . . . .	20
<b>Chapter 3: Blind Source Separation – Partial Least Squares Regression Model for Quantification of Target Species in Liquid Low-Activity Waste Simulants<sup>2</sup> . . . . .</b>		<b>22</b>
3.1	Introduction . . . . .	22
3.2	Methods . . . . .	24
3.2.1	Experimental . . . . .	24
3.2.2	Modeling . . . . .	24
3.3	Results and Discussion . . . . .	32
3.3.1	Simulated Data . . . . .	33
3.3.2	Experimental Data . . . . .	39
3.4	Conclusions . . . . .	54
<b>Chapter 4: Effect of Ion Interactions on the Raman Spectrum of Nitrate: To- ward Monitoring of Low-Activity Nuclear Waste at Hanford<sup>3</sup> . . . . .</b>		<b>57</b>
4.1	Introduction . . . . .	57
4.2	Ionic Interactions in Aqueous Solutions . . . . .	58
4.2.1	Ion Association Theory . . . . .	58
4.2.2	Overview of Ion Pairing in Nitrate Solutions . . . . .	59
4.2.3	Peak Fitting . . . . .	60
4.3	Materials and Methods . . . . .	63

4.4	Experimental Results . . . . .	64
4.4.1	Subpeaks Separation . . . . .	65
4.4.2	Nitrate Peak Analysis . . . . .	66
4.5	Conclusions . . . . .	77
<b>Chapter 5: Expansion of <i>In-Situ</i> Monitoring and Data-Driven Modeling Across Multiple DFLAW Sampling Locations<sup>4</sup></b> . . . . .		79
5.1	Introduction . . . . .	79
5.2	Methods . . . . .	80
5.2.1	In-line spectroscopy . . . . .	80
5.2.2	Blind source separation – partial least squares regression (BSS – PLSR) modeling . . . . .	81
5.3	Results and Discussion . . . . .	82
5.3.1	Low activity waste (LAW) feed simulant . . . . .	82
5.3.2	Concentrate receipt vessel (CRV) simulant . . . . .	86
5.3.3	Effluent management facility (EMF) simulants . . . . .	87
5.4	Conclusions . . . . .	88
<b>Chapter 6: IR Spectroscopy Monitoring coupled with Machine Learning for Defense Waste Processing Facility (DWPF) Applications at the Savannah River Site<sup>5</sup></b> . . . . .		90
6.1	Introduction . . . . .	90
6.2	Materials and Methods . . . . .	91
6.2.1	Overview of the waste . . . . .	91
6.2.2	Design of Experiments . . . . .	93
6.3	Results and Discussion . . . . .	94

6.4	Conclusions . . . . .	99
<b>Chapter 7: Conclusions and Future Directions . . . . .</b>		<b>102</b>
7.1	Conclusions . . . . .	102
7.2	Future Directions . . . . .	107
7.2.1	Future work with Blind Source Separation (BSS) . . . . .	108
7.2.2	Understanding the nonlinear phenomena in the spectra and its im- plication on waste quantification . . . . .	109
7.2.3	Using ATR-FTIR to elucidate the chemistry during Defense Waste Processing at SRS . . . . .	109
<b>Appendices . . . . .</b>		<b>112</b>
Appendix A: Prediction errors for BSS-PLSR . . . . .		113
Appendix B: Experimental data for the nitrate shift system . . . . .		123
<b>References . . . . .</b>		<b>138</b>

## LIST OF TABLES

2.1	Experimental design for the training set. The molal concentrations are calculated as the moles of the target species per kilogram of solvent (water). . .	14
2.2	Experimental design for the test set. The molal concentrations are calculated as the moles of the target species per kilogram of solvent (water). . . .	15
3.1	Experimental design for the test set. Each mixture contains the four target species and a different combination of the non-target species. . . . .	28
4.1	Table of parameters of the Gaussian functions used to deconvolve the nitrate peak, from Yu and co-workers [52]. . . . .	62
4.2	Table of salt combinations explored in the experiments. . . . .	64
4.3	Table of fitted subpeak areas for spectra in Figure 4.3, using the Gaussian model from Section 4.2.3. The subpeak areas have been normalized by the total area. . . . .	67
5.1	Target species concentrations in mol/L, quantified with ion chromatography (IC) at SRNL. . . . .	83
6.1	Tank 40 sludge waste nonradioactive simulant composition, reported by Woodham and co-workers [75]. The wt% total solids (TS) represents the solids that do not dissolve after heating to 110°C, while the wt% insoluble solids (IS) represents the difference between the measured total solids and dissolved solids. . . . .	93
6.2	Ion composition in the nonradioactive simulated Tank 40-8 SRAT/SME product slurry, reported by Woodham and co-workers [75]. . . . .	94
6.3	Anion composition measured by IC for a nonradioactive lab run of SRAT/SME in the RC1 experimental set-up. The samples match the steps in Figure 6.2.	95

6.4	Training data set for the target anion species and solvent (water). . . . .	95
6.5	Table of the concentration predictions using PLSR, matching the sample numbers for the IC measurements from Table 6.3. . . . .	101
6.6	Table of the concentration predictions using BSS-PLSR, matching the sample numbers for the IC measurements from Table 6.3. . . . .	101
A.1	Absolute values of the percent errors for the simulated Raman data set using a 0th derivative in the PLSR model. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species. . . . .	114
A.2	Absolute values of the percent errors for the simulated Raman data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 7. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species. . . . .	115
A.3	Absolute values of the percent errors for the simulated Raman data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 27. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species. . . . .	116
A.4	Absolute values of the percent errors for the simulated IR data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 5. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species. . . . .	118
A.5	Absolute values of the percent errors for the experimental Raman data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 7. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species. . . . .	120

A.6	Absolute values of the percent errors for the experimental IR data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 5. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species. . . . .	122
B.1	Molal concentrations for Exp. 1 from Table 4.2. . . . .	124
B.2	Molal concentrations for Exp. 2 Table 4.2. . . . .	124
B.3	Molal concentrations for Exp. 3 from Table 4.2. . . . .	124
B.4	Molal concentrations for Exp. 4 from Table 4.2. . . . .	125
B.5	Molal concentrations for Exp. 5 from Table 4.2. . . . .	125
B.6	Molal concentrations for Exp. 6 from Table 4.2. . . . .	125
B.7	Molal concentrations for Exp. 7 from Table 4.2. . . . .	126
B.8	Molal concentrations for Exp. 8 from Table 4.2. . . . .	126
B.9	Molal concentrations for Exp. 9 from Table 4.2. . . . .	126
B.10	Molal concentrations for Exp. 10 from Table 4.2. . . . .	127
B.11	Molal concentrations for Exp. 11 from Table 4.2. . . . .	127
B.12	Molal concentrations for Exp. 12 from Table 4.2. . . . .	127
B.13	Molal concentrations for Exp. 13 from Table 4.2. . . . .	128
B.14	Molal concentrations for Exp. 14 from Table 4.2. . . . .	129
B.15	Molal concentrations for Exp. 15 from Table 4.2. . . . .	130

## LIST OF FIGURES

2.1	Infrared absorption and Raman scattering phenomena [19]. . . . .	8
2.2	Experimental set-up. . . . .	13
2.3	Reference spectra of the species in our model system: a) Raman spectroscopy; b) ATR-FTIR spectroscopy. The data show the characteristic peaks of each species. . . . .	15
2.4	Spectral measurements of the training data set mixtures for a) Raman spectroscopy and b) ATR-FTIR spectroscopy. Concentrations for each measurement are shown in Table 2.1. . . . .	16
2.5	Spectral measurements of the test data set mixtures for a) Raman spectroscopy and b) ATR-FTIR spectroscopy. Concentrations that match each measurement are shown in Table 2.2. . . . .	17
2.6	PCR, PLSR, ILSR and CLSR model predictions using a) Raman and b) ATR-FTIR data. The solid black line represents unity and the dashed black lines represents $\pm 20\%$ deviations, which are based on the uncertainty at the Hanford analytical laboratory [11]. . . . .	17
2.7	Updated ATR-FTIR model results (including water) showing predictions for a) the full range of concentrations and b) the analyte range. The solid black line represents unity and the dashed black lines represents $\pm 20\%$ deviations, which are based on the uncertainty at the Hanford analytical laboratory [11]. . . . .	18
2.8	PCR, PLSR, ILSR and CLSR model predictions on ATR-FTIR data for a) nitrate, b) nitrite, c) carbonate and d) sulfate. The solid black line represents unity and the dashed black lines represents $\pm 20\%$ deviations, which are based on the uncertainty at the Hanford analytical laboratory [11]. . . . .	19

2.9	PCR, PLSR, ILSR and CLSR model predictions on Raman data for a) nitrate, b) nitrite, c) carbonate and d) sulfate. The solid black line represents unity and the dashed black lines represents $\pm 20\%$ deviations, which are based on the uncertainty at the Hanford analytical laboratory [11]. . . . .	20
2.10	Updated Raman model results (including water) showing predictions for a) the full range of concentrations using five principal components in the PLSR and PCR models and b) the full range of concentrations using six principal components in the PLSR and PCR models. The solid black line represents unity and the dashed black lines represents $\pm 20\%$ deviations, which are based on the uncertainty at the Hanford analytical laboratory [11].	21
3.2	Reference spectra of the species in our model system: a) Raman spectroscopy; b) ATR-FTIR spectroscopy. The spectra show the extent of overlap among the target and non-target species. . . . .	33
3.3	Simulated test data set mixtures for a) Raman spectroscopy with $SNR_{avg} = 40$ dB and b) ATR-FTIR spectroscopy with $SNR_{avg} = 55$ dB. Concentrations for the 21 mixtures are given in Table 3.1. . . . .	35
3.4	Simulated training data set mixtures for a) Raman spectroscopy with $SNR_{avg} = 40$ dB and b) ATR-FTIR spectroscopy with $SNR_{avg} = 55$ dB. The concentrations that match each measurement are shown in Table 2.1. . . . .	36
3.5	Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last three rows show the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed). The BSS algorithm did not identify a unique source for water. . . . .	37
3.6	Removing the signals of the non-target species for Raman: a) original test set mixture data (blue), containing both target and non-target species, b) signals of the non-target species (yellow) that are subtracted to generate the preprocessed spectra (red) in c). The intensity of the yellow signals is lower relative to the spectra of the targets. . . . .	38
3.7	Parity plot comparing the Raman predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman: a) 0th derivative spectra; b) 1st derivative spectra using a Savitzky-Golay filter with a second order polynomial and window length size of 7; and c) 1st derivative spectra using a Savitzky-Golay filter with a second order polynomial and window length size of 27. . . . .	39

3.8	Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last three rows show the normalized reference spectra of the non-target species. The normalized IR sources identified by the blind source separation algorithm are plotted in black (dashed). . . . .	40
3.9	Removing the signals of the non-target species for IR: a) original mixture data (blue), containing both target and non-target species, b) signals of the non-target species (yellow) that are subtracted to generate the preprocessed spectra (red) in c). The intensity of the yellow signals is lower relative to the spectra of the targets. . . . .	41
3.10	Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR. . . . .	42
3.11	Spectral measurements of the test data set mixtures for a) Raman spectroscopy and b) ATR-FTIR spectroscopy. Concentrations that match each measurement are shown in Table 3.1. . . . .	42
3.12	Spectral measurements of the training data set mixtures for a) Raman spectroscopy and b) ATR-FTIR spectroscopy. Concentrations for each measurement are shown in Table 2.1. . . . .	43
3.13	Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last two rows show the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed). The BSS algorithm did not identify water as a unique source. . . . .	44
3.14	Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last three rows show the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed). The "nitrate blue shift" source from row six was reclassified as a target species. . . . .	45
3.15	Overlay of the original, subtracted and preprocessed spectra for Raman, when the nitrate shift source is classified as a non-target species: the yellow signal in the dashed region has been subtracted during BSS-preprocessing. . . . .	47

3.16	Parity plot comparing the Raman predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman experimental data, while the nitrate shift source was classified as a non-target species. . . . .	47
3.17	a) Removing the signals of the non-target species for Raman. b) Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman. . . . .	48
3.18	Parity plot comparing the Raman predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman experimental data, using 0th derivative spectra. . . . .	48
3.19	Reference spectra-sources correlation using $L_t$ in the data input for BSS: the first six rows show the normalized reference spectra of the target species, while the last row shows the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed). . . . .	50
3.20	a) Removing the signals of the non-target species for Raman, using $L_t$ in the data input for BSS. b) Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman, using $L_t$ in the data input for BSS. . . . .	51
3.21	Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last two rows show the normalized reference spectra of the non-target species. The normalized IR sources identified by the blind source separation algorithm are plotted in black. . . . .	52
3.22	a) Removing the signals of the non-target species for IR. b) Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR. . . . .	53
3.23	Parity plot comparing the IR predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR experimental data, using 0th derivative spectra. . . . .	54
3.24	Reference spectra-sources correlation using $L_t$ in the data input for BSS: the first five rows shows the normalized reference spectra of the target species, while the last two rows show the normalized reference spectra of the non-target species. The normalized IR sources identified by the blind source separation algorithm are plotted in black. . . . .	55

3.25	a) Removing the signals of the non-target species for IR, using $L_t$ in the data input for BSS. b) Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR, using $L_t$ in the data input for BSS. . . . .	56
4.1	(a) The reference Raman spectrum of $\text{NaNO}_3$ in water at $1 \text{ mol kg}^{-1}$ . The red label highlights the peak due to the symmetric stretch of the nitrate anion, $\nu_1$ , investigated in this work. (b) The blue shift of the $\nu_1$ nitrate peak in aqueous solutions with increasing sodium nitrate concentration (which equals to the ionic strength of the solution for this experiment). The blue dotted lines highlight the shift of the peak maximum from 1048 to 1052 $\text{cm}^{-1}$ . The spectra were normalized by dividing by the maximum intensity and plotted with an offset of 0.3 to highlight the peak shift. . . . .	59
4.2	The areas of the subpeaks indicate the type of ionic interactions prevailing in each unary system at (a) $0.7 \text{ m Na}^+$ and (b) $9.5 \text{ m Na}^+$ . The measurements (solid black lines) were fit using the Gaussian model from Section 4.2.3 (dashed and solid magenta lines). The errors were calculated using Equation 4.4. The prevailing subpeak changes from SIP to CIP as the concentration increases, corresponding with the formation of contact ion pairs in the presence of greater sodium nitrate concentrations. The CIA subpeak also becomes relevant at high concentration, indicating the presence of complex ion aggregates. . . . .	66
4.3	The experimentally measured $\nu_1$ nitrate peaks for (a) unary solution of sodium nitrate (Exp. 1), (b) binary solution of sodium nitrate and nitrite (Exp. 6), (c) binary solution of sodium nitrate and carbonate (Exp. 7) and (d) binary solution of sodium nitrate and sulfate (Exp. 5). The measurements (solid lines) were fit using the Gaussian model from Section 4.2.3 (dashed lines). The errors were calculated using Equation 4.4. . . . .	67
4.4	The blue shift of the $\nu_1$ peak in the presence of three different cations as a function of (a) ionic strength and (b) cation concentration. . . . .	69
4.5	The blue shift of the main peak of the nitrate anion as a function of ionic strength, $I$ , which is equal to cation and anion concentration in these experiments. . . . .	70
4.6	The relative fractions, $x_A$ , of the subpeak areas associated with each type of ionic interaction in Exps. 1 and 4. . . . .	70

4.7	The extent of blue shift in the binary systems plotted against (a) the ionic strength, $I$ , (b) the concentration of anions, $C^-$ , and (c) the concentration of the cation, $C^+$ . The sodium nitrate system is included in the comparison, as a reference. . . . .	71
4.8	The relative areas of the subpeaks, $x_A$ , indicate the type of ionic interactions prevailing in each binary system. . . . .	72
4.9	Subtraction of the carbonate contribution using blind source separation [33].	73
4.10	Change in (left) total area of the $\nu_1$ peak of nitrate and (right) maximum height of the nitrate peak as a function of the cation concentration. . . . .	74
4.11	Change in (left) total area of the $\nu_1$ peak of nitrate and (right) maximum height of the nitrate peak as a function of the cation concentration. The values were scaled by the concentration of nitrate at each point, to account for the nitrate salt additions to the solution. . . . .	74
4.12	The extent of blue shift in the ternary systems plotted against (a) the ionic strength, $I$ , (b) the concentration of anions, $C^-$ , and (c) the concentration of the cation, $C^+$ . The sodium nitrate system is included in the comparison, as a reference. . . . .	75
4.13	The relative areas of the subpeaks, $x_A$ , indicate the type of ionic interactions prevailing in each ternary system. . . . .	75
4.14	The extent of blue shift in the simulants, including LAW train ( $\text{NaNO}_3$ , $\text{Na}_2\text{SO}_4$ , $\text{NaNO}_2$ and $\text{Na}_2\text{CO}_3$ ), LAW test ( $\text{NaNO}_3$ , $\text{Na}_2\text{SO}_4$ , $\text{NaNO}_2$ , $\text{Na}_2\text{CO}_3$ , $\text{Na}_3\text{PO}_4$ , $\text{NaCH}_3\text{COO}$ and $\text{Na}_2\text{C}_2\text{O}_4$ ) and LAW complete feed ( $\text{Na}^+$ , $\text{NO}_3^-$ , $\text{NO}_2^-$ , $\text{CO}_3^{2-}$ , $\text{OH}^-$ , $\text{Al}(\text{OH})_4^-$ , $\text{K}^+$ , $\text{SO}_4^{2-}$ , $\text{Cl}^-$ , $\text{PO}_4^{3-}$ , $\text{CrO}_4^{2-}$ , $\text{C}_2\text{O}_4^{2-}$ , $\text{F}^-$ and total organic carbon (TOC) species at $pH > 13$ ) plotted against (a) the ionic strength, $I$ , (b) the concentration of anions, $C^-$ , and (c) the concentration of the cation, $C^+$ . The sodium nitrate system is included in the comparison, as a reference. . . . .	77
4.15	The relative areas of the subpeaks, $x_A$ , indicate the type of ionic interactions prevailing in each LAW system. . . . .	77
B.1	Spectra for Exps. 1 – 3 in Table 4.2. . . . .	131
B.2	Spectra for Exps. 4 – 6 in Table 4.2. . . . .	132
B.3	Spectra for Exps. 7 – 9 in Table 4.2. . . . .	133

B.4	Spectra for Exps. 10 – 12 in Table 4.2. . . . .	134
B.5	Spectra for Exps. 13 – 15 in Table 4.2. . . . .	135
B.6	Model fit for a low and high concentration of sodium nitrate from Exp. 1 in Table 4.2. A comparison was made between a model with five and three subpeaks. . . . .	136
B.7	Model fit for a low and high concentration of sodium nitrate from Exp. 1 in Table 4.2. The effect of removing one subpeak at a time was evaluated. .	137

## **LIST OF ACRONYMS AND SYMBOLS**

**ATR-FTIR** Attenuated Total Reflectance-Fourier Transform Infrared

**BSS-PLSR** Blind Source Separation-Partial Least Squares Regression

**DFLAW** Direct Feed Low Activity Waste

**DOE-EM** Department of Energy's Office of Environmental Management

**DWPF** Defense Waste Processing Facility

**GAO** Government Accountability Office

**HLW** High Level Waste

**ICA** Independent Component Analysis

**LAW** Low Activity Waste

**LIBS** Laser Induced Breakdown Spectroscopy

**PAT** Process Analytical Technology

**PLSR** Partial Least-Squares Regression

**RTIM** Real-Time In-Line Monitoring

**SRNL** Savannah River National Laboratory

## SUMMARY

Approximately 55 million gallons of nuclear and chemical waste are stored at the Hanford site in the United States. The nuclear waste will be stabilized into a glass form via vitrification. One of the main challenges associated with Hanford waste treatment is the complexity of the waste, which calls for extensive sampling during processing.

In this thesis, the use of Process Analytical Technology (PAT), including *in-situ* infrared and Raman spectroscopy, is demonstrated for Real-Time In-Line Monitoring (RTIM) of simulated nuclear waste. *In-situ* monitoring will facilitate the continuous operation of the waste treatment facility and help reduce employee exposure to hazardous working conditions. However, in order to extract meaningful information from the spectroscopy measurements, the IR and/or Raman spectra need to be converted into species concentrations. The concentration values are used to evaluate the composition of the nuclear waste feed and determine the amount of glass-forming chemicals that should be added to vitrify the waste into a stable glass form.

The work in this thesis bridges the gap between spectroscopy data and concentration outputs. A flexible spectra-to-composition modeling framework is developed to address the varying complexity in the simulated waste mixtures. The initial work incorporates linear multivariate regression models, such as Partial Least-Squares Regression (PLSR), to quantify the concentrations of process-relevant (target) species in the waste system, which include nitrate, nitrite and sulfate. As the complexity of the waste system increases, which is represented by the presence of additional species in the waste system, advanced signal separation preprocessing techniques are incorporated in the modeling framework. The signal separation methods include blind source separation, such as Independent Component Analysis (ICA), which is used to preprocess the spectra and classify the contributions of the nuclear waste components into target (process-relevant species) and non-target species. In this scenario, which is applicable to low-activity liquid waste processing at Hanford, the

non-target subset represents species that are present in the waste, but whose concentration measurements may not be required for process operation and control. By implementing the blind source separation preprocessing step, the signal contributions of the non-target species are removed before the spectra are analyzed with PLSR. This allows for increased robustness of the PLSR model, which only requires mixtures of the target species as training data.

Another aspect of the thesis work is the analysis of non-linear phenomena occurring in Raman spectra, centered around the nitrate peak which exhibited peak shifting during certain conditions. The extent of the non-linear peak shifting behaviour is analyzed in the presence of tank waste species of varying composition.

Finally, while the work in this thesis is mainly focused on applications for liquid waste processing at the Hanford site, the BSS preprocessing and PLSR quantification framework (BSS-PLSR) is also used to analyze mixture data for applications at the Savannah River site, which holds approximately 35 million gallons of high-level radioactive waste. This thesis showcases the successful use of *in-situ* infrared spectroscopy, in the form of Attenuated Total Reflectance-Fourier Transform Infrared (ATR-FTIR), in conjunction with the BSS-PLSR framework to track the concentrations of target species during slurry processing at the Savannah River site.

Altogether, the work in this thesis is focused on enabling real-time monitoring of the chemical composition of nuclear waste. The work spans both sample measurements with *in-situ* spectroscopy and subsequent spectroscopy data-driven model development. The work enables robust and efficient quantification of target chemical species concentrations from multidimensional spectroscopy measurements.

# CHAPTER 1

## INTRODUCTION AND BACKGROUND

Advances in Process Analytical Technology (PAT) have enabled collection of data in real-time for all stages of process development [1]. The use of PAT can improve our understanding of a process during the early phases of development, which has been especially beneficial for the pharmaceutical industry. Other applications include process monitoring and verification that can facilitate real-time quality control, resulting in increased efficiency. Another important application of *in-situ* probe technology is collection of data in hard-to-access locations or hazardous environments, such as nuclear waste tanks. Compared to offline analysis, the online approach offers rapid collection of data with little to no sample preparation, which can be used to control the process, while shielding operators from hazards.

While PATs are capable of collecting rich data sets, useful for understanding and controlling a given process, the bottleneck lies in interpreting the collected information. The nature of the process dictates the number of PATs to be deployed, sometimes further increasing the dimensionality and thus complexity of the data.

In this thesis, the challenges associated with complex and multidimensional data are addressed by carefully designing training data sets of the target species, followed by application of dimensionality reduction and blind source separation techniques. The focus of the work is to enable RTIM for the processing of liquid low activity waste at the Hanford site in the State of Washington. In the remainder of Chapter 1, the various aspects of nuclear waste remediation are discussed, with a focus on the Hanford site. In Chapter 2, we introduce the system of target species (process-relevant species whose concentrations need to be quantified to enable waste processing) and analyze the data with multivariate linear models. In Chapter 3, the complexity of the system is increased by introducing non-target

species to the simulated waste mixtures. These species are present in the real low-activity waste at Hanford, but are not used for decision making during waste processing. A novel Blind Source Separation-Partial Least Squares Regression (BSS-PLSR) framework is developed to address the presence of non-target species, whose peaks may overlap with those of the targets. The BSS-PLSR framework consists of two stages; first, the contributions of the non-target species are identified and isolated using blind source separation techniques, second, the concentrations of the target species are quantified using a partial least-squares regression model. Our work from Chapter 3 revealed non-linear phenomena associated with the Raman nitrate peak, which are further studied in Chapter 4. To understand the phenomena, we measured the extent of nonlinear behaviour (in the form of a blue shift) in the presence of various anions and cations relevant for nuclear waste processing. In Chapter 5, the *in-situ* probes are used to measure simulated low-activity waste at different DFLAW locations. Moving across the DFLAW flowsheet introduces additional complexities, such as the presence of solids. In Chapter 6, we switch nuclear waste sites and focus on the Defense Waste Processing Facility (DWPF) at Savannah River National Laboratory (SRNL). The work is focused on the potential deployment of an *in-situ* IR probe for measurement of anion species in the highly radioactive slurry waste. Finally, Chapter 7 provides conclusions on the thesis work and offers a perspective on the future of *in-situ* spectroscopy technology and chemometric modeling in the nuclear waste industry.

## **1.1 Legacy Nuclear Waste in the United States**

The Department of Energy's Office of Environmental Management (DOE-EM) has been tasked with the clean-up of nuclear waste across 107 sites in the United States. Significant progress has been made, including the completion of clean-up at 91 sites. However, according to a 2019 U.S. Government Accountability Office (GAO) report [2], the remaining sites are some of the most challenging to clean-up. Out of the 90 million gallons of radioactive and hazardous waste, approximately 55 million gallons are stored at the Hanford

Site in Washington State. The waste originated from the production of plutonium as part of the Manhattan Project. Throughout the years, the waste has been reprocessed, leading to high complexity in chemical and physical composition. There are a total of 177 single and double-shell waste tanks. Within the tanks, the properties of the waste have been altered further via other activities, such as chemical processing and evaporation. This has resulted in non-homogeneous waste categorized in three types: supernatant liquid, saltcake solids, and sludge. The supernatant liquid and the dissolvable saltcake have been classified as Low Activity Waste (LAW) that contains small amounts of radionuclides, while the water-insoluble sludge is classified as High Level Waste (HLW). Under the current timeline, vitrification of the low-activity waste will be implemented first. Remediation efforts are expected to continue for another 50 years. The goal is to immobilize all waste in a borosilicate glass form, which will be cast in stainless steel canisters [3, 4, 5, 6].

## **1.2 Direct Feed Low Activity Waste (DFLAW) Processing**

The majority of the waste at Hanford falls under the LAW classification. The DOE has initiated the Direct Feed Low Activity Waste (DFLAW) project to expedite clean-up of low activity waste. DFLAW aims to directly vitrify supernatant low activity waste, as shown in Figure 1.1 [7]. The initial steps include feeding the waste into the Tank Side Cesium Removal (TSCR) pretreatment facility to reduce Cs-137 levels and remove insoluble solids. The preprocessed waste will be sent to the LAW vitrification facility. The waste will be mixed with Glass Forming Chemicals (GFCs) to create a melter feed slurry, which will be fed to a melter that operates at high temperatures to convert the feed to glass. The overall glass formulation depends upon the composition of the waste. Secondary waste will be treated in the Effluent Management Facility (EMF), and the concentrate will be recycled.

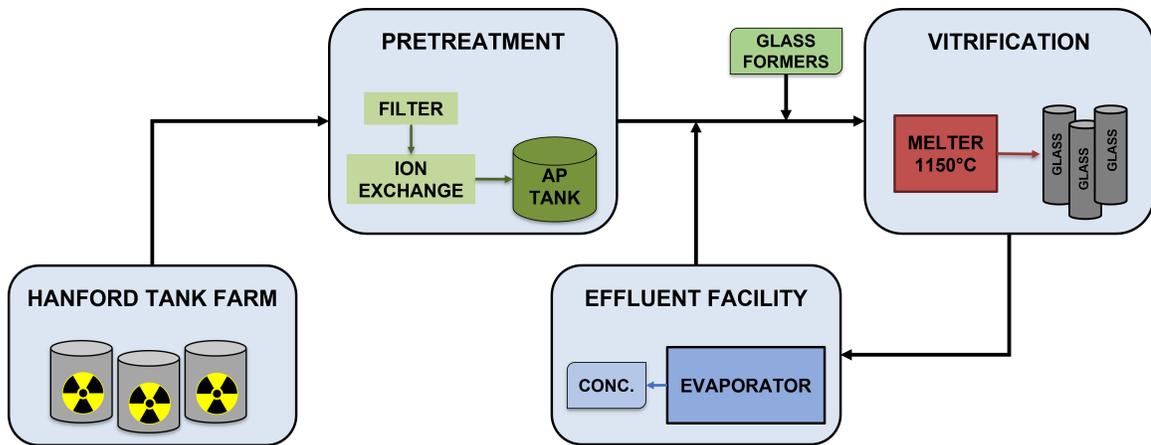


Figure 1.1: Simplified flowsheet for Direct Feed Low Activity Waste (DFLAW) processing [7].

### 1.3 Real-Time In-Line Monitoring (RTIM) at Hanford

The current Hanford mission is to process, immobilize and dispose of waste. The complexity of the waste and its hazardous nature call for extensive sampling at different locations across DFLAW to facilitate processing. It has been projected that thousands of samples per year are needed to maintain environmental compliance, product quality, and operating efficiency [8, 7]. In order to optimize the current sampling approach, the RTIM initiative seeks to evaluate if any of the existing offline analytical techniques can be replaced by online monitoring. *In-situ* monitoring can help minimize sample turn-around times, compared to offline analysis of grab samples. Since the sampling would be performed at or near the tank waste, the amount of secondary waste generated in the analytical laboratories will be reduced. Online monitoring will also help reduce employee exposure to radiation and other hazards associated with collecting samples for offline analysis. Finally, online analysis can provide data that can be used for online quality control and process control at different sampling locations [9, 10, 11].

As mentioned previously, the low-activity waste at Hanford will be stabilized into glass under the DFLAW process. The Best Basis Inventory Maintenance (BBIM) tool is an on-line resource that estimates and tracks chemical and radionuclide components from the

Hanford tank farm [12, 13]. According to the BBIM, the supernatant waste from the AP tank farm, which will be processed under DFLAW, is primarily made up of nonradioactive species, including H<sub>2</sub>O, and various cations and anions, such as Na<sup>+</sup>, NO<sub>3</sub><sup>-</sup>, NO<sub>2</sub><sup>-</sup>, CO<sub>3</sub><sup>2-</sup>, OH<sup>-</sup>, Al(OH)<sub>4</sub><sup>-</sup>, K<sup>+</sup>, SO<sub>4</sub><sup>2-</sup>, Cl<sup>-</sup>, PO<sub>4</sub><sup>3-</sup>, CrO<sub>4</sub><sup>2-</sup>, C<sub>2</sub>O<sub>4</sub><sup>2-</sup>, F<sup>-</sup> and additional total organic carbon (TOC) species (citrate, formate, acetate and ethylenediaminetetraacetic acid or EDTA). It is important to note that while the waste consists of many species, a few stand out as the most abundant ones, as shown in Figure 1.2. These species are currently analyzed offline using Ion Chromatography (IC) methods [14]. Having the capability to measure anion concentrations online using spectroscopic techniques would help optimize processing under DFLAW.

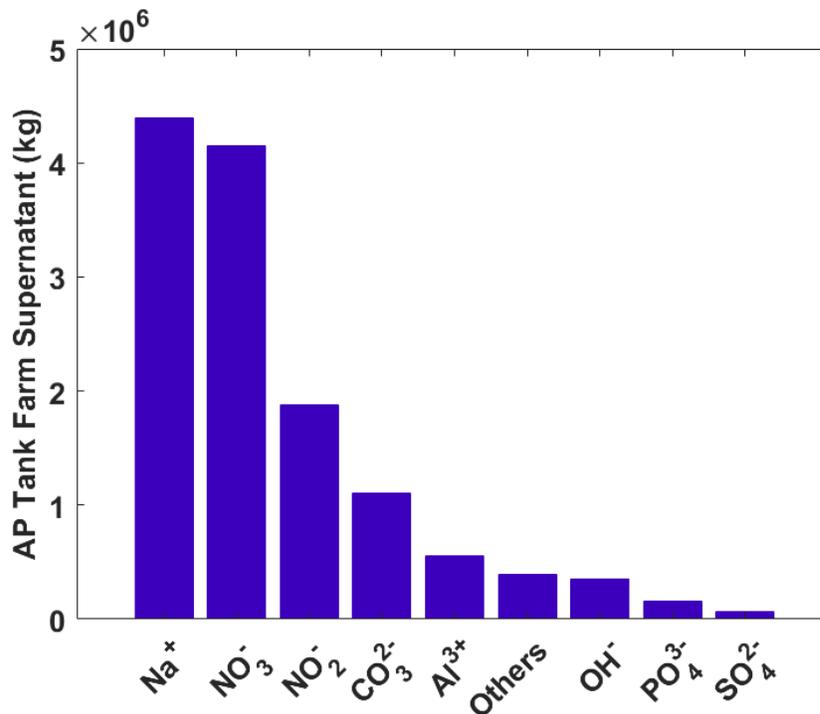


Figure 1.2: LAW Supernatant Composition by the Best Basis Maintenance Inventory (BBIM) tool [12, 13].

The goal of the Hanford RTIM program is to select and evaluate *in-situ* instrumentation for monitoring of low-activity waste. Selection criteria include cost of development, operation and maintenance and ability to measure analytes of interest within the ±20%

uncertainty level for the Hanford analytical laboratory [11]. The waste contains elemental, molecular and radioactive species, thus a variety of *in-situ* instruments have been proposed for use in DFLAW, including Raman spectroscopy, ATR-FTIR spectroscopy and Laser Induced Breakdown Spectroscopy (LIBS). Raman and ATR-FTIR have been prioritized over ion chromatography for molecular analysis, since they do not require any sample preparation or consumables, while LIBS is being evaluated for elemental analysis [11].

## CHAPTER 2

### MULTIVARIATE REGRESSION MODELS FOR *IN-SITU* MONITORING OF SIMPLIFIED LIQUID LOW-ACTIVITY WASTE SIMULANTS<sup>1</sup>

#### 2.1 Introduction

The liquid waste at Hanford, which is largely made up of water and sodium salts, such as nitrate, nitrite, carbonate, sulfate, aluminate, hydroxide, phosphate, oxalate, chromate, acetate, fluoride and other minor species, will be processed via vitrification under the direct-feed low-activity waste (DFLAW) procedure [11]. *In-situ* monitoring techniques, such as Raman and infrared (IR) spectroscopy can provide real-time information on the mixtures, which is required for continuous operation of the melter during DFLAW. Online monitoring tools can reduce the need for sample collection for off-line analysis, decreasing employee exposure to hazardous working conditions and long analysis periods.

Despite the presence of tens of molecular species in the liquid low-activity waste, a few stand out as the process-relevant species whose concentrations are important for the formulation of stable glass. The key spectroscopically active species that need to be monitored to facilitate DFLAW processing include nitrate and nitrite due to their prevalence in low-activity waste, and sulfate since it can interfere with the vitrification process by accumulating on the melt surface and corroding the melter components [11, 16]. In order to evaluate the performance of *in-situ* spectroscopy probes for liquid low-activity waste at Hanford, a system consisting of the process-relevant molecular species is studied in this chapter. The focus on this chapter is both on the instruments, comparing the measurements of Raman and ATR-FTIR spectroscopy probes, and the model development.

---

<sup>1</sup>Adapted from "S. Kocavska, R. W. Rousseau, M. A. Grover and G. M. Maggioni, "Multivariate Linear Models for In Situ Monitoring of Low-Activity Nuclear Waste Using Infrared and Raman Spectroscopy," in Proceedings of the Waste Management Symposia, 2020." [15]

## 2.2 Methods

### 2.2.1 Spectroscopic Methods

Spectroscopic methods are based on interactions between electromagnetic radiation and matter. Throughout this work, we use two spectroscopic methods, infrared (in the form of Attenuated Total Reflectance-Fourier Transform Infrared or ATR-FTIR) and Raman spectroscopy. Absorption spectroscopy, such as ATR-FTIR, occurs when a molecule absorbs light energy [17, 18]. A molecule is IR-active if there is a change in the dipole moment during molecular vibrations, which generates a field that can interact with the electric field from light radiation. Energy is absorbed if the natural frequency of the vibration matches the frequency of the radiation.

Scattering spectroscopy, such as Raman spectroscopy (inelastic scattering), occurs when the light radiation is scattered by matter. In Raman spectroscopy, the light radiation exchanges energy with the molecule, which results in a shift in the scattered radiation, known as the Raman effect [18]. An overview of the spectroscopic methods is shown in Figure 2.1.

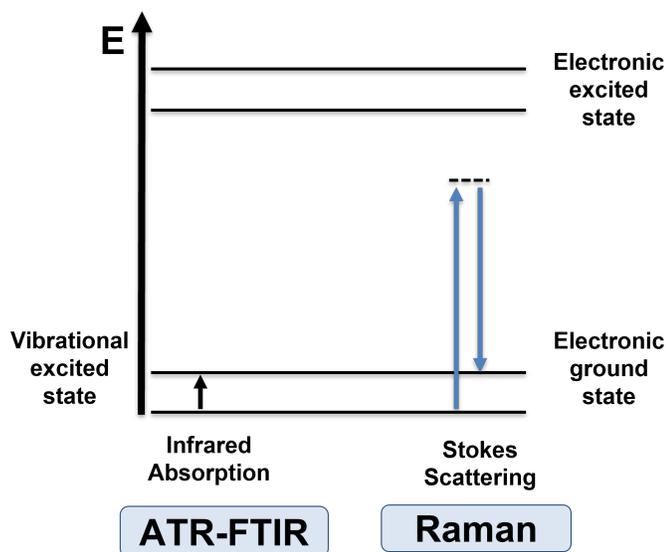


Figure 2.1: Infrared absorption and Raman scattering phenomena [19].

Raman-active molecules exhibit changes in polarizability. In the case when molecular bonds are electrically symmetrical, they do not absorb infrared energy and are thus IR-inactive, but can still be detected using Raman spectroscopy. Similarly, molecules that are weakly polarizable may be Raman-inactive, but IR-active. Due to the different measurement mechanisms, both IR and Raman were used to probe the multicomponent systems in this work. These techniques can also be used for verification and validation purposes if the species of interest are both IR and Raman active.

The intensity of the measured light radiation in both ATR-FTIR and Raman spectroscopy can be related to the concentration of the measured species. Assuming a linear dependence, the intensity can be expressed as a function of the concentration using the Beer-Lambert law:

$$\mathbf{X} = \mathbf{C}\mathbf{L} \quad (2.1)$$

where  $\mathbf{X} \in \mathbb{R}^{n_N \times n_L}$  represents the measured spectra,  $\mathbf{C} \in \mathbb{R}^{n_N \times n_K}$  the concentrations and  $\mathbf{L} \in \mathbb{R}^{n_K \times n_L}$  the pure component contributions. Note that  $n_N$  is the number of measurements,  $n_L$  the number of wavenumbers and  $n_K$  the total number of species.

### 2.2.2 Chemometric Methods

We selected four common chemometrics techniques to study the potential of ATR-FTIR and Raman spectra in quantifying the composition of key analytes in nuclear waste [20]. The first two methods are variations of a multiple-least squares regression (MLR), where the sum-squared error between the measured and reconstructed signal is minimized. The number of species must be known for both techniques, which is one of the limitations of this approach.

#### *Classical Least-Squares Regression (CLSR)*

In Classical Least-Squares Regression (CLSR) [20, 21, 22], also known as the K-Matrix method, the molar attenuation coefficients are calculated using calibration data:

$$\mathbf{K} = \mathbf{C}_c^{-1} \mathbf{X}_c \quad (2.2)$$

where the subscript  $c$  is used to distinguish the calibration data,  $\mathbf{X}_c \in \mathbb{R}^{n_{N,c} \times n_{L,c}}$  represents the measured spectra,  $\mathbf{C}_c \in \mathbb{R}^{n_{N,c} \times n_{K,c}}$  the concentrations and  $\mathbf{K} \in \mathbb{R}^{n_{K,c} \times n_{L,c}}$  the pure component contributions. Note that  $n_{N,c}$  is the number of measurements,  $n_{L,c}$  the number of wavenumbers and  $n_{K,c}$  the total number of species for the calibration data.

The coefficients from  $\mathbf{K}$  can be used to analyze mixtures with unknown composition, using:

$$\mathbf{X}_m = \mathbf{K} \mathbf{C}_m \quad (2.3)$$

where the subscript  $m$  is used to distinguish the mixture data and  $\mathbf{X}_m$  represents the spectral measurements of the unknown mixtures.

In CLSR, the number of wavelengths must be equal to or larger than the number of components. One of the main limitations of this method is the accuracy of the calibration set, which is affected by noise. The presence of noise may be misinterpreted as spectral activity, which will be stored in  $\mathbf{K}$ . The calibration set should therefore include the active spectral region for each of the components to minimize the effect of noise. Other limitations include the assumption of linearity and superposition, and the presence of peak overlap and collinear wavenumbers.

#### *Inverse Least-Squares Regression (ILSR)*

Inverse Least-Squares Regression (ILSR) [20, 21, 22], also known as the P-matrix method, is based on Beer-Lambert's law, but while CLSR models the spectral signal from information on the concentration and molar attenuation coefficients, ILSR uses the reverse approach to model the concentration as the product of the spectral signal and calibration coefficient matrix  $\mathbf{P}$ :

$$\mathbf{C}_c = \mathbf{P}\mathbf{X}_c \quad (2.4)$$

The coefficient matrix  $\mathbf{P}$  is calculated from calibration data and relates the concentration of each component with the spectral signal at each wavelength. In this case, the number of samples used in the calibration must be greater than the number of wavelengths used in  $\mathbf{X}_c$ . Once the coefficient matrix has been calculated, the concentration of an unknown mixture can be determined using:

$$\mathbf{C}_m = \mathbf{P}\mathbf{X}_m \quad (2.5)$$

ILSR is limited by the accuracy of the calibration set and the selection of wavelengths. Using fewer wavelengths may minimize the effect of noise, but will not account for minor spectral features that differentiate overlapping peaks. On the other hand, increasing the number of samples will allow us to add more wavelengths to the model, but may lead to collinearities that result in correlated regression coefficients.

Data reduction, such as Principal Component Analysis (PCA) [21, 23], is used to mitigate collinearity effects. PCA finds principal components that represent linear combinations of the original variables for a given observation. This is accomplished by projecting the original variables onto a new, smaller set of variables that are not correlated with each other. The magnitudes of the principal components (scores)  $\mathbf{T}$  are defined as:

$$\mathbf{T} = \mathbf{X}\mathbf{P} \quad (2.6)$$

where  $\mathbf{P}$  stores the direction vectors of the principal components (loadings).

### *Principal Component Regression (PCR)*

Principal Component Regression (PCR) is a technique that combines PCA with MLR [21, 24]. First, the data set is reduced using PCA and then MLR is applied to regress the reduced set.

### *Partial Least-Squares Regression (PLSR)*

Partial Least-Squares Regression (PLSR) is another regression tool that performs data reduction and regression simultaneously, utilizing both the independent and dependent data as input [21, 24]. PLSR often has superior predictive power compared to PCR, because it considers the input-output relationship in the variable reduction. More detailed explanations and derivations of these algorithms can be found elsewhere [25, 26, 27].

### 2.2.3 Experimental Design

The model system is based on a variation of the 5.6 M Na low-activity waste pretreatment system (LAWPS) simulant [28, 29], and includes water and four sodium salts: nitrate, nitrite, carbonate and sulfate. The species with higher concentrations,  $\text{NO}_3^-$  and  $\text{NO}_2^-$ , as well as  $\text{SO}_4^{2-}$ , which may interfere with the operation of the melter [11, 16], were designated as target species. In addition,  $\text{CO}_3^{2-}$  and  $\text{H}_2\text{O}$  were also treated as target species:  $\text{CO}_3^{2-}$  due to its abundance in the waste and signal overlap with nitrate for both Raman and IR spectra, and  $\text{H}_2\text{O}$  as the common solvent. Measurements were obtained simultaneously using *in-situ* Raman and infrared (attenuated total reflectance-Fourier transform infrared or ATR-FTIR) instruments. The temperature in the vessel was kept at 25°C. Raman spectra were collected using a Mettler Toledo ReactRaman 785 system equipped with a 785 nm laser, at 300 mW laser power and 0.75 s exposure time. IR spectra were collected using a Mettler Toledo ReactIR 10 system. Data acquisition was performed using iC Raman and iC IR software from Mettler Toledo. The experimental set-up is shown in Figure 2.2.

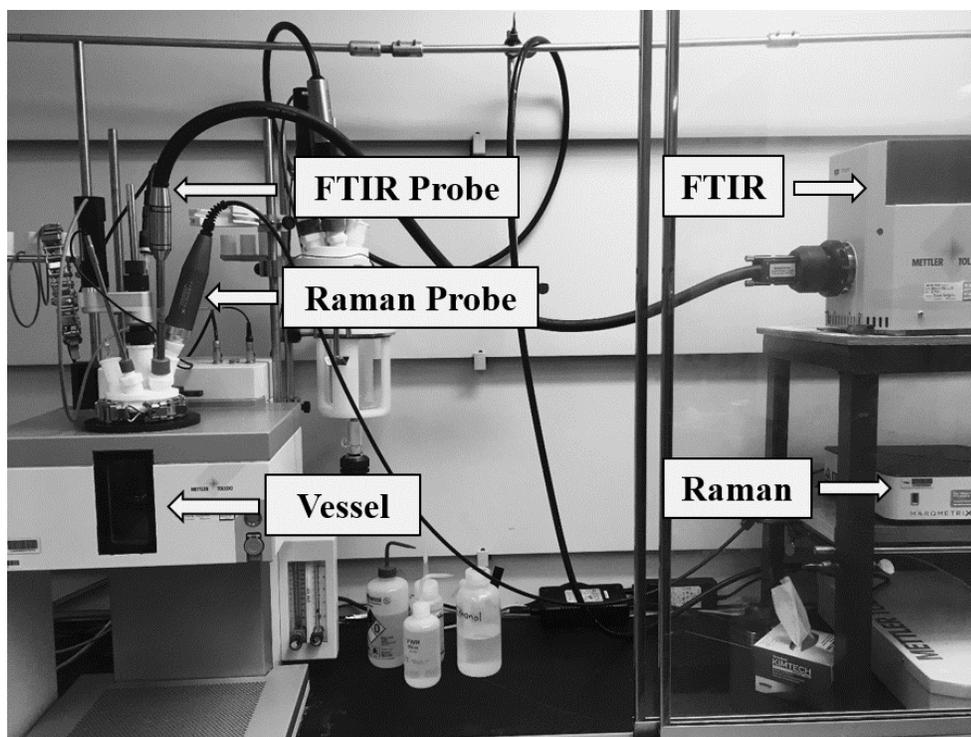


Figure 2.2: Experimental set-up.

The experimental design consisted of creating a training and test data sets of the mixtures. The wavenumber range was restricted to  $548 - 1774 \text{ cm}^{-1}$  for Raman and  $914 - 1824 \text{ cm}^{-1}$  for IR, which represent the regions of spectral activity for all species. A central composite design with four factors (nitrate, nitrite, carbonate and sulfate) at three levels (low, medium and high) was used to create the training data. Note that the solvent (water), is not counted as a factor despite being a target species. Overall, 36 mixtures were generated, including 12 repetitions of the center point based on a face centered design [30, 31]. The molal concentrations are given in Table 2.1. Each mixture in the test set contains a different combination of the target constituents. A total of 12 mixtures were generated; their concentrations are summarized in Table 2.2.

Table 2.1: Experimental design for the training set. The molal concentrations are calculated as the moles of the target species per kilogram of solvent (water).

<b>Sample</b>	<b>NO<sub>3</sub><sup>-</sup> mol kg<sup>-1</sup></b>	<b>NO<sub>2</sub><sup>-</sup> mol kg<sup>-1</sup></b>	<b>SO<sub>4</sub><sup>2-</sup> mol kg<sup>-1</sup></b>	<b>CO<sub>3</sub><sup>2-</sup> mol kg<sup>-1</sup></b>	<b>H<sub>2</sub>O mol kg<sup>-1</sup></b>
<b>1</b>	0.00	0.00	0.00	0.00	55.51
<b>2</b>	0.00	0.00	0.00	0.61	55.51
<b>3</b>	0.00	0.00	0.19	0.61	55.51
<b>4</b>	0.00	1.26	0.19	0.61	55.51
<b>5</b>	1.49	1.26	0.19	0.61	55.51
<b>6</b>	0.00	0.00	0.19	0.00	55.51
<b>7</b>	0.00	1.26	0.19	0.00	55.51
<b>8</b>	1.50	1.26	0.19	0.00	55.51
<b>9</b>	0.00	1.27	0.00	0.00	55.51
<b>10</b>	0.00	1.27	0.00	0.61	55.51
<b>11</b>	1.50	1.27	0.00	0.61	55.51
<b>12</b>	1.48	0.00	0.00	0.00	55.51
<b>13</b>	1.48	0.00	0.19	0.00	55.51
<b>14</b>	1.48	0.00	0.19	0.61	55.51
<b>15</b>	1.48	0.00	0.00	0.61	55.51
<b>16</b>	1.47	1.25	0.00	0.00	55.51
<b>17</b>	0.00	0.63	0.10	0.31	55.51
<b>18</b>	1.49	0.63	0.10	0.31	55.51
<b>19</b>	0.75	0.00	0.10	0.31	55.51
<b>20</b>	0.75	1.27	0.10	0.31	55.51
<b>21</b>	0.74	0.62	0.00	0.31	55.51
<b>22</b>	0.74	0.62	0.19	0.31	55.51
<b>23</b>	0.74	0.63	0.10	0.00	55.51
<b>24</b>	0.74	0.63	0.10	0.61	55.51
<b>25</b>	0.75	0.63	0.10	0.31	55.51
<b>26</b>	0.74	0.63	0.10	0.31	55.51
<b>27</b>	0.75	0.63	0.09	0.31	55.51
<b>28</b>	0.75	0.63	0.10	0.31	55.51
<b>29</b>	0.75	0.63	0.10	0.31	55.51
<b>30</b>	0.75	0.63	0.10	0.31	55.51
<b>31</b>	0.75	0.63	0.10	0.31	55.51
<b>32</b>	0.75	0.63	0.10	0.31	55.51
<b>33</b>	0.75	0.63	0.10	0.31	55.51
<b>34</b>	0.75	0.63	0.09	0.30	55.51
<b>35</b>	0.75	0.63	0.09	0.31	55.51
<b>36</b>	0.75	0.63	0.10	0.31	55.51

Table 2.2: Experimental design for the test set. The molal concentrations are calculated as the moles of the target species per kilogram of solvent (water).

Sample	$\text{NO}_3^-$ $\text{mol kg}^{-1}$	$\text{NO}_2^-$ $\text{mol kg}^{-1}$	$\text{SO}_4^{2-}$ $\text{mol kg}^{-1}$	$\text{CO}_3^{2-}$ $\text{mol kg}^{-1}$	$\text{H}_2\text{O}$ $\text{mol kg}^{-1}$
1	1.17	0.97	0.14	0.50	55.51
2	0.55	0.39	0.05	0.13	55.51
3	1.44	1.16	0.10	0.58	55.51
4	1.22	0.97	0.03	0.59	55.51
5	1.45	0.76	0.10	0.45	55.51
6	1.22	0.76	0.06	0.30	55.51
7	1.45	1.17	0.10	0.30	55.51
8	1.00	1.17	0.06	0.59	55.51
9	1.22	0.97	0.03	0.45	55.51
10	1.00	0.97	0.10	0.45	55.51
11	1.00	1.17	0.03	0.45	55.51
12	1.44	0.75	0.10	0.44	55.51

### 2.3 Results and Discussion

Spectroscopic fingerprints of the components are shown in Figure 2.3. The overlap between the nitrate and carbonate peak is apparent in both Raman and IR spectra, but is more apparent in the IR region, which also overlaps with nitrite.

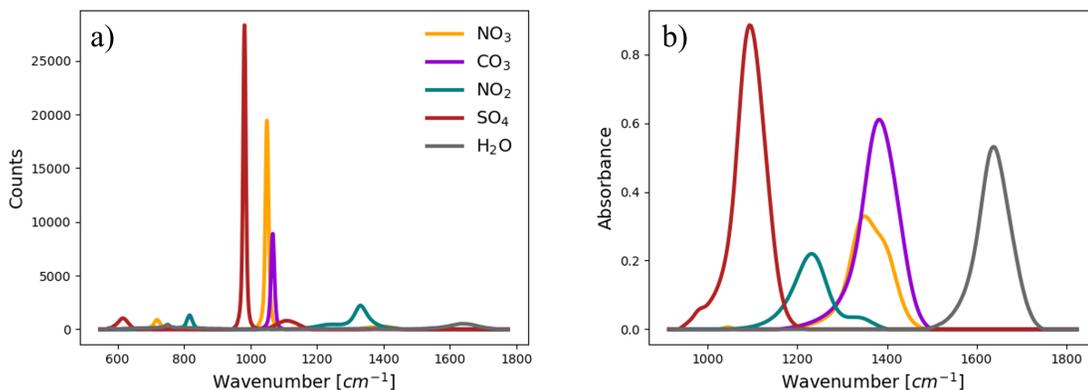


Figure 2.3: Reference spectra of the species in our model system: a) Raman spectroscopy; b) ATR-FTIR spectroscopy. The data show the characteristic peaks of each species.

The training and test data for the chemometric models are shown in Figure 2.4 and Figure 2.5. The Raman and IR spectra were preprocessed by smoothing using a Savitzky-Golay filter [22]. The Raman spectra were further preprocessed by performing a baseline correction [32]. Note that the solvent (water) spectrum was not subtracted from the Raman/IR spectra during the preprocessing steps. Past work has showed that the water baseline in IR is sensitive to many disturbance variables, including the position of the probe and agitation speed, and therefore may introduce error in the spectra when subtracted [17].

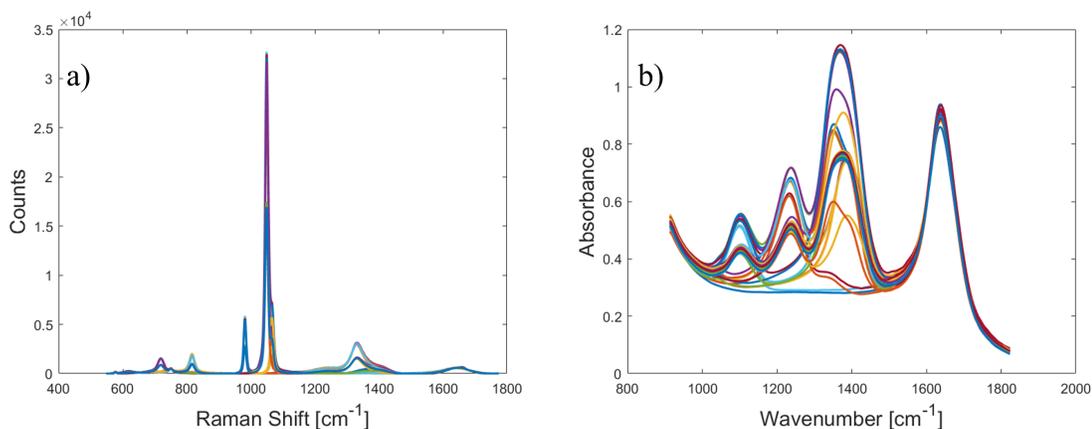


Figure 2.4: Spectral measurements of the training data set mixtures for a) Raman spectroscopy and b) ATR-FTIR spectroscopy. Concentrations for each measurement are shown in Table 2.1.

The initial model comparison for both Raman and ATR-FTIR, using five principal components in the PLSR and PCR models, is shown in Figure 2.6. The chemometric models were designed using only molalities of the four analytes and excluded the solvent concentration. The measured concentrations were calculated with a gravimetric analysis of the amounts of solute used in each mixture.

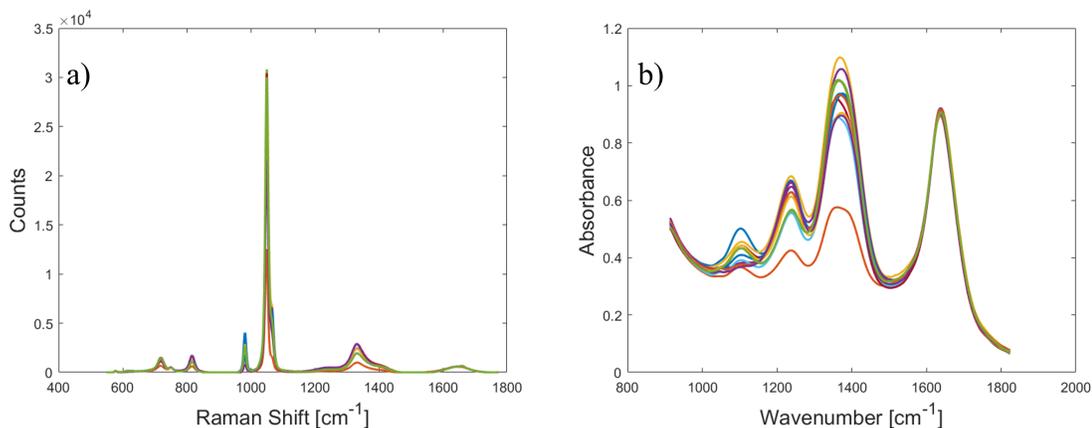


Figure 2.5: Spectral measurements of the test data set mixtures for a) Raman spectroscopy and b) ATR-FTIR spectroscopy. Concentrations that match each measurement are shown in Table 2.2.

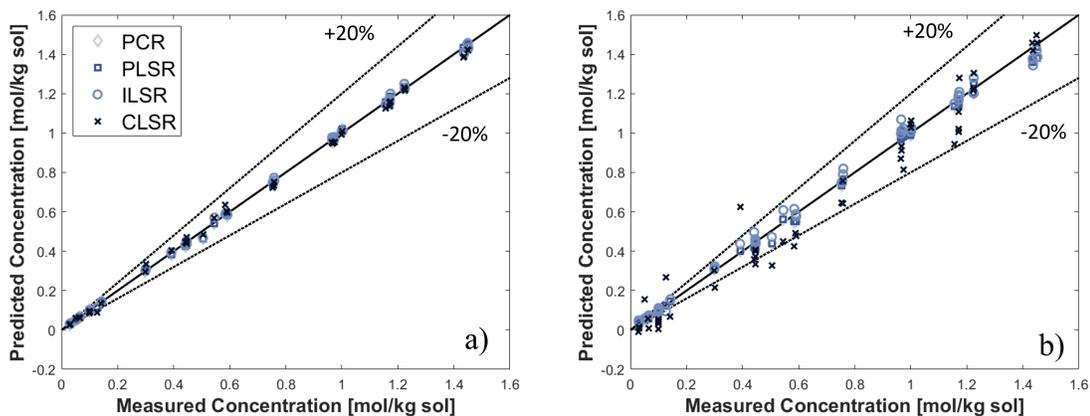


Figure 2.6: PCR, PLSR, ILSR and CLSR model predictions using a) Raman and b) ATR-FTIR data. The solid black line represents unity and the dashed black lines represents  $\pm 20\%$  deviations, which are based on the uncertainty at the Hanford analytical laboratory [11].

### 2.3.1 ATR-FTIR Data - Closer Look

The parity plots from Figure 2.6 facilitate some general observations, as well as trends that vary for each spectroscopic technique. The main observation from Figure 2.6 is that the ATR-FTIR predictions are much less accurate than those made using the same models on Raman data. Two possible explanation for prediction errors in the ATR-FTIR data include the peak overlap among the species (which is more pronounced in IR compared to the Raman spectra), and the water baseline in IR, since water absorbs more strongly in

IR and exhibits a distinct peak in the  $1600\text{ cm}^{-1}$  region. With the current model, the IR peak of water is present in the spectra, but not accounted for in the concentration matrix  $C$ , which only contains molalities of the four analytes (nitrate, nitrite, sulfate and carbonate). In order to incorporate the water concentration data while training the models, the concentration matrix  $C$  was augmented to include the water concentrations. The updated results are shown in Figure 2.7. The results from Figure 2.7 show that incorporating the water concentrations in the training data for ATR-FTIR has a positive effect on the model predictions of the concentrations, across all models. The predicted concentrations of the individual analytes are shown in Figure 2.8.

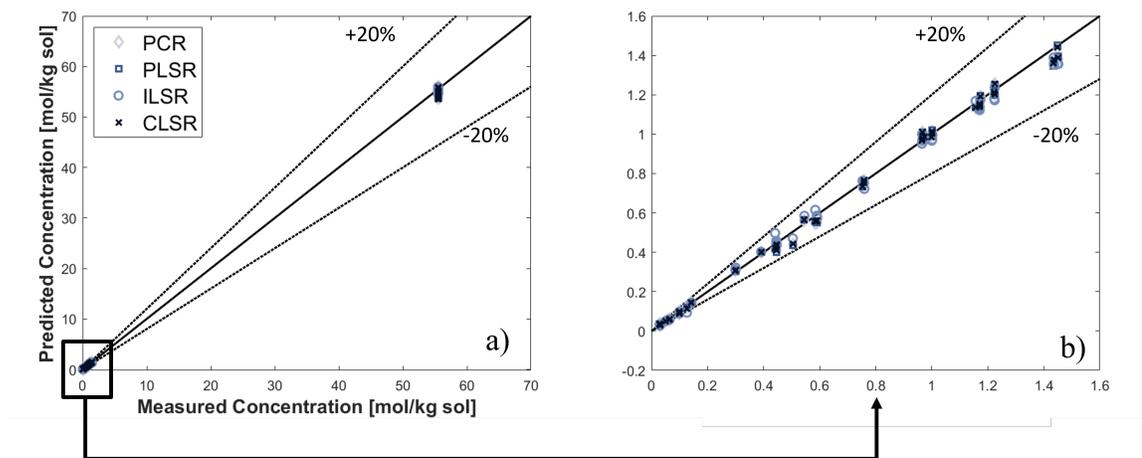


Figure 2.7: Updated ATR-FTIR model results (including water) showing predictions for a) the full range of concentrations and b) the analyte range. The solid black line represents unity and the dashed black lines represents  $\pm 20\%$  deviations, which are based on the uncertainty at the Hanford analytical laboratory [11].

### 2.3.2 Raman Data - Closer Look

All four linear models were able to predict the analyte concentrations well using Raman spectra, as shown in Figure 2.6. The predicted concentrations of the individual analytes are shown in Figure 2.9.

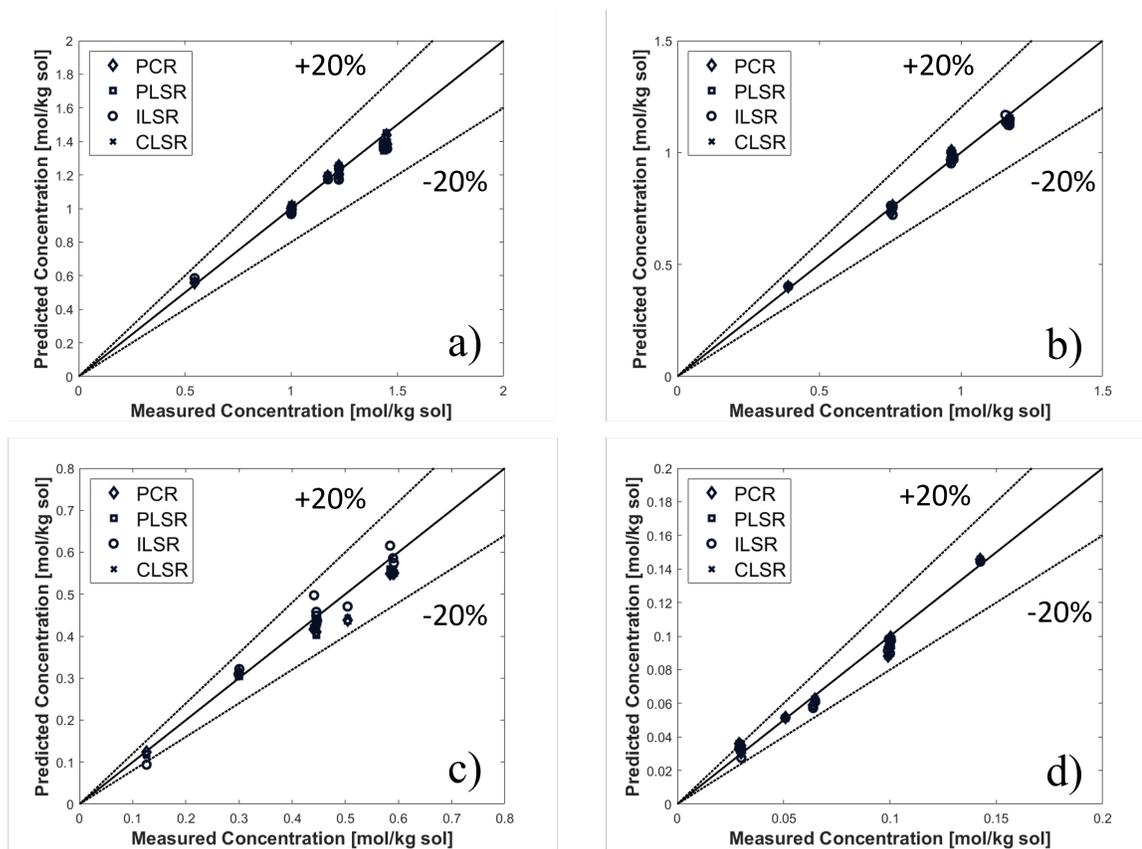


Figure 2.8: PCR, PLSR, ILSR and CLSR model predictions on ATR-FTIR data for a) nitrate, b) nitrite, c) carbonate and d) sulfate. The solid black line represents unity and the dashed black lines represents  $\pm 20\%$  deviations, which are based on the uncertainty at the Hanford analytical laboratory [11].

Since the water concentrations were incorporated in the ATR-FTIR models, we also ran a scenario where the Raman models included the water concentrations in the training data. Note that the number of principal components (5) remained unchanged for both the ATR-FTIR and Raman models. The Raman results are shown in Figure 2.10 (a). The results indicate that the models were not able to predict the concentration of water as well as the ATR-FTIR models. A potential explanation may involve the appearance of the water peak, which only shows up as a minor peak in the  $1600\text{ cm}^{-1}$  region. Another potential reason is the number of principal components, which may not be sufficient for the Raman system (a nonlinearity associated with a blue shift of the nitrate peak is discussed in more detail

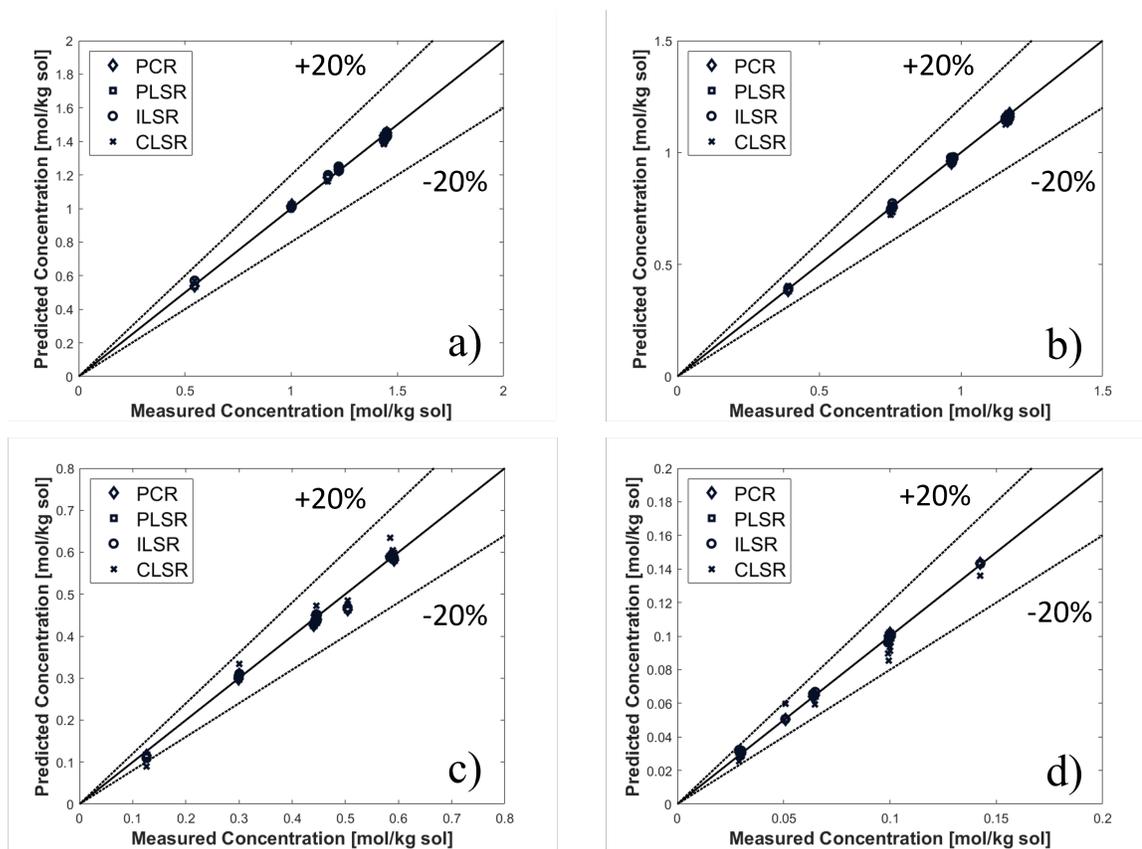


Figure 2.9: PCR, PLSR, ILSR and CLSR model predictions on Raman data for a) nitrate, b) nitrite, c) carbonate and d) sulfate. The solid black line represents unity and the dashed black lines represents  $\pm 20\%$  deviations, which are based on the uncertainty at the Hanford analytical laboratory [11].

in Chapters 3 and 4). Once we introduced an additional principal component, the water concentration predictions improved for the ILSR, PLSR and PCR models, which is shown in Figure 2.10 (b).

## 2.4 Conclusions

The work in this chapter has revealed key features of ATR-FTIR and Raman spectroscopy in the context of online monitoring of salt species in nuclear waste. Both ATR-FTIR and Raman spectroscopy were able to detect the process-relevant species from DFLAW. The study showed that the ATR-FTIR technique, characterized by larger peaks that exhibit more

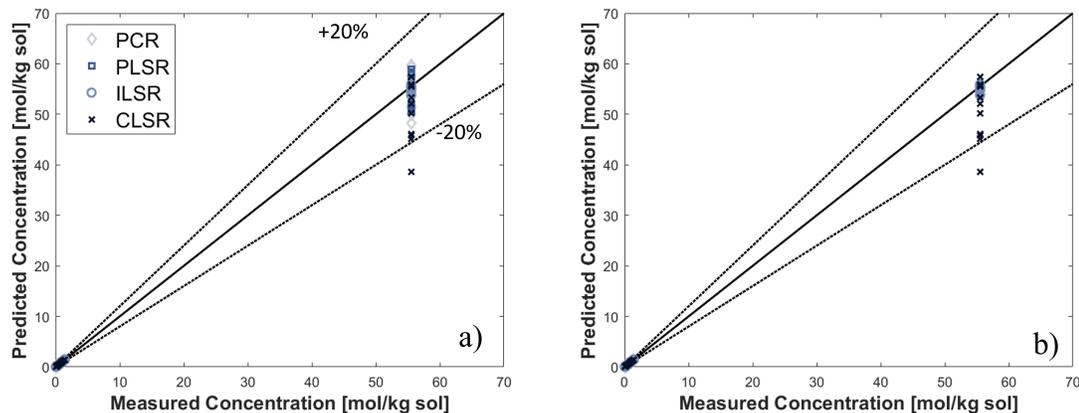


Figure 2.10: Updated Raman model results (including water) showing predictions for a) the full range of concentrations using five principal components in the PLSR and PCR models and b) the full range of concentrations using six principal components in the PLSR and PCR models. The solid black line represents unity and the dashed black lines represents  $\pm 20\%$  deviations, which are based on the uncertainty at the Hanford analytical laboratory [11].

pronounced overlap, is highly sensitive to the presence of water in the system. On the other hand, Raman may be susceptible to nonlinear phenomena which affect the model and will be discussed in the upcoming chapters. The PLSR and PCR models generally provided the most accurate estimates of the concentrations of species, since they model the change in spectral activity as a result of the change in concentration and are thus able to deconvolute overlapping peaks more effectively. However the ILSR model also performed well, even though it only uses a few key wavelengths, which shows that sometimes the simpler model may provide more robust estimates. ILSR is highly dependent on the wavenumber selection, so these estimates will change if different wavenumbers are selected in the calibration, while PLSR and PCR reduce the full spectrum and give more consistent predictions overall. A final note is that these models may not provide such accurate estimates in the presence of additional species that are not part of the training set, which is highly probable at Hanford, considering the waste is made up of many target and non-target species, indicating that we cannot rely solely on calibration approaches for accurate estimates of the composition of low-activity waste.

**CHAPTER 3**  
**BLIND SOURCE SEPARATION – PARTIAL LEAST SQUARES REGRESSION**  
**MODEL FOR QUANTIFICATION OF TARGET SPECIES IN LIQUID**  
**LOW-ACTIVITY WASTE SIMULANTS<sup>1</sup>**

### **3.1 Introduction**

Current approaches to measuring the composition of multicomponent mixtures include developing spectra-to-composition chemometrics models, which need to be trained with calibration data [17, 1]. The calibration data may be sensitive to process fluctuations, such as changes in the temperature and pH. Another challenge, specifically associated with complex mixtures having many constituents, is the size of the calibration data set. For example, clean-up efforts at the Hanford site are highly dependent on the composition of the nuclear waste, which contains 25 chemical constituents and 46 radionuclides, and is differentiated in three categories (liquid, saltcake and sludge) [5].

Previous attempts at calculating the composition of the liquid waste consist of sample-collection procedures on the order of hundreds of samples to train a chemometrics model, such as partial least-squares regression (PLSR) [34, 14, 35]. Even though this method can quantify the concentrations of the target species, it also generates extra information on many minor species that may not be necessary to operate the process. Furthermore, the calibration procedure may need to be repeated or the training data set extended in the presence of any variations with respect to the species. This means that tailored models and new calibrations are needed any time the new test data varies in composition from the set of designated training data. This may occur when moving across different sampling locations

---

<sup>1</sup>Adapted from "S. Kocavska, G. M. Maggioni, R. W. Rousseau, and M. A. Grover, "Spectroscopic Quantification of Target Species in a Complex Mixture Using Blind Source Separation and Partial Least-Squares Regression: A Case Study on Hanford Waste," *Ind. Eng. Chem. Res.*, vol. 60, no. 27, pp. 9885–9896, 2021. Copyright 2021 American Chemical Society." [33]

in the DFLAW plant or even if there are variations in the feed.

On the other hand, blind source separation (BSS) techniques, such as independent component analysis, which can decompose mixtures into its constituents with little or no prior information, are emerging in the field of analytical chemistry [36]. Similarly to the more recognized spectra-to-composition chemometric models (such as PLSR), BSS models, including Independent Component Analysis (ICA), rely on linearity and superimposition assumptions in the Beer-Lambert law. When applied to spectroscopy data, the goal of BSS is to approximate the signal of the original components found in the mixtures, as well as their contributions to the mixture. However, since both the mixture identity and composition are unknown at the start, BSS is limited in quantifying the concentrations of the components in the mixtures [37, 38, 39, 40].

To both mitigate the limitations of BSS and take advantage of its efficient identification of independent components from mixture spectra, we propose a new methodology that uses blind source separation techniques to classify and preprocess the mixture spectra, followed by a regression to quantify the concentration of the target species. First, BSS is applied to identify the independent components (sources) in the mixture, and then a matrix of correlation coefficients is calculated to classify the components as targets (if they correlate with reference spectra of the target species) or non-targets. This information is used to preprocess the mixture data by removing the signals of the non-target components.

While the class of target species usually comprises a small number of process-relevant constituents, the class of non-target species can be as large as necessary. Other than nuclear waste, this approach is applicable to any mixture system in which non-target species may interfere with the spectroscopic signals of the target species, such as quantification of nitrate in wastewater [41, 42]. In the final step, the concentrations of the target species are quantified using partial least-squares regression (PLSR), performed on the BSS-preprocessed data.

## 3.2 Methods

### 3.2.1 Experimental

The model system is based on a variation of the 5.6 M Na low-activity waste pretreatment system (LAWPS) simulant [28, 29], and includes water and seven sodium salts: nitrate, nitrite, carbonate, sulfate, phosphate, oxalate and acetate. For demonstration and practical purposes, the species with higher concentrations,  $\text{NO}_3^-$  and  $\text{NO}_2^-$ , as well as  $\text{SO}_4^{2-}$  which may interfere with the operation of the melter [11, 16], were labeled as target species. In addition,  $\text{CO}_3^{2-}$  and  $\text{H}_2\text{O}$  were also treated as target species:  $\text{CO}_3^{2-}$  due to its abundance in the waste and signal overlap with nitrate for both Raman and IR spectra, and  $\text{H}_2\text{O}$  as the common solvent. The remaining species with lower concentrations,  $\text{PO}_4^{3-}$ ,  $\text{C}_2\text{O}_4^{2-}$  and  $\text{CH}_3\text{COO}^-$ , were labeled as the non-target species. However, the classification of target and non-target species could be different in practice, based on process expectations and goals.

Measurements were obtained simultaneously using *in-situ* Raman and infrared (attenuated total reflectance-Fourier transform infrared or ATR-FTIR) instruments. The temperature in the vessel was kept at 25°C. Raman spectra were collected using a Mettler Toledo ReactRaman 785 system equipped with a 785 nm laser, at 300 mW laser power and 0.75 s exposure time. IR spectra were collected using a Mettler Toledo ReactIR 10 system. Data acquisition was performed using iC Raman and iC IR software from Mettler Toledo.

### 3.2.2 Modeling

As discussed previously, the performance of common chemometric models is limited by the extent and complexity of the calibration. For complex systems, such as nuclear waste mixtures, a training data set that covers every possible combination of the spectroscopically active species requires many independent measurements. Design of experiments (DOE) could be used to reduce the size of the training set, but even DOE would be limited

in optimizing a data set with a large number of species (DOE factors). Another common approach is variable selection, where wavelength regions that correspond to the target analytes are carefully selected to improve the model performance [43, 44]. This idea can also be extended to develop local models [45]. While variable selection is commonly used as a data-reduction technique, similar to principal component analysis (PCA), it could also be used to simplify the measurements. Since specific wavelengths are to be used to quantify the concentrations, the model can be trained with only the target species. For the aforementioned DFLAW nuclear waste with four factors at three levels, the number of measurements for a full factorial design would be 81. However, this simplification could only work if there is no overlap between the spectra of the target and non-target species, so that the signals of the non-targets do not interfere with those of the target species. Since Hanford tank waste mixtures contain species whose spectroscopic signals overlap, a full calibration that accounts for the complexity of the data set is employed for accurate spectra-to-composition calculations [34]. In this paper, we bypass the need for a full calibration by removing the signals of the non-target species with BSS.

### *Overview of Blind Source Separation*

Independent Component Analysis (ICA), one of the most widely used methods to perform BSS, seeks to identify the independent sources from mixture data [37, 39, 38]. The goal of the algorithm is to solve:

$$\mathbf{X} = \mathbf{C}\mathbf{L} = \mathbf{A}\mathbf{S} \quad (3.1)$$

where  $\mathbf{X} \in \mathbb{R}^{n_N \times n_L}$  represents the measured spectra,  $\mathbf{C} \in \mathbb{R}^{n_N \times n_K}$  the concentrations and  $\mathbf{L} \in \mathbb{R}^{n_K \times n_L}$  the pure component contributions. Note that  $n_N$  is the number of measurements,  $n_L$  the number of wavenumbers and  $n_K$  the total number of species. To find the approximate solution, ICA identifies a mixing matrix  $\mathbf{A} \in \mathbb{R}^{n_N \times \hat{n}_K}$  and a sources matrix  $\mathbf{S} \in \mathbb{R}^{\hat{n}_K \times n_L}$ , where  $\hat{n}_K$  is the number of independent components or species estimated by the algorithm. The orthogonal rotation of the mixture spectra seeks to minimize the mutual

information or maximize the non-Gaussianity in the data, which have been preprocessed by centering, whitening and dimensionality reduction. Another popular BSS technique is Multiple Curve Resolution-Alternating Least Squares (MCR-ALS) [46], which similarly to ICA, also decomposes the data into source and mixing matrices, but without relying on mutual independence. MCR-ALS seeks a solution to:

$$\min_{\mathbf{A}, \mathbf{S}} \|\mathbf{X} - \mathbf{AS}\| \quad (3.2)$$

The MCR-ALS algorithm obtains a sources matrix—indicative of the pure components, and a mixing matrix—indicative of the composition. Furthermore, MCR-ALS imposes a constraint, such that the elements in the mixing matrix  $\mathbf{A} \in \mathbb{R}^{n_N \times \hat{n}_K}$  are positive numbers. Since MCR-ALS requires an initial guess, in this work, as well as previous works, we combined ICA and MCR-ALS using the results from ICA to initiate MCR-ALS. Detailed explanations and derivations of BSS techniques can be found elsewhere [40]. To perform the ICA algorithm, we used FastICA [37], which maximizes the non-Gaussianity of the sources. The MCR-ALS algorithm was performed using pyMCR, developed by NIST [47].

### *Design of Experiments*

The experimental design consisted of creating a library of the reference spectra, as well as test and training data sets of the mixtures. The use of each data set is explained in Section 3.2.2. To create a library of the reference spectra, each analyte was dissolved in water and its characteristic signal measured. The wavenumber range was restricted to  $548 - 1774 \text{ cm}^{-1}$  for Raman and  $914 - 1824 \text{ cm}^{-1}$  for IR, which represent the regions of spectral activity for all species. Next, the background signal of the solvent was subtracted, to isolate the peaks of the pure components. A baseline correction was also performed [32], and the signal intensity in spectrally inactive regions was set to zero. Finally, each spectrum underwent smoothing using a Savitzky-Golay filter [22], and was divided by its

associated concentration to obtain unimolal references, shown in Figure 3.2.

Next, a test data set,  $\mathbf{X}_{ts}$ , was defined, representing the DFLAW mixtures that need to be quantified and therefore contains both the target and non-target species. More specifically, each mixture in  $\mathbf{X}_{ts}$  contains a different combination of the non-target constituents, to include varying levels of interference on the signal of the targets. A total of 21 mixtures were generated; their concentrations are summarized in Table 3.1.

Table 3.1: Experimental design for the test set. Each mixture contains the four target species and a different combination of the non-target species.

Mixture Index	$\text{NO}_3^-$ [mol kg $^{-1}$ ]	$\text{NO}_2^-$ [mol kg $^{-1}$ ]	$\text{SO}_4^{2-}$ [mol kg $^{-1}$ ]	$\text{CO}_3^{2-}$ [mol kg $^{-1}$ ]	$\text{H}_2\text{O}$ [mol kg $^{-1}$ ]	$\text{PO}_4^{3-}$ [mol kg $^{-1}$ ]	$\text{CH}_3\text{COO}^-$ [mol kg $^{-1}$ ]	$\text{C}_2\text{O}_4^{2-}$ [mol kg $^{-1}$ ]
1	1.17	0.97	0.14	0.50	55.51	0.00	0.00	0.02
2	1.17	0.97	0.14	0.50	55.36	0.00	0.08	0.02
3	1.14	0.95	0.14	0.49	54.37	0.07	0.19	0.02
4	0.54	0.39	0.05	0.12	54.75	0.07	0.00	0.00
5	0.54	0.38	0.05	0.12	54.60	0.07	0.07	0.00
6	0.54	0.38	0.05	0.12	54.60	0.07	0.07	0.05
7	0.54	0.38	0.05	0.13	54.60	0.07	0.07	0.05
8	0.96	0.85	0.14	0.41	54.60	0.07	0.07	0.05
9	0.75	0.63	0.10	0.31	55.51	0.00	0.00	0.03
10	0.73	0.63	0.10	0.25	55.22	0.00	0.14	0.00
11	0.72	0.61	0.09	0.30	54.20	0.12	0.00	0.00
12	1.39	1.12	0.10	0.57	54.07	0.10	0.19	0.01
13	1.21	0.95	0.03	0.58	54.77	0.05	0.10	0.01
14	1.43	0.75	0.10	0.44	54.77	0.05	0.10	0.00
15	1.21	0.75	0.06	0.30	54.96	0.05	0.00	0.00
16	1.41	1.13	0.10	0.29	54.05	0.10	0.19	0.01
17	0.99	1.16	0.06	0.59	55.31	0.00	0.10	0.01
18	1.20	0.95	0.03	0.44	54.76	0.05	0.10	0.01
19	0.98	0.94	0.10	0.43	54.25	0.10	0.10	0.00
20	1.00	1.17	0.03	0.45	55.51	0.00	0.00	0.01
21	1.39	0.73	0.10	0.43	54.07	0.10	0.19	0.00

A training data set,  $\mathbf{X}_{tr}$ , was also defined, which is used to train the PLSR model, and is made of mixtures of the target species only. A central composite design with four factors (nitrate, nitrite, carbonate and sulfate) at three levels (low, medium and high) was used to create the training data. Note that the solvent (water), is not counted as a factor despite being a target species. Overall, 36 mixtures were generated, including 12 repetitions of the center point based on a face centered design [30, 31]. The concentrations are given in Table 2.1 (the same training data set was used in Chapter 2). For this study, training the model without the non-target species has decreased the number of required training measurements from 100 (face centered design with seven factors) to 36 (face centered design with four factors). The simplification is only possible because the BSS framework will remove the signals of the non-target species from mixtures in  $\mathbf{X}_{ts}$ .

Both the Raman and IR mixture spectra in the test and training data sets were subjected to smoothing. In addition, background correction was performed for the Raman data.

### *BSS-PLSR Framework*

The preprocessing and quantification framework can be outlined in four steps, of which the first three preprocess the data based on BSS and the last one quantifies the target concentrations using PLSR. The Python code for BSS-PLSR can be found on GitHub: <https://github.com/skocevaska/BSS-PLSR-for-Spectroscopy-Analysis>.

1. **Identifying the independent components (sources):** First, BSS methods outlined in Section 3.2.2, ICA and MCR-ALS, are used to analyze the mixture data  $\mathbf{X}$  and identify the independent components or sources,  $\mathbf{S} \in \mathbb{R}^{\hat{n}_K \times n_L}$ . The number of independent components,  $\hat{n}_K$ , is determined using an automated analysis of the singular value decomposition (SVD) of the mixture data, as reported in our previous work, which also includes more details on the ICA and MCR-ALS procedures [40]. Typically BSS is applied to the mixture data that needs to be quantified, which in our case is the test data set,  $\mathbf{X}_{ts}$ . However, in this work, the mixtures in the test data set

exhibit collinearities (caused by insufficient variability in the concentrations of the target species), which hinders the correct identification of independent sources by the BSS algorithm. To mitigate the collinearity effect, the BSS input is augmented with the training data set, which contains sufficient variability of the target species. Finally, the BSS input is also augmented with the library of reference spectra (targets and non-targets),  $\mathbf{L}$ , to further ensure that each species will be identified, including the subset of non-targets. Therefore, the data input to BSS in Step 1 of the framework is:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{ts} \\ \mathbf{X}_{tr} \\ \mathbf{L} \end{bmatrix} \quad (3.3)$$

The prior information we have provided to the algorithm should be readily available in practice, considering that the training data set is originally created to train the PLSR model, and a working spectral library of the species expected to be in the mixtures can be collected prior to analysis. It is important to note that the BSS input was augmented with additional spectral information specific to the model system for the sole purpose of correctly identifying the independent components or sources. The remaining analyses focus on quantifying the mixtures in the test data set only ( $\mathbf{X}_{ts}$ ).

2. **Correlating the components with target species:** In the second step, a library of the target species,  $\mathbf{L}_t \in \mathbb{R}^{\tilde{n}_K \times n_L}$ , is introduced to facilitate identification of components and subsequent classification.  $\mathbf{L}_t$  is a subset of the full library,  $\mathbf{L}$ , and is made of reference spectra for nitrate, nitrite, sulfate, carbonate and water, thus  $\tilde{n}_K$  is set to five. Next, a matrix of Pearson correlation coefficients [48],  $\mathbf{R}$ , is generated to compare the normalized spectra from the target library with the normalized independent components generated by BSS. Each reference spectrum in  $\mathbf{L}_t$  is correlated with each independent component (source) in  $\mathbf{S}$ . The resulting matrix of correlation

coefficients has five rows (for each target species) and  $\hat{n}_K$  columns (for each source spectrum identified by BSS). The maximum value in each row identifies the source spectrum that correlates with the target species associated with the row. The sources which do not correlate with the target reference spectra are classified as the non-target species, and their total number is the difference between  $\hat{n}_K$  and  $\tilde{n}_K$ .

- 3. Removing the signal of non-target components:** In the third step, the signal of the non-target sources is calculated using the rows from  $\mathbf{S}$  that did not correlate with  $\mathbf{L}_t$  in Step 2 and the corresponding columns in the mixing matrix  $\mathbf{A}$ , following Equation 3.1. Once identified, their contributions are simply subtracted from the mixture spectra. An overview of the preprocessing steps is shown in Figure 3.1.
- 4. Quantifying the concentrations of the target species:** After preprocessing, the concentrations of the target species are quantified using a partial least-squares regression with five principal components. The PLSR model is popular in chemometrics, since it uses both the independent and dependent variables (concentration and spectral intensity, respectively), and the relationship between the two to determine the regression coefficients. The PLSR approach is described in more detail elsewhere [25, 26]. The test data set in the PLSR model is comprised of the preprocessed mixture spectra from Step 3 of the BSS-PLSR framework. The training data set is comprised of measurements of the target species only ( $\mathbf{X}_{tr}$ ), based on the assumption that the BSS preprocessing steps have completely removed the signals of the non-target species from the test data set. Therefore, the goodness of the PLSR quantification of the target species depends on the results of the BSS algorithm. In order to minimize any baseline effects and to resolve any overlapping peaks, derivative spectroscopy was used in the PLSR model [49]. More specifically, the first derivative of the test and training data set spectra was used, since it improves the resolution of overlapping peaks, but the loss in signal-to-noise ratio is not as prominent as when the second

derivative is used. In addition, higher order derivatives tend to discriminate against wider peaks, such as that of nitrite, which further motivated the use of the first derivative of the signal [50]. Derivative spectroscopy was applied using a Savitzky-Golay filter with a second order polynomial, which also smooths the data [22]. The size of the Savitzky-Golay window length was seven for Raman spectra and five for IR spectra.

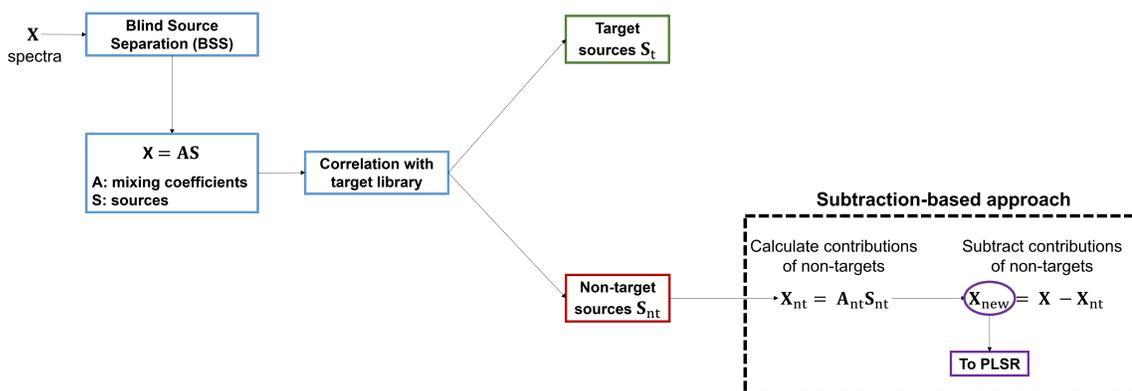


Figure 3.1: Blind source separation (BSS) preprocessing framework, based on a subtraction-based approach.

In order to evaluate the potential of the BSS-PLSR framework, both BSS-processed and unprocessed data were analyzed with the same PLSR model. This comparison shows how the BSS-processing steps can affect the results of the PLSR model, which was not trained with any non-target species. In this work, BSS-preprocessing is used to improve the flexibility of the PLSR model, but other conventional methods may include using DOE to introduce information on the non-target species or restricting the wavenumber region, as discussed in Section 3.2.2.

### 3.3 Results and Discussion

Figure 3.2 shows that the most significant overlap in Raman occurs in the nitrate region, at around  $1050\text{ cm}^{-1}$  (between nitrate and carbonate), and in the region around  $1400\text{ cm}^{-1}$  (between nitrite, acetate and oxalate). Additionally, there is overlap in the region around

950  $\text{cm}^{-1}$ , among phosphate, acetate and oxalate, and to a lesser degree, among phosphate and sulfate. The most prominent overlap in ATR-FTIR is in the region between 1250 and 1650  $\text{cm}^{-1}$ , where both target and non-target species have their characteristic peaks, including nitrite, nitrate, carbonate, oxalate, acetate and water. Phosphate and sulfate also overlap in the region around 1050  $\text{cm}^{-1}$ , but to a lesser degree. It is notable that IR exhibits significant overlap between the target and non-target species, while the overlap in Raman is more pronounced in each individual subset (*i.e.*, the non-target species overlap among themselves, and there is less overlap of this subset with the targets).

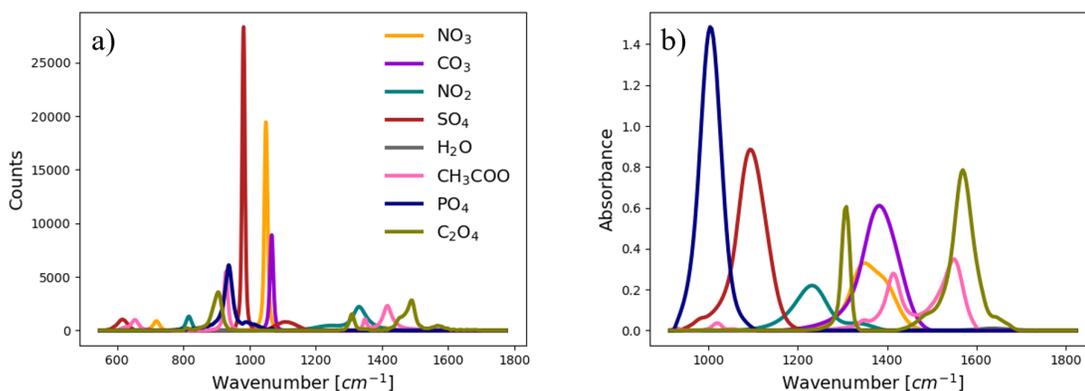


Figure 3.2: Reference spectra of the species in our model system: a) Raman spectroscopy; b) ATR-FTIR spectroscopy. The spectra show the extent of overlap among the target and non-target species.

### 3.3.1 Simulated Data

Prior to performing physical experiments, spectra of the mixtures were simulated according to the experimental design. The *in silico* data were subjected to the BSS-PLSR framework, to evaluate how it performs in ideal conditions. The spectra are simulated by multiplying each unimolar reference spectrum by the desired concentration, and adding the contributions of each species to create a mixture with a given composition. Next, to simulate the presence of noise in the experimental measurements, a predetermined level of white, Gaus-

sian noise is added to the spectra. For example, to simulate the test set from Table 3.1, using the eight reference spectra from Figure 3.2, one can evaluate:

$$\mathbf{X}_{ts,s} = \sum_{k=1}^{n_K} \mathbf{C}_{ts,s}^k \mathbf{L}^k + \boldsymbol{\eta} \quad (3.4)$$

where the subscript  $s$  is used to distinguish the simulated data, the superscript  $k$  represents each species,  $\mathbf{C}_{ts,s} \in \mathbb{R}^{n_N \times n_K}$  represents the concentrations,  $\mathbf{L} \in \mathbb{R}^{n_K \times n_L}$  is the library of the reference spectra and  $\boldsymbol{\eta} \in \mathbb{R}^{n_K \times n_L}$  represents the noise, which is assumed to be additive to the signal. The level of added noise was based on the measurement noise associated with the experimental system, represented by the variance of the noise,  $\sigma_\eta^2$ . This value was used to calculate the signal-to-noise ratio ( $SNR$ ). The local  $SNR$  (labeled with subscript  $l$ ) for measurement  $i$  at a specific wavenumber  $j$  can be calculated as:

$$SNR_l^{ij} = \frac{X_{ij}^2}{\sigma_\eta^2} \quad (3.5)$$

The average  $SNR$  (labeled with subscript  $avg$ ) for measurement  $i$  can be calculated as:

$$SNR_{avg}^i = \frac{1}{n_L} \sum_{j=1}^{n_L} SNR_l^{ij} \quad (3.6)$$

All  $SNR$  values were converted to decibel (dB), defined as  $10 \log_{10}(SNR)$ , for better resolution at both lower and higher levels. In this work, the  $SNR$  values for Raman and IR were calculated from experiments and similar levels were used in the simulated data, so that the simulations resemble the conditions of real experiments. The same approach was used to simulate the training data,  $\mathbf{X}_{tr,s}$ , using the experimental design from Table 2.1.

### *Raman Spectroscopy*

The simulated Raman spectra for  $\mathbf{X}_{ts,s}$  and  $\mathbf{X}_{tr,s}$  are shown in Figure 3.3(a) and Figure 3.4(a). Note that noise that corresponds to a  $SNR$  value of 40 dB was added to the

simulated data. The algorithm identified seven independent components, which were compared against the reference spectra of the five target species. Figure 3.5 shows the correlation results. While sources for nitrate, nitrite, sulfate and carbonate were identified correctly, water was not identified as a unique source; instead the peak of water is present in the source which correlates highly with nitrite. In order to prevent any of the remaining sources (non-targets) to be identified as water in the correlation matrix, a cut-off value of 0.95 was imposed. The intensity of water is low in Raman, which could have contributed to the discrepancy.

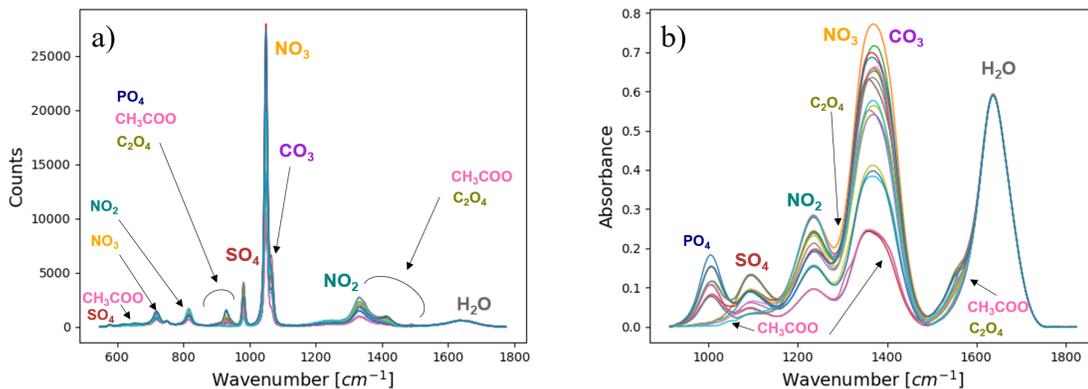


Figure 3.3: Simulated test data set mixtures for a) Raman spectroscopy with  $SNR_{avg} = 40$  dB and b) ATR-FTIR spectroscopy with  $SNR_{avg} = 55$  dB. Concentrations for the 21 mixtures are given in Table 3.1.

The sources from Figure 3.5 that did not correlate with the target reference spectra, were classified as the non-target species. Their contributions to the mixtures in  $\mathbf{X}_{t,s,s}$  were subtracted, as shown in Figure 3.6. The resulting preprocessed spectra for  $\mathbf{X}_{t,s,s}$ , shown in red, only include the signals of the target species.

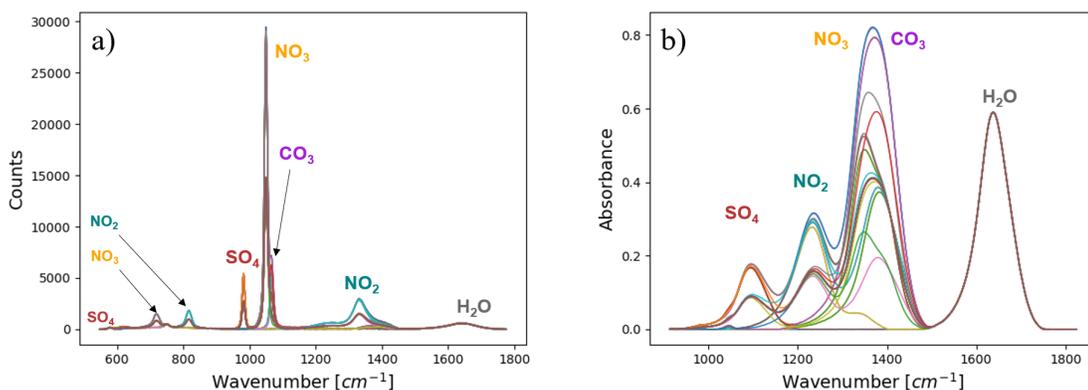


Figure 3.4: Simulated training data set mixtures for a) Raman spectroscopy with  $SNR_{avg} = 40$  dB and b) ATR-FTIR spectroscopy with  $SNR_{avg} = 55$  dB. The concentrations that match each measurement are shown in Table 2.1.

The preprocessed test data for  $\mathbf{X}_{t,s,s}$  were used in Step 4 of the framework, to calculate the concentrations of the target species using a PLSR model. The predicted concentrations for the original and preprocessed data (using the same PLSR model) are compared to the real values in Figure 3.7. The parity plots show that generally, the prediction errors improved for the preprocessed data set, but the outcome of both PLSR and BSS-preprocessed PLSR also depends on the data treatment of the spectra, such as the order of derivation and window length for the Savitzky-Golay filter. A summary of the percent errors for each point in the original and preprocessed data sets is shown in Section A.1, for different values for the order of derivation and Savitzky-Golay smoothing window length.

### *ATR-FTIR Spectroscopy*

Similar simulation analyses were performed for ATR-FTIR, for a  $SNR$  value of 55 dB, based on the level of noise present in experimental IR data. While the IR spectra are less noisy compared to Raman, they feature more significant overlap among the target and non-target species, which complicates the identification of BSS components. For example, the

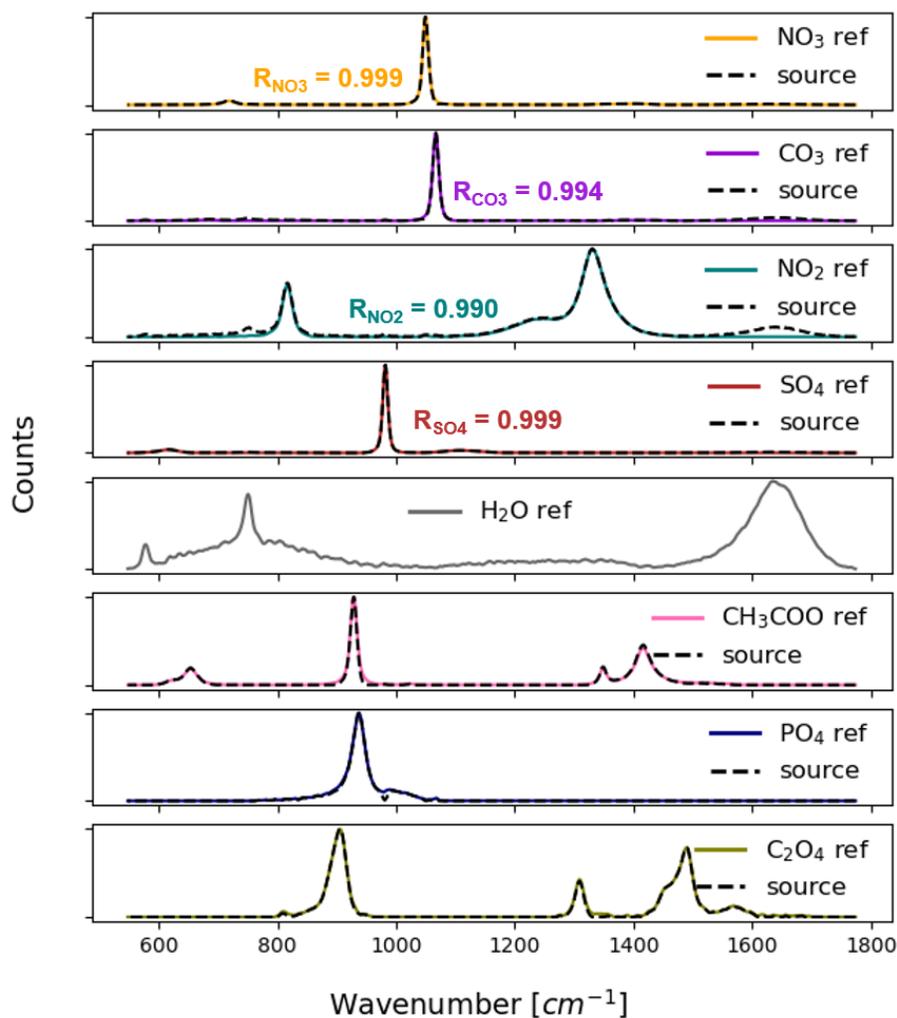


Figure 3.5: Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last three rows show the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed). The BSS algorithm did not identify a unique source for water.

carbonate and nitrate overlap with each other almost completely, but also with acetate, and to a lower extent with oxalate and nitrite. At the same time, the signals of both the target and non-target species are strong, which could help with identification, compared to Raman, where the signals of phosphate, water, oxalate and acetate are weaker and therefore harder

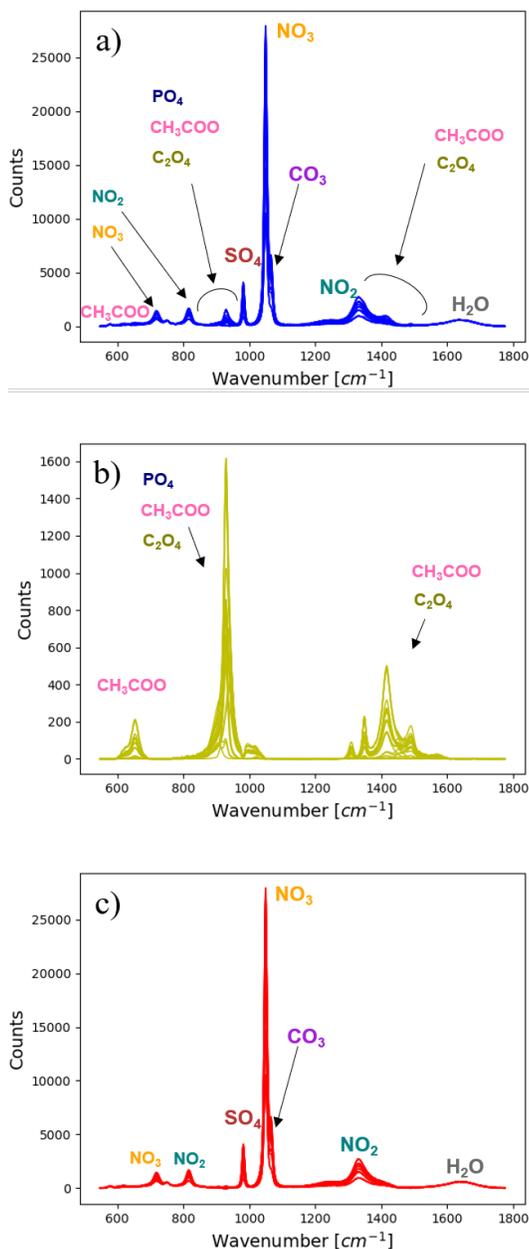


Figure 3.6: Removing the signals of the non-target species for Raman: a) original test set mixture data (blue), containing both target and non-target species, b) signals of the non-target species (yellow) that are subtracted to generate the preprocessed spectra (red) in c). The intensity of the yellow signals is lower relative to the spectra of the targets.

to identify. The test and training data sets are plotted in Figure 3.3(b) and Figure 3.4(b). The preprocessing procedure identified eight independent sources, shown in Figure 3.8. The sources that correlated with nitrate, sulfate and water preserved their shapes completely,

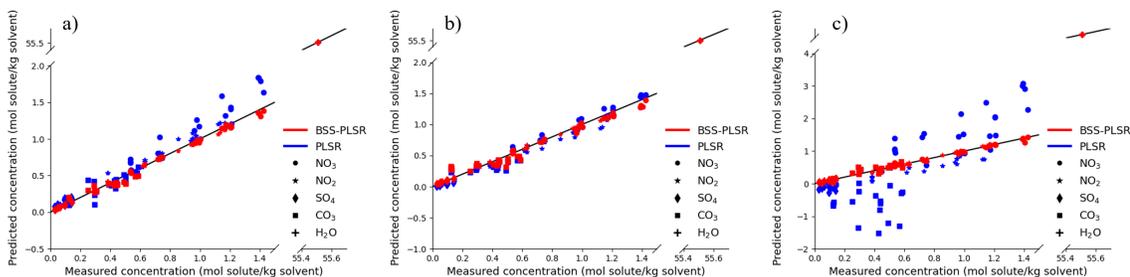


Figure 3.7: Parity plot comparing the Raman predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman: a) 0th derivative spectra; b) 1st derivative spectra using a Savitzky-Golay filter with a second order polynomial and window length size of 7; and c) 1st derivative spectra using a Savitzky-Golay filter with a second order polynomial and window length size of 27.

while carbonate and nitrite show minor deviations. Even with the deviations, the correlation coefficients are relatively high, as shown in Figure 3.8.

Next, the test data were preprocessed to remove the signals associated with the non-target sources, shown in Figure 3.9. The parity plot that compares the predicted concentrations for the original and preprocessed data shows improvements for the preprocessed mixtures (Figure 3.10). A summary of the percent errors for each point in the original and preprocessed data sets is shown in Section A.2.

### 3.3.2 Experimental Data

The simulation study showed that the BSS-PLSR framework can improve the concentration prediction accuracy for simulated data, which adhere to the assumptions in the Beer-Lambert Law (linearity and superimposition). The accuracy of the PLSR model also depends on the data treatment, such as the use of derivative spectroscopy and extent of

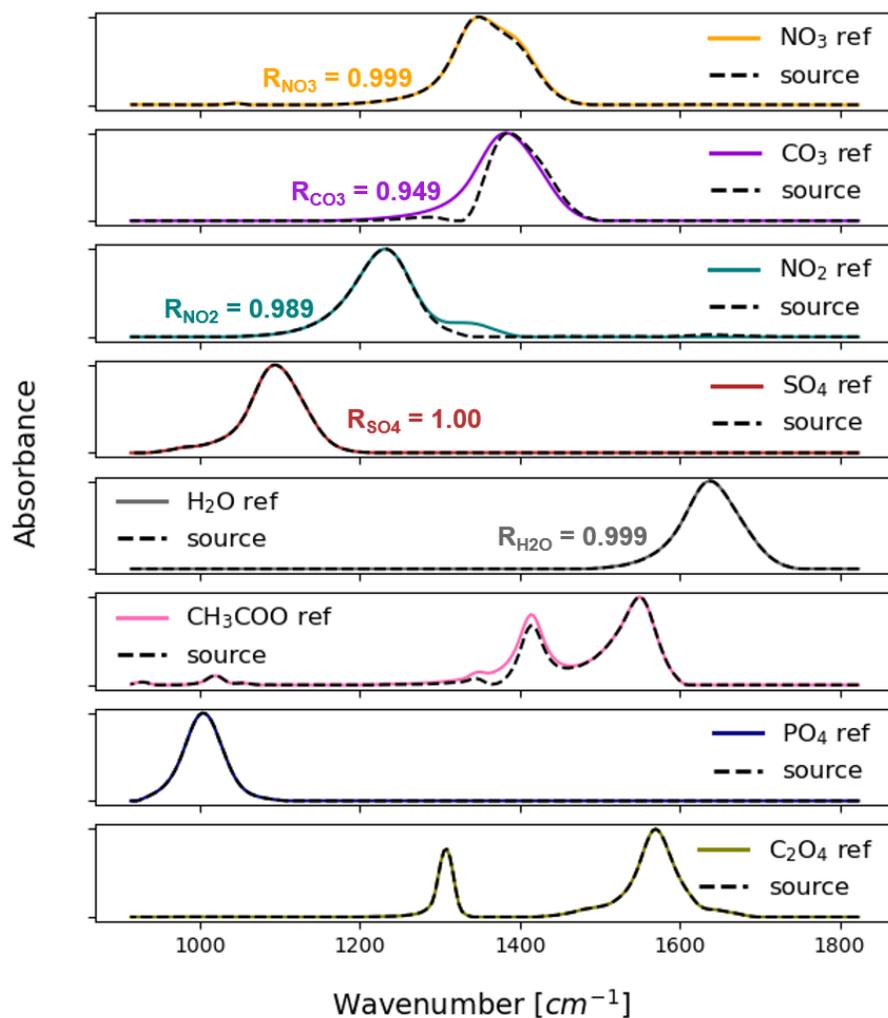


Figure 3.8: Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last three rows show the normalized reference spectra of the non-target species. The normalized IR sources identified by the blind source separation algorithm are plotted in black (dashed).

smoothing. The main challenge in the Raman spectra is the carbonate peak, which appears as a shoulder on the main peak of nitrate, and exhibited higher prediction errors.

Experimental measurements of complex mixtures can be further complicated by baseline changes (due to temperature fluctuations or drifting baseline due to fluorescence in Raman), and nonlinear behaviours due to species interactions, or changes in the ionic strength,

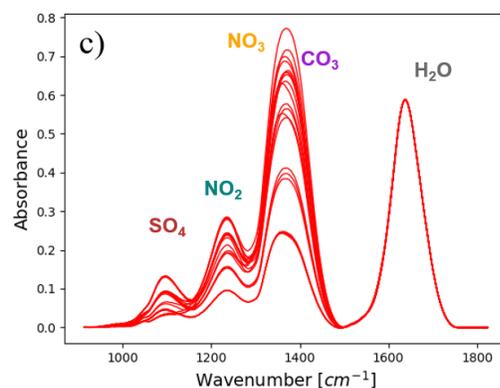
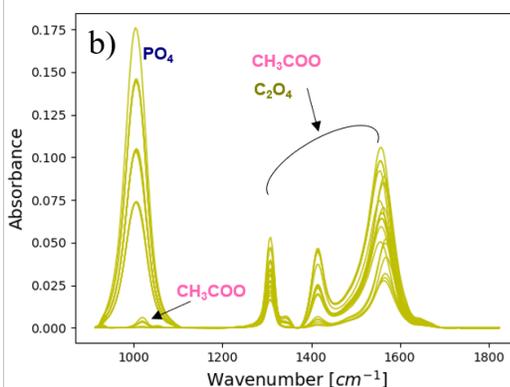
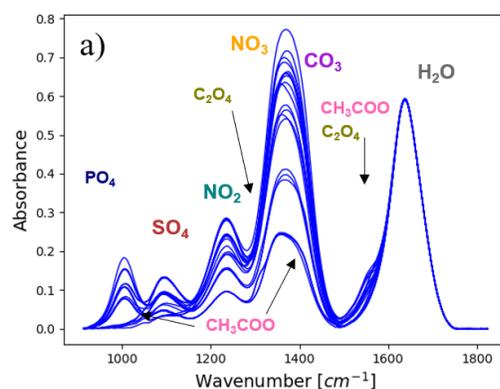


Figure 3.9: Removing the signals of the non-target species for IR: a) original mixture data (blue), containing both target and non-target species, b) signals of the non-target species (yellow) that are subtracted to generate the preprocessed spectra (red) in c). The intensity of the yellow signals is lower relative to the spectra of the targets.

pH, etc. In order to evaluate the applicability of the BSS-PLSR framework to real spectra, the mixtures were measured experimentally. Mixtures matching the concentrations from Table 3.1 were prepared and measured to create the test data set,  $\mathbf{X}_{ts}$ , shown in Figure 3.11.

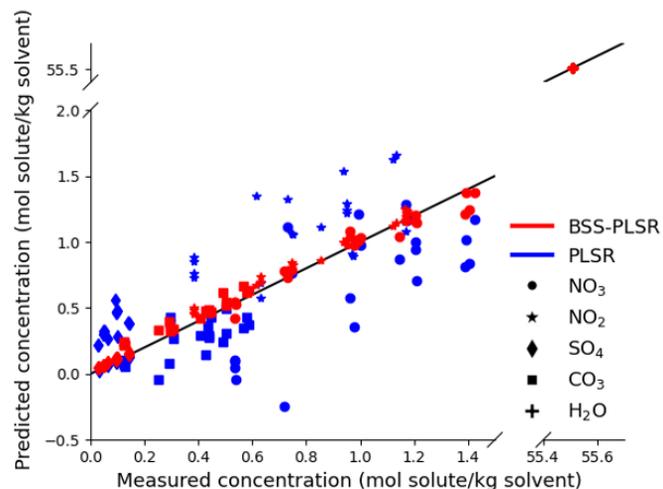


Figure 3.10: Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR.

The training data set,  $X_{tr}$ , shown in Figure 3.12 was created in a similar manner. The data preparation and experimental design are described in more details in Section 3.2.2.

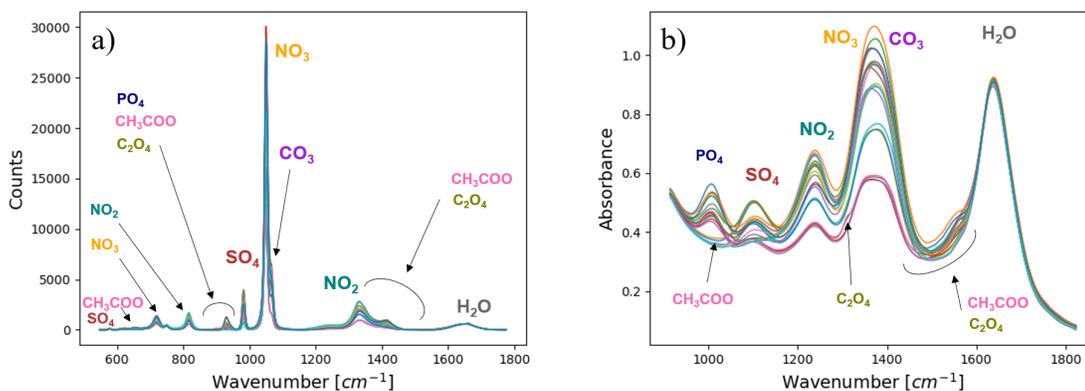


Figure 3.11: Spectral measurements of the test data set mixtures for a) Raman spectroscopy and b) ATR-FTIR spectroscopy. Concentrations that match each measurement are shown in Table 3.1.

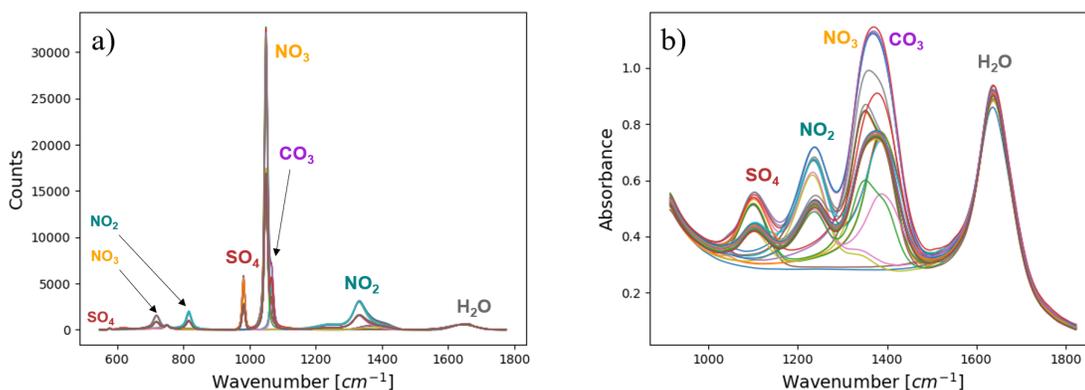


Figure 3.12: Spectral measurements of the training data set mixtures for a) Raman spectroscopy and b) ATR-FTIR spectroscopy. Concentrations for each measurement are shown in Table 2.1.

### *Raman Spectroscopy*

Initially, the BSS-PLSR algorithm identified seven independent sources, which were classified as either target or non-target species according to their correlation coefficients, shown in Figure 3.13. Note that there are eight chemical species present in the mixtures. While sources for nitrate, nitrite, carbonate and sulfate were identified, the water was not identified as an independent component. The non-target sources included an independent source for oxalate, and a lumped source for phosphate and acetate. Finally, the algorithm identified a source associated with a blue shift of the nitrate peak, which will be revisited.

Water is an example of a linearly-dependent species, since it is present in each measurement. The water peak in the 1600-1700  $\text{cm}^{-1}$  region has a low intensity (the maximum height is approximately 1640 when baselined) and the unimolar reference spectrum used in the library data set has an even lower intensity. In order to improve the identification of the water source, the water reference spectrum was amplified until a unique source associated with water appeared (100-fold). The identification of an independent water source also improved the identification of the remaining minor species. After this modification,

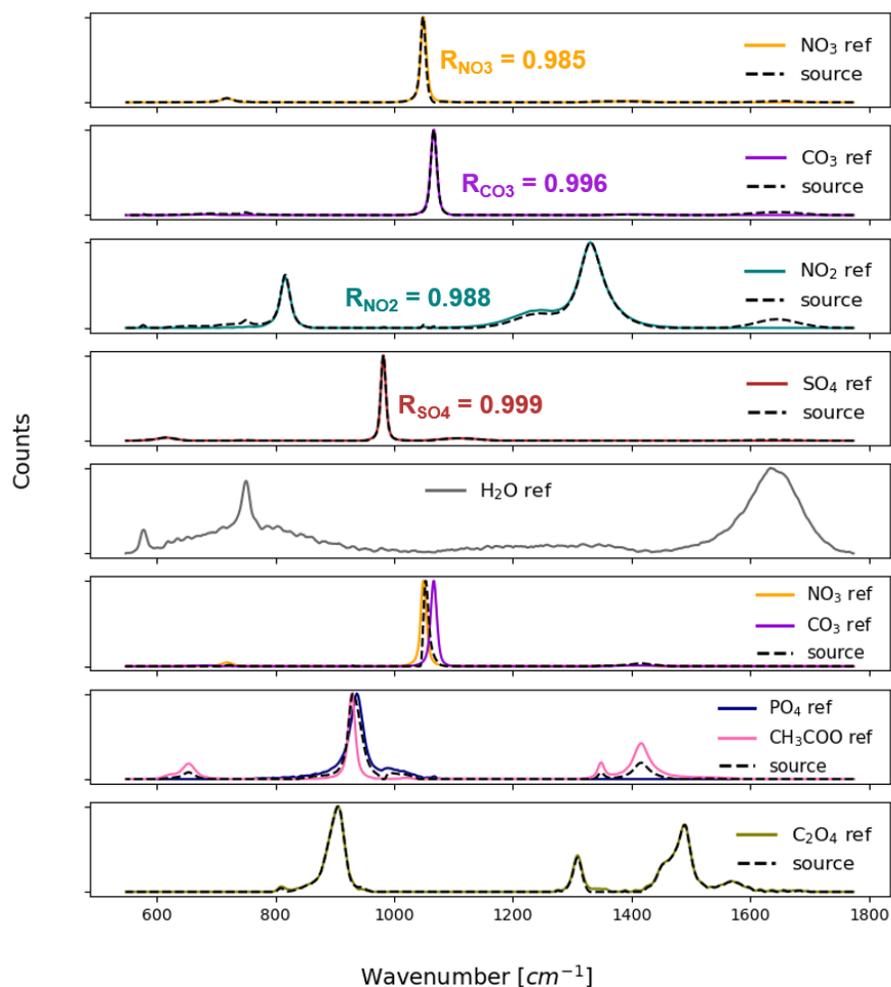


Figure 3.13: Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last two rows show the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed). The BSS algorithm did not identify water as a unique source.

a total of nine source were identified (one for each chemical species plus the additional source associated with a blue shift of nitrate). This modification highlights the importance of introducing linearly-independent spectra into BSS, in order to identify sources that have physical relevance (i.e. correspond to chemical species). The sources which had the highest correlation with the five target species are shown as the first five rows in Figure 3.14. The

remaining four sources are also shown in Figure 3.14 and were classified as the non-target species.

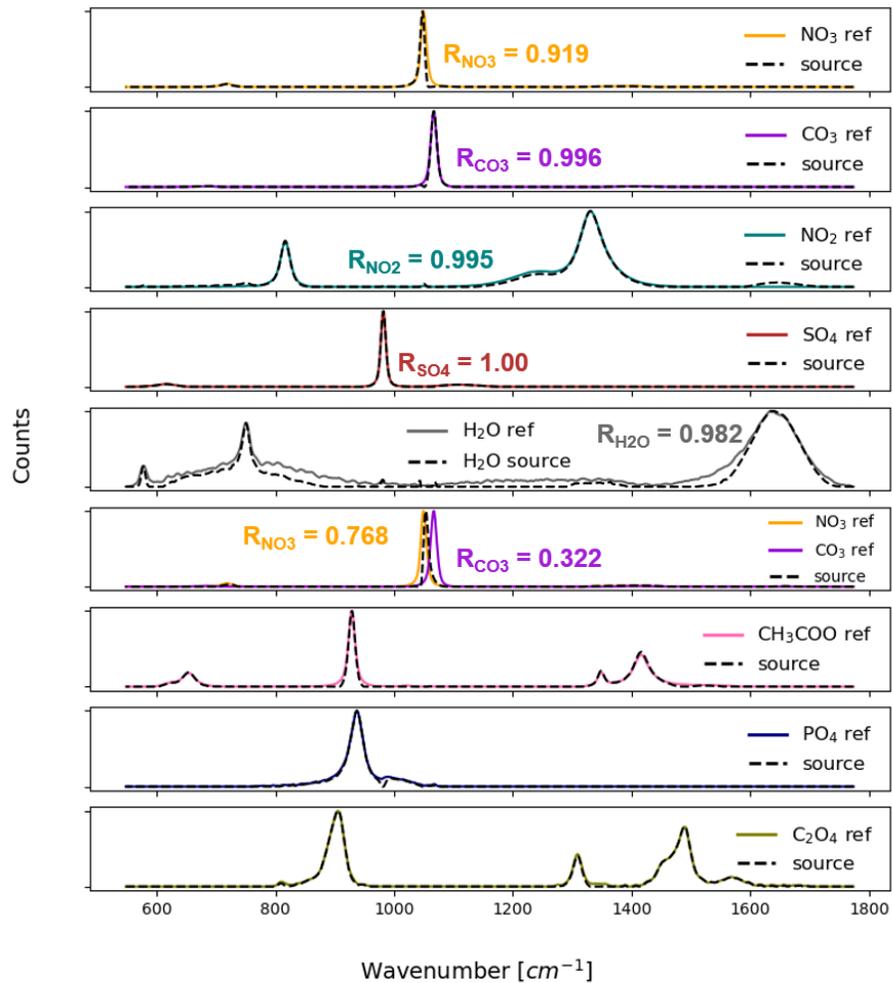


Figure 3.14: Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last three rows show the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed). The "nitrate blue shift" source from row six was reclassified as a target species.

The additional source from row six (counting from the top down) in Figure 3.14 cor-

relates with the reference spectra of nitrate and carbonate. Figure 3.15 shows how the nitrate-carbonate region is affected when the source is classified as a non-target species and removed. Previous work has shown a blue shift of the main Raman peak of nitrate at  $1050\text{ cm}^{-1}$  towards higher wavenumbers and, thus, shifting it closer to the main carbonate peak, in the presence of increasing concentrations of certain anions and cations [51, 52, 33]. A similar source was identified in our previous work, using mixtures of nitrate, nitrite, carbonate, sulfate and phosphate in water [40]. Considering this information, the source was reclassified as an additional target species, by introducing a threshold value of 0.70 for the coefficients of correlation. This approach could also be used in practice, if a source needs to be reclassified. The parity plot of the concentrations before re-classification is shown in Figure 3.16, where its visible that the predicted concentration of nitrate is below the parity line, especially for higher concentrations. The updated preprocessing procedure in Figure 3.17 a) shows that the nitrate-carbonate region is unaltered once the source has been reclassified as a target species. The classification of the nitrate shift source can affect the concentration prediction for nitrate. The predicted concentrations for nitrate increased once the nitrate shift source was reclassified as a target species.

The parity plot in Figure 3.17 b) shows that the preprocessing procedure improved the concentration predictions for nitrate and nitrite compared to the results from the original data. However, the prediction errors for the concentrations for sulfate and carbonate remained higher than the acceptance criteria specification imposed at the Hanford analytical lab ( $\pm 20\%$ ) [11]. The percent errors for each point are summarized in Section A.3. The performance of the partial least-squares model is affected by spectral overlap and species interactions, and the training and test data sets may need to be expanded to account for the complexity of the data. The concentrations in the PLSR model were predicted on first

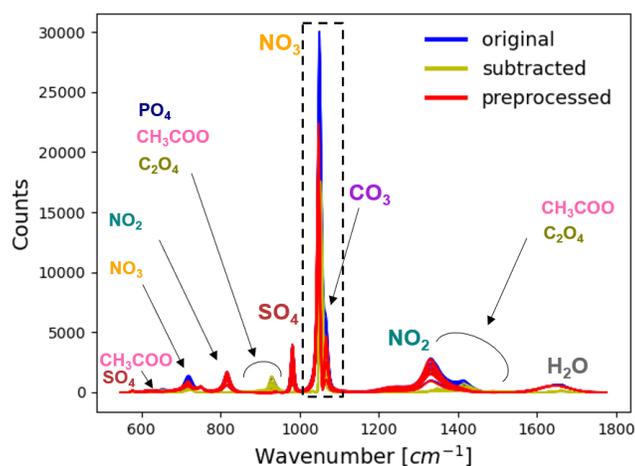


Figure 3.15: Overlay of the original, subtracted and preprocessed spectra for Raman, when the nitrate shift source is classified as a non-target species: the yellow signal in the dashed region has been subtracted during BSS-preprocessing.

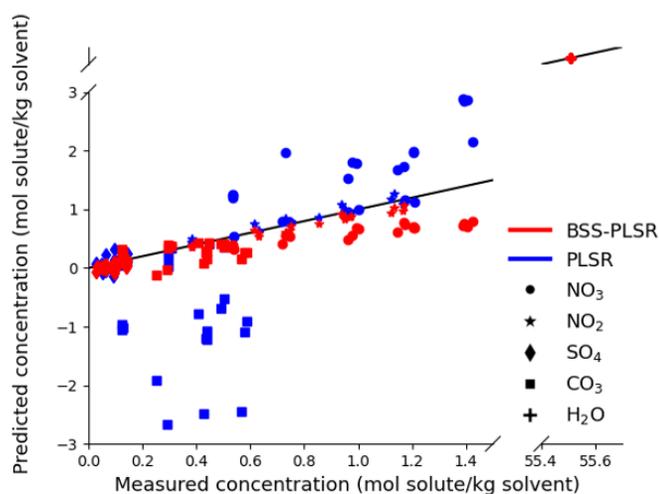


Figure 3.16: Parity plot comparing the Raman predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman experimental data, while the nitrate shift source was classified as a non-target species.

derivative spectra, which should improve the spectral resolution of overlapping peaks. In addition to first derivative spectra, we also performed the analysis for 0th derivative spectra; the concentration predictions are shown in Figure 3.18.

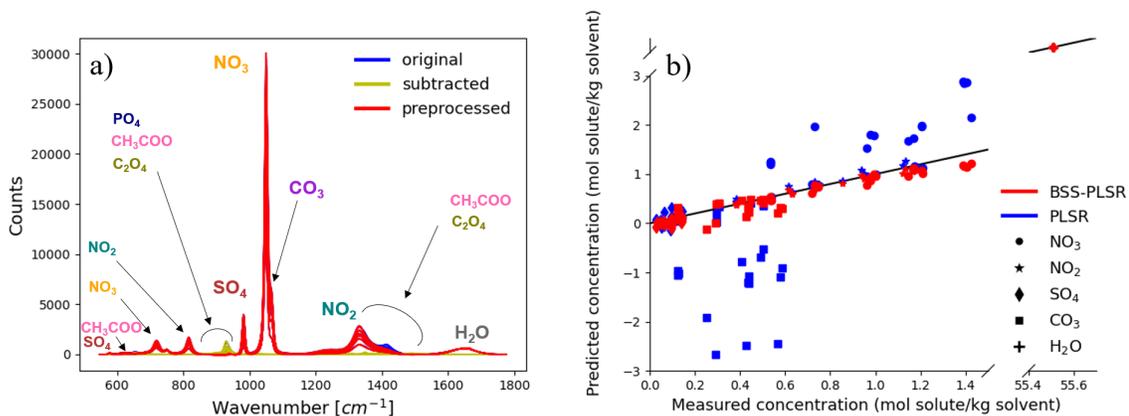


Figure 3.17: a) Removing the signals of the non-target species for Raman. b) Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman.

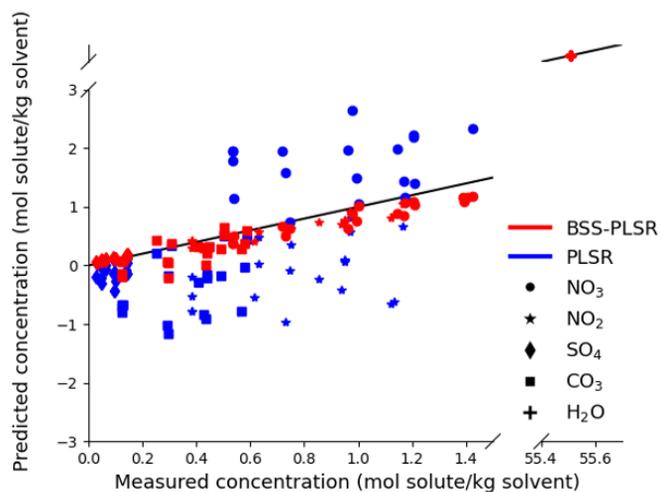


Figure 3.18: Parity plot comparing the Raman predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman experimental data, using 0th derivative spectra.

Finally, assuming that information on the non-target species may not be available, we ran an additional scenario in which only the reference spectra of the target species were

provided. The data input to BSS in Step 1 of the framework consisted of:

$$\mathbf{X}_s = \begin{bmatrix} \mathbf{X}_{ts,s} \\ \mathbf{X}_{tr,s} \\ \mathbf{L}_t \end{bmatrix} \quad (3.7)$$

so that  $\mathbf{L}$  from Equation 3.3 is replaced with a library of the target species only,  $\mathbf{L}_t$ . The correlation data is shown in Figure 3.19. The target sources were identified successfully and with minor discrepancies, including the sources for nitrate, nitrite, sulfate, carbonate, water and nitrate blue shift. Only one source was identified for the three non-target species, mostly resembling acetate. The non-target source also has a peak in the nitrate-carbonate region, and therefore the subtracted signals shown in Figure 3.20 contain some contributions from this region. The results indicate that for the most accurate identification, the BSS-PLSR model should be supplied with reference spectra of all (or at least as many as possible) species expected to be found in the mixtures. Note that even with providing the reference spectra, the information needed to run BSS-PLSR requires less prior information compared to more conventional methods, such as DOE, as explained in Section 3.2.2. Alternatively, the BSS-preprocessing could be updated to identify the contributions associated with the target species, instead of identifying the non-target species contributions to be subtracted (Step 3 of the BSS-PLSR framework). This method could also lead to errors, if any of the target sources have contributions from the non-targets. However, this approach may be more suitable for unknown mixtures with a large number of non-target species and will be revisited.

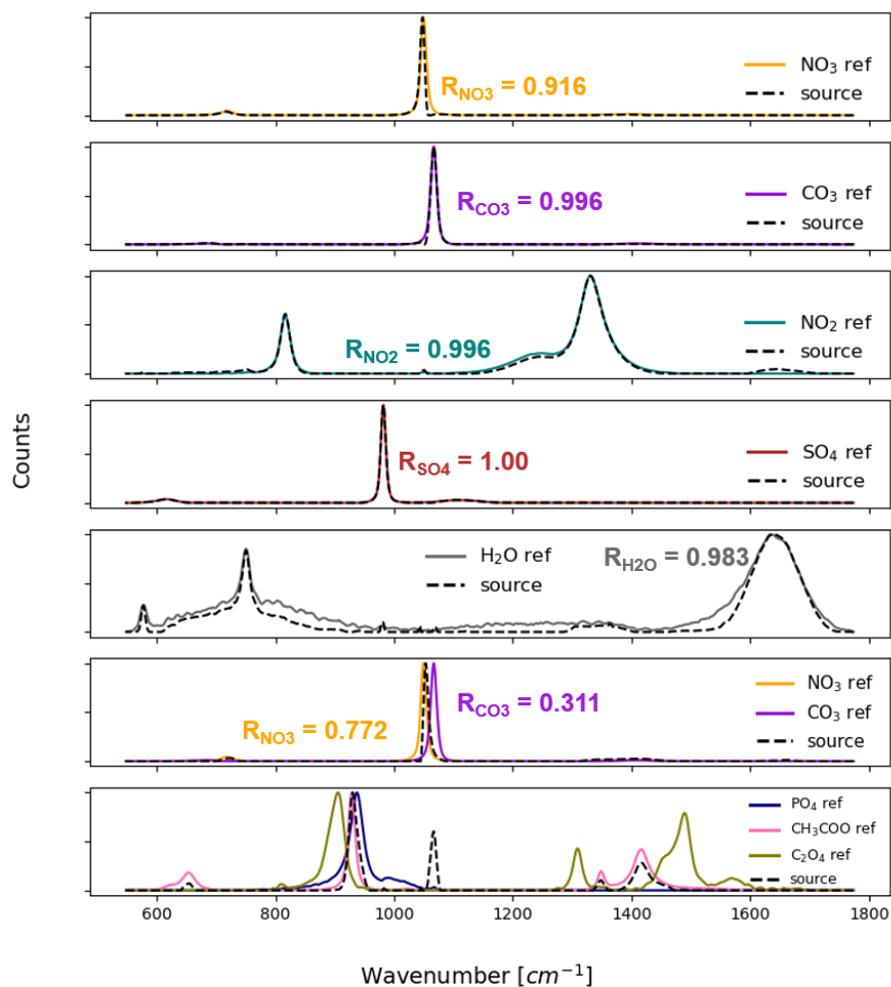


Figure 3.19: Reference spectra-sources correlation using  $L_t$  in the data input for BSS: the first six rows show the normalized reference spectra of the target species, while the last row shows the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed).

### ATR-FTIR Spectroscopy

Similar to the Raman spectroscopy analysis, the BSS-PLSR framework was applied to preprocess the mixture spectra and quantify the concentrations of the target species. Based on trial and error, we determined that providing reference spectra that contain water in the library,  $L$ , in Equation 3.3 provided better estimates of the independent sources in Step 1 of

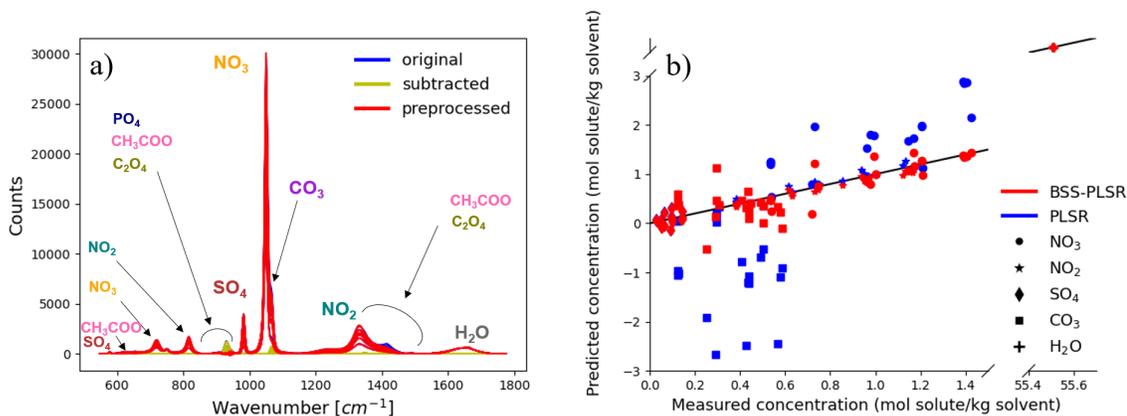


Figure 3.20: a) Removing the signals of the non-target species for Raman, using  $L_t$  in the data input for BSS. b) Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman, using  $L_t$  in the data input for BSS.

the framework. Since water exhibits a high absorbance in IR, coupled with a raised baseline in the region where other chemical species are IR-active, it is possible that removing the water background can introduce error that affects the accuracy of Step 1 of the BSS-PLSR framework. The water activity was not a challenge for collecting Raman spectra since the Raman signal of water is mostly contained in the  $1600 - 1700\text{ cm}^{-1}$  region, so the subtraction of water from the reference spectra in  $L$  does not affect the identification of sources from Raman experimental data. The BSS algorithm identified seven independent components, shown in Figure 3.21. Note that two of the sources correlated highly with the reference spectrum of carbonate (second row from top), one being the source identified as nitrate. Since the coefficient of correlation for the nitrate source was slightly higher (0.915), the second source was not originally identified as a target species by the automated procedure. The discrepancy was addressed by reclassifying the second source as a target species, by introducing a threshold value of 0.70. Other minor discrepancies include residual peaks in the sources for nitrate, nitrite, carbonate and sulfate, including a peak at  $1640\text{ cm}^{-1}$ , associated with water. The IR source of water exhibits a raised baseline in the  $900-1500\text{ cm}^{-1}$  region, which caused the mismatch with the water reference spectrum, whose baseline has been removed. The remaining two sources were classified as the

non-target species. One of the non-target sources resembles phosphate, while the second source resembles acetate. The oxalate was not recognized as a unique source in IR, since the concentrations of oxalate in the test mixtures were low. The oxalate was recognized in Raman, which generally has lower limits of detection.

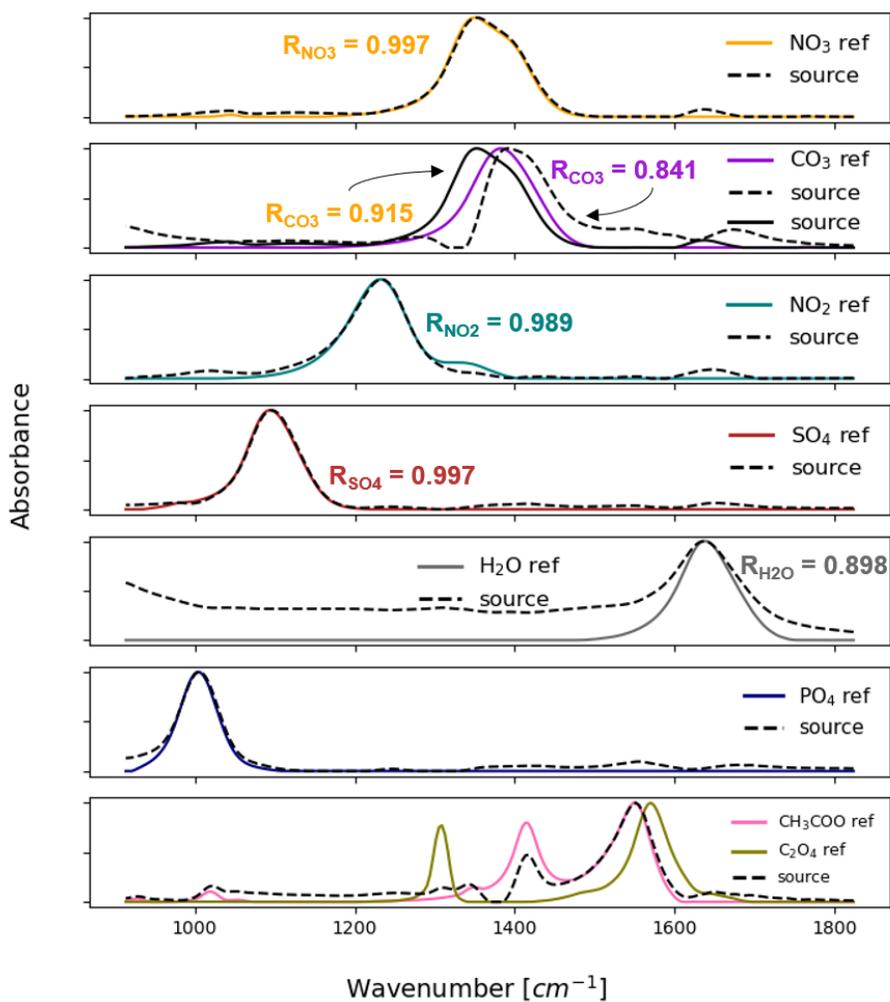


Figure 3.21: Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last two rows show the normalized reference spectra of the non-target species. The normalized IR sources identified by the blind source separation algorithm are plotted in black.

Despite the discrepancies in the non-target sources detection, the BSS-preprocessing improved the concentration predictions, which can be seen in the parity plot in Figure Figure 3.22(b) that compares the PLSR results for the original and preprocessed data. Overall, the average percent errors for the target species are below the  $\pm 20\%$  limits, except for carbonate (23.4%). The percent errors for each point are given in Section 3.2.1 in the Supplementary Information. The ATR-FTIR data set was less noisy, with an average *SNR* value of 57 dB, compared to the Raman data with average *SNR* of 42 dB, and no peak shifting as a result of species interactions was observed experimentally. While the concentrations in the PLSR model were predicted on first derivative spectra, the results for 0th derivative spectra are also shown below (Figure 3.23).

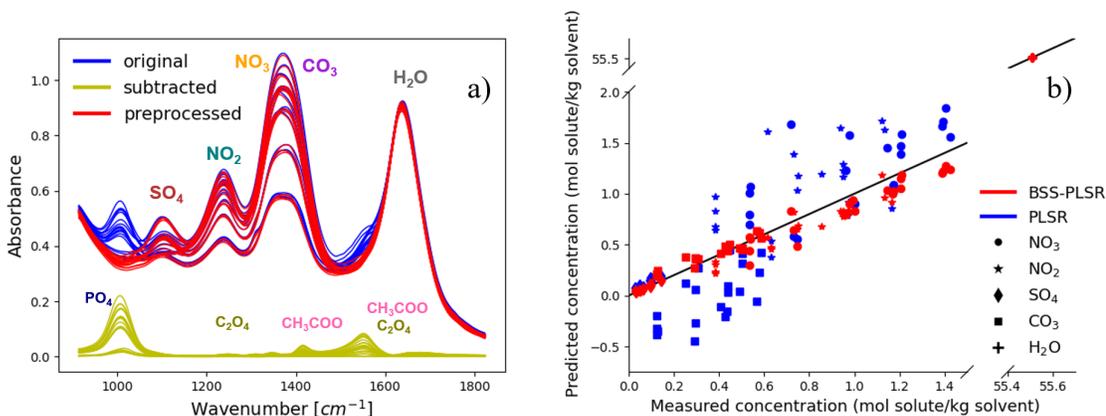


Figure 3.22: a) Removing the signals of the non-target species for IR. b) Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR.

Finally, the importance of including reference spectra of the non-target species was tested by replacing  $L$  with  $L_t$  in the BSS input. Figure 3.24 shows that there are strong correlations between the reference spectra of the target species and the sources classified as targets. The sources classified as non-targets also include residual peaks of the target

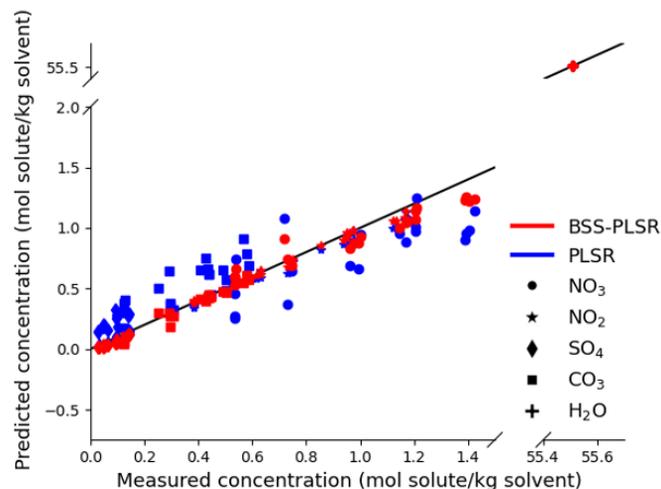


Figure 3.23: Parity plot comparing the IR predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR experimental data, using 0th derivative spectra.

species. Even though the prediction results were not as accurate as those from Figure 3.22, there were improvements following the BSS preprocessing procedure, as shown in Figure 3.25.

### 3.4 Conclusions

This work shows the utility of coupling blind source separation (BSS) with more established chemometric models, such as partial least-squares regression (PLSR), for analyses of multicomponent mixtures. The BSS preprocessing in BSS-PLSR removes the dependence of the PLSR model on the non-target species, thus increasing its robustness. Another feature of the BSS-PLSR framework is the increase in process efficiency and productivity, since the design space of the training set is reduced to the target species only.

Correct identification of the independent sources is the first step in the BSS-PLSR framework and therefore it affects the accuracy of subsequent steps. Should the avail-

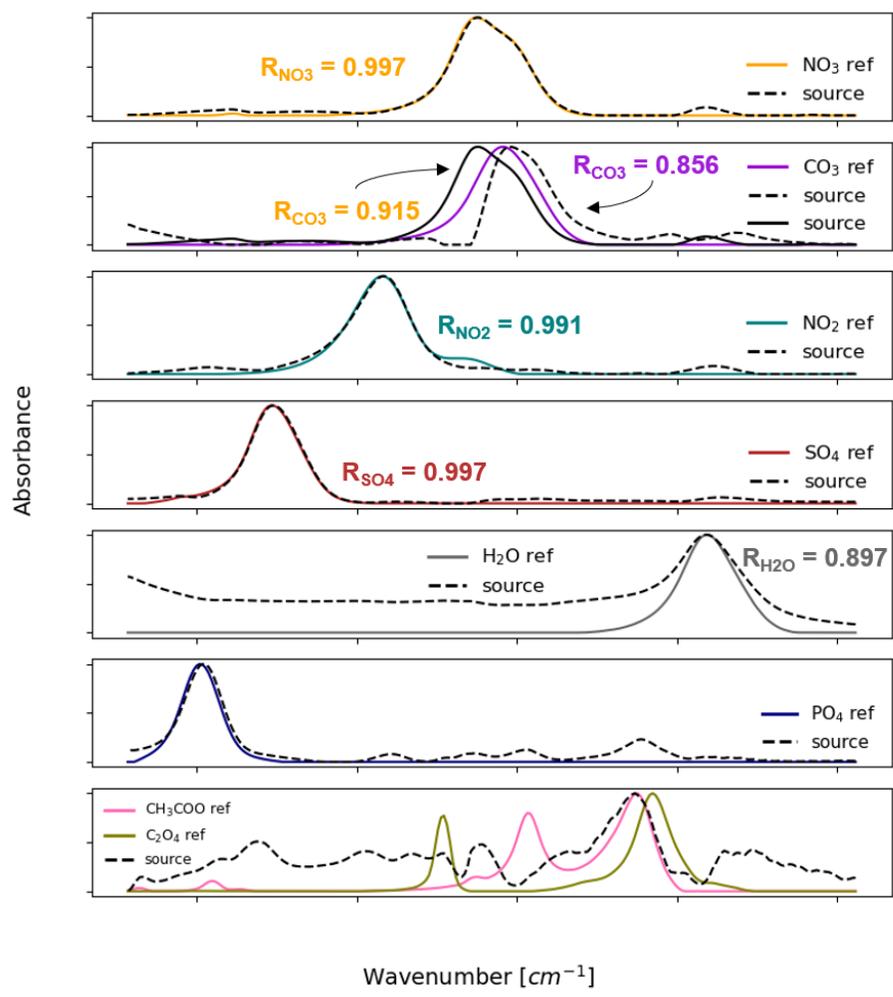


Figure 3.24: Reference spectra-sources correlation using  $L_t$  in the data input for BSS: the first five rows shows the normalized reference spectra of the target species, while the last two rows show the normalized reference spectra of the non-target species. The normalized IR sources identified by the blind source separation algorithm are plotted in black.

able samples from a process have insufficient variation in compositions to facilitate their identification by BSS, as was the case with the DFLAW mixtures in the present study, additional data may be required. Supplementing process samples with mixtures having a range of concentrations from the PLSR training data set and library of reference spectra greatly improved identification of the independent sources by BSS. This step may not be required

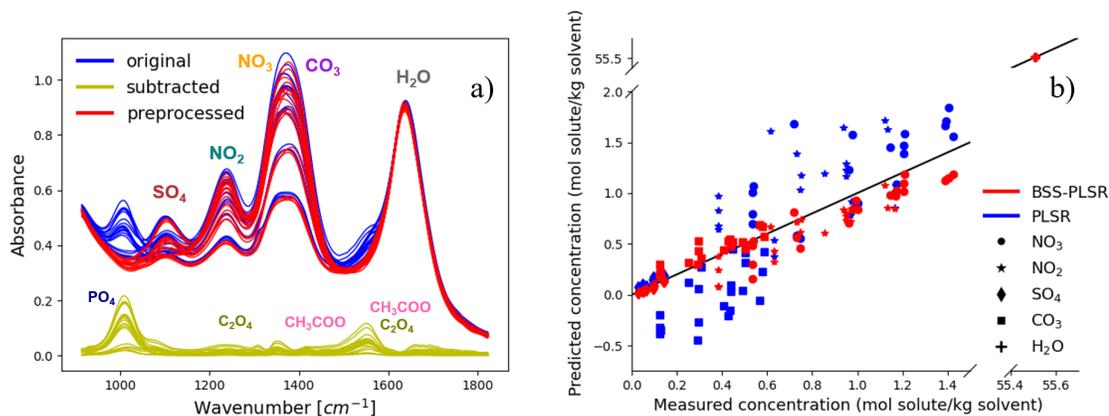


Figure 3.25: a) Removing the signals of the non-target species for IR, using  $L_t$  in the data input for BSS. b) Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR, using  $L_t$  in the data input for BSS.

if the unknown mixture spectra are expected to have variations in their concentrations.

The results show that preprocessing the data with BSS generally improved the PLSR predictions for Raman and IR, despite using relatively small data sets. The analysis of Raman data was challenging due to weaker signals for some of the species (water, phosphate, oxalate and acetate), and deviations from the Beer-Lambert law due to species interactions. The first challenge was partially addressed with using reference spectra in the BSS input, while more research is needed to study the mechanism of species interactions and the extent of deviation from Beer-Lambert law. While the IR data did not exhibit any deviations from linearity, the overlap among target and non-target species was more prominent, and reference spectra were provided in the BSS input. The developed algorithm is not restricted to Raman and mid-IR, but in principle is applicable to any quantitative spectroscopic technique complying with analogous physico-mathematical assumptions, such as the Beer-Lambert law.

**CHAPTER 4**  
**EFFECT OF ION INTERACTIONS ON THE RAMAN SPECTRUM OF**  
**NITRATE: TOWARD MONITORING OF LOW-ACTIVITY NUCLEAR WASTE**  
**AT HANFORD<sup>1</sup>**

## **4.1 Introduction**

As discussed previously, the DFLAW supernatant waste is made of dissolved sodium salts, including nitrate, nitrite, carbonate, sulfate, phosphate, hydroxide, aluminate, oxalate, acetate, fluoride, chloride, and others. The number and concentrations of species that constitute the waste can complicate the development of spectra-to-composition chemometric models. Linear chemometric models rely on three main assumptions: (1) each substance is characterized by a unique spectrum; (2) the spectrum of a multicomponent system is given by the linear superposition of the spectra of the individual components; and (3) a linear dependence relates the intensities of the peaks in a spectrum with the concentrations of the species, following the Beer-Lambert law. A common spectra-to-composition model used extensively in chemometrics is Partial Least Squares (PLS) regression, which models the relationship between the independent variable (concentration) and dependent variable (spectral intensity), maximizing their covariance [27]. A PLS model was used in Chapters 2 and 3 of this thesis.

However, nuclear waste simulants are complex mixtures that may not meet the linearity assumptions of PLS. In particular, the waste mixtures contain both polyatomic species with distinct Raman signals and monoatomic species that are Raman inactive but contribute to the solution concentration. Experiments dating back to the 1960s [54, 55], discussed

---

<sup>1</sup>Adapted from "S. Kocевska, G. M. Maggioni, S. H. Crouse, R. Prasad, R. W. Rousseau, and M. A. Grover, Effect of Ion Interactions on the Raman Spectrum of  $\text{NO}_3^-$ : Toward Monitoring of Low-Activity Nuclear Waste at Hanford, *Chemical Engineering Research and Design*, vol. 181, no. 1, pp. 173–194, 2022." [53]

in more detail in Section 4.2, describe spectral changes in the Raman peaks of oxyanion species, including nitrate, resulting from ion pairing. The sodium,  $\text{Na}^+$ , and nitrate,  $\text{NO}_3^-$ , ions constitute a large portion of the waste, so that monitoring these two species correctly is important for facilitating processing under DFLAW [5, 11, 29]. Considering the importance of nitrate in the vitrification process, this work aims to present experimental data relevant for DFLAW and quantify the change in location and shape of the main peak of nitrate as it relates to the presence of other DFLAW constituents.

The chapter is structured as follows: Section 4.2 provides a theoretical background of the main types of ionic interactions in aqueous solutions and describes how these interactions can be individually identified from the main peak of nitrate to yield a more accurate estimate of the peak shift. Section 4.3 reports the experimental materials and methods and Section 4.4 details the experiments performed in this study. In particular, Section 4.4.1 discusses the quality of the deconvolution model against the data, while Section 4.4.2 analyzes the shift of nitrate spectrum with different ionic species. Finally, Section 4.5 provides a summary of the results and final remarks on the nonlinear phenomena.

## **4.2 Ionic Interactions in Aqueous Solutions**

### 4.2.1 Ion Association Theory

Raman spectroscopy has been widely used for characterization of ion association in electrolyte solutions. The main types of ion association include: solvent sharing ion pairs (such as SIP), contact ion pairs (CIP), and complex ion aggregates (CIA) [56, 57, 52]. In solvent sharing ion pairs (SIP), the anion and cation share a single solvation layer. Loss of the shared solvation layer leads to formation of contact ion pairs (CIP), where the anion and cation are in direct mutual contact surrounded by a common solvation layer. Finally, complex ion aggregates (CIA) can form via further association of contact ion pairs and possibly free ions, resulting in a complex network [56, 52]. The possibility of complex interactions among solute ions becomes greater with increased solution concentrations.

## 4.2.2 Overview of Ion Pairing in Nitrate Solutions

Raman activity depends on the polarizability of the molecule [58]. The sodium ion ( $\text{Na}^+$ ) is Raman inactive; the nitrate ion ( $\text{NO}_3^-$ ), on the other hand, possesses vibration and stretching modes. The modes produce three peaks in the Raman spectrum, shown in Figure 4.1 (a). The most prominent peak is due to the first mode of symmetric stretch, and is referred to as the  $\nu_1$  mode. Located at approximately  $1048 \text{ cm}^{-1}$  in aqueous solutions, the  $\nu_1$  peak is the focus of this work. The other minor peaks are due to an asymmetric stretch and bending vibration modes,  $\nu_3$  and  $\nu_4$ , and are located at  $1400 \text{ cm}^{-1}$  and  $718 \text{ cm}^{-1}$ , respectively [51, 54, 55].

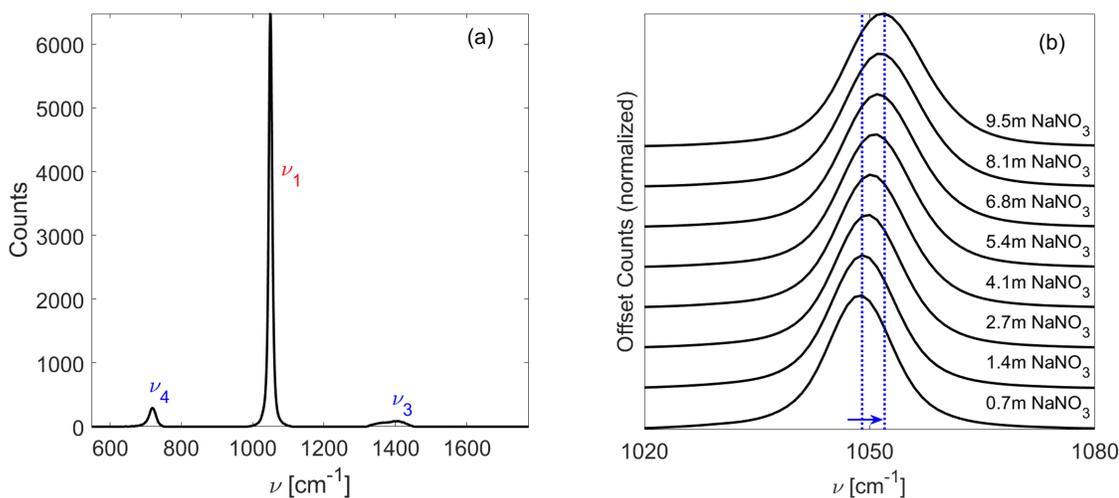


Figure 4.1: (a) The reference Raman spectrum of  $\text{NaNO}_3$  in water at  $1 \text{ mol kg}^{-1}$ . The red label highlights the peak due to the symmetric stretch of the nitrate anion,  $\nu_1$ , investigated in this work. (b) The blue shift of the  $\nu_1$  nitrate peak in aqueous solutions with increasing sodium nitrate concentration (which equals to the ionic strength of the solution for this experiment). The blue dotted lines highlight the shift of the peak maximum from  $1048$  to  $1052 \text{ cm}^{-1}$ . The spectra were normalized by dividing by the maximum intensity and plotted with an offset of  $0.3$  to highlight the peak shift.

One of the initial observations of a shift in the  $\nu_1$  peak, resulting from ion pairing, was reported by Vollmar in 1963, who focused on the effect of several cations on the position

and width of the  $\nu_1$  peak [54]. Subsequently, Miller and Macklin studied the behavior of several oxyanions found in low activity waste at Hanford as a function of sodium content [51]. Frost and James later studied the association equilibrium for unary aqueous solutions of nitrate salts and identified the types of ionic interactions (free aquated nitrate ion, solvent-separated ion pair, contact ion pair and ion aggregates) [57]. They also studied the effect of ionic strength, temperature, and added halides on the association equilibrium for binary aqueous solutions [59]. More recently, Yu *et al.* studied the variations of the  $\nu_1$  peak as a result of  $\text{NaNO}_3$  concentration and temperature changes, identifying the ionic interactions responsible for the peak shifting [52]. Rudolph *et al.* analyzed sodium nitrate solutions in water using Raman and infrared spectroscopy, concluding that the nitrate is fully dissociated at very low concentrations, while ion pair formation occurs at higher concentrations [60]. While previous work with sodium nitrate systems was focused on unary and binary solutions, we study the nitrate ion association in solutions of increasing complexity, culminating with complex nuclear waste simulants that contain two cation species and more than ten anion species.

### 4.2.3 Peak Fitting

Each type of ionic interaction produces a characteristic Raman response, *i.e.* a specific subpeak: it is the linear superposition of these subpeaks that produces the observed peak in a spectrum. The linear superposition is simply the sum:

$$X_c(\nu) = \sum_{j=1}^5 X_s(\nu; \bar{\nu}_j, \omega_j, A_j) \quad (4.1)$$

where  $\nu$  is the wavenumber,  $X_c \in \mathbb{R}^{1 \times n_L}$  is the superposition peak,  $X_s \in \mathbb{R}^{1 \times n_L}$  is the function representing the subpeaks, and  $n_L$  is the number of wavenumbers. Usually, the functions  $X_s$  are modeled by standard distributions, using Gaussian, [59] Lorentzian, [61] or mixed Gaussian-Lorentzian [62] basis functions. The basis functions are uniquely de-

fined by the location and shape parameters,  $\bar{\nu}_j$  and  $\omega_j$ , respectively, up to a multiplicative constant. The constant  $A_j$  in Equation 4.1 represents the area of the subpeak  $j$  and is proportional to the strength of the interaction.

The ionic interactions described in Section 4.2.1 are present in aqueous solutions of nitrate and determine the actual shape of its measured  $\nu_1$  peak [57, 59, 56, 52]. Using Equation 4.1 one can study the spectral shift in detail, since the blue shift (*i.e.*, the shift towards higher wavenumbers) of  $\text{NO}_3^-$  observed in the experiments reflects the changes in the types of interactions between the nitrate and the other species in solution [57, 52]. As an example, the extent of peak shifting for a system consisting of sodium nitrate salt dissolved in water is plotted in Figure 4.1 (b) and discussed in more detail in Section 4.4.2.

To separate the main peak into subpeaks we have opted for a Gaussian model with location and shape parameters presented by Yu *et al.*, shown in Table 4.1 [52]. In addition to the subpeaks for free solvated nitrate anion (SNA), solvent-shared ion pairs (SIP), contact ion pairs (CIP), and contact ion aggregates (CIA), Yu *et al.* used *ab initio* calculations to assign a subpeak (at  $1040 \text{ cm}^{-1}$ ) to the coupling vibrations of water near  $\text{NO}_3^-$  (CPW) [52]. While the model of Yu and co-workers is not the only one reported in the literature [51, 62], it describes the data well and thoroughly characterizes the position and width of the subpeaks using a large, consistent set of experiments [52]. The Gaussian basis function,  $X_s$ , representing a subpeak is:

$$X_s(\nu; \bar{\nu}_j, \omega_j, A_j) = \frac{A_j}{\omega_j \sqrt{\pi/2}} \exp\left(-\frac{2(\nu - \bar{\nu}_j)^2}{\omega_j^2}\right) \quad (4.2)$$

where  $\bar{\nu}_j$  and  $\omega_j$  represent the location (mean) and the shape (standard deviation) of the peak, while  $A_j$  is the peak area. Note that  $\bar{\nu}_j$  increases with the complexity of the interaction, hence the subpeak associated with the coupling vibrations of water near  $\text{NO}_3^-$  is the lowest one, while that associated with complex ion aggregates is the highest one.

The only parameters that require estimation are the areas of the subpeaks,  $\mathbf{A} = [A_1, A_2, A_3, A_4, A_5]$ ,

Table 4.1: Table of parameters of the Gaussian functions used to deconvolve the nitrate peak, from Yu and co-workers [52].

Interaction	$\bar{\nu}$ [cm <sup>-1</sup> ]	$\omega$ [cm <sup>-1</sup> ]
coupling vibrations of water near NO <sub>3</sub> <sup>-</sup> (CPW)	1040	20
free solvated NO <sub>3</sub> <sup>-</sup> (SNA)	1043	5.76
solvent-shared ion pairs (SIP)	1048.5	5.9
contact ion pairs (CIP)	1053.5	8.83
complex ion aggregates (CIA)	1057	17.2

corresponding to CPW, SNA, SIP, CIP and CIA, since they depend on the measured spectra.

$\mathbf{A}$  was estimated by minimizing the following objective function:

$$\Phi(\mathbf{A}) = \left\| \hat{\mathbf{X}} - \sum_{j=1}^5 X_{s,j}(\boldsymbol{\nu}; \bar{\nu}_j, \omega_j, A_j) \right\|_2 \quad (4.3)$$

where  $\mathbf{A}$  is the vector of the areas defined above and  $\hat{\mathbf{X}}$  the discretized measured spectra. The minimization was performed using the *lsqnonlin* function in the optimization toolbox of MATLAB R2021b<sup>©</sup>. The numerical values of  $\mathbf{A}$  can then be used to determine the relative effect of the interactions as well as their dependence on the system conditions.

The fitting quality was assessed by defining an error,  $E_i$ , based on the  $L_1$  norm that is applicable to any individual measured spectrum:

$$E_i = \frac{\left\| \hat{X}_i - X_{ci} \right\|_1}{\left\| \hat{X}_i \right\|_1} \quad (4.4)$$

where  $\hat{X}_i$  is the  $i^{th}$  Raman spectrum measured during an experiment (from Table 4.2) with  $N$  total measurements,  $X_{ci} = X_{ci}(\boldsymbol{\nu})$  is its Gaussian fitting at the  $n$  discretized wavenumbers, and  $E_i$  is its associated error. The definition of  $E_i$  allows it to be interpreted as a generalized relative error between two vectors.

### 4.3 Materials and Methods

Several important features of the LAW to be processed at Hanford are well reproduced by simulant aqueous solutions containing specific proportions of the main salts found in nuclear waste. The mixtures measured for this work are shown in Table 4.2. The mixtures increase in complexity, starting with unary solutions of a nitrate salt in water to examine the effect of the counter-ion ( $\text{Na}^+$ ,  $\text{K}^+$  and  $\text{Al}^{3+}$ ). While  $\text{Al}^{3+}$  was included in the experimental design to study its effect on the nitrate peak, the predominant cations in the nuclear waste at Hanford are limited to  $\text{Na}^+$  and  $\text{K}^+$  [29]. Next, we measure binary and ternary solutions of sodium nitrate and one or two other salts, to examine the effect of a monoatomic anion ( $\text{Cl}^-$ ) and polyatomic anions, such as  $\text{SO}_4^{2-}$ ,  $\text{NO}_2^-$ , and  $\text{CO}_3^{2-}$ . Lastly, we measure low-activity waste simulants. The simulants for Exps. 13 and 14 were prepared in our previous work and include  $\text{NaNO}_3$ ,  $\text{Na}_2\text{SO}_4$ ,  $\text{NaNO}_2$ ,  $\text{Na}_2\text{CO}_3$ ,  $\text{Na}_3\text{PO}_4$ ,  $\text{NaCH}_3\text{COO}$  and  $\text{Na}_2\text{C}_2\text{O}_4$  [33]. The simulants for Exp. 15 were prepared at Savannah River National Laboratory (SRNL) and contain a significant number of waste species ( $\text{Na}^+$ ,  $\text{NO}_3^-$ ,  $\text{NO}_2^-$ ,  $\text{CO}_3^{2-}$ ,  $\text{OH}^-$ ,  $\text{Al}(\text{OH})_4^-$ ,  $\text{K}^+$ ,  $\text{SO}_4^{2-}$ ,  $\text{Cl}^-$ ,  $\text{PO}_4^{3-}$ ,  $\text{CrO}_4^{2-}$ ,  $\text{C}_2\text{O}_4^{2-}$ ,  $\text{F}^-$  and additional total organic carbon (TOC) species (citrate, formate, acetate and ethylenediaminetetraacetic acid or EDTA)) [29].

The experiments were conducted at constant temperature ( $T = 298 \text{ K}$ ) in a 250-mL vessel, stirred at 400 rpm to ensure well-mixed conditions. Raman spectra of the simulant mixtures were measured using a Mettler Toledo ReactRaman system with a 785 nm laser at laser power of 300 mW. The exposure time varied from 0.25 seconds for Exps. 1–12, to 0.75 seconds for Exps. 13–15. Spectral preprocessing consisted of a baseline correction [32], followed by a smoothing filter, according to our previous work [33]. Spectra collected at 0.75 seconds exposure time were also divided by a conversion factor (3) to account for the change in exposure time.

Table 4.2: Table of salt combinations explored in the experiments.

Exp.	NaNO <sub>3</sub> [mol kg <sup>-1</sup> ]	2 <sup>nd</sup> Salt [mol kg <sup>-1</sup> ]	3 <sup>rd</sup> Salt [mol kg <sup>-1</sup> ]
1	[0.68 - 9.47]	-	-
2	-	KNO <sub>3</sub> [0.07 - 3.28]	-
3	-	(AlNO <sub>3</sub> ) <sub>3</sub> [0.13 - 1.20]	-
4	[2.02]	NaCl [0 - 4.78]	-
5	[2.03]	Na <sub>2</sub> SO <sub>4</sub> [0 - 1.79]	-
6	[2.03]	NaNO <sub>2</sub> [0 - 5.73]	-
7	[2.03]	Na <sub>2</sub> CO <sub>3</sub> [0 - 2.57]	-
8	[2.03]	Na <sub>3</sub> PO <sub>4</sub> [0 - 0.27]	-
9	[2.03]	NaNO <sub>2</sub> [0 - 3.42]	Na <sub>2</sub> SO <sub>4</sub> [0 - 1.02]
10	[2.03]	NaNO <sub>2</sub> [0 - 3.43]	Na <sub>2</sub> CO <sub>3</sub> [0 - 1.89]
11	[2.03]	Na <sub>2</sub> SO <sub>4</sub> [0 - 1.67]	Na <sub>2</sub> CO <sub>3</sub> [0 - 1.92]
12	[2.03]	NaNO <sub>2</sub> [0 - 3.43]	NaCl [0 - 3.89]
	NaNO <sub>3</sub> [mol kg <sup>-1</sup> ]	Other Salts [mol kg <sup>-1</sup> ]	
13	[0.74 - 1.50]	Na <sub>2</sub> SO <sub>4</sub> [0 - 0.19], NaNO <sub>2</sub> [0 - 1.27], Na <sub>2</sub> CO <sub>3</sub> [0 - 0.61] [33]	
14	[0.54 - 1.43]	Na <sub>2</sub> SO <sub>4</sub> [0.03 - 0.14], NaNO <sub>2</sub> [0.38 - 1.17], Na <sub>2</sub> CO <sub>3</sub> [0.12 - 0.59], Na <sub>3</sub> PO <sub>4</sub> [0 - 0.12], NaCH <sub>3</sub> COO [0 - 0.19], Na <sub>2</sub> C <sub>2</sub> O <sub>4</sub> [0 - 0.05] [33]	
15	✓	LAW feed simulant components [29]	

#### 4.4 Experimental Results

Presently, to the best of our knowledge, there is no theory-based model capable of predicting all the interactions taking place among the ions in Hanford wastes, and consequently how such interactions affect the measured Raman spectra. We have conducted a series of experiments, using Exp. 1 (increasing concentrations of sodium nitrate in water) as the reference case.

The results are reported in terms of ion concentration, as well as ionic strength,  $I$ , which measures the electrolytic character of a solution and accounts for the type and concentration of all ions in the system:

$$I = \frac{1}{2} \sum_k z_k^2 m_k \quad (4.5)$$

where  $m_k$  and  $z_k$  are the molal concentration and charge number of the ion  $k$ ; the summation is over all ionic species in the system.

#### 4.4.1 Subpeaks Separation

A direct measurement of the blue shift occurring at different conditions is limited by the resolution of the Raman apparatus used in the experiments, *i.e.* by its sampling rate in the wavenumber space. The Raman device used in this work has a spectral resolution of  $6 \text{ cm}^{-1}$ , resulting in wavenumber discretization with a step of  $\Delta\nu = 1 \text{ cm}^{-1}$ . Thus, changes smaller than this value are not resolved, due to the effect of rounding and of noise. Nevertheless, as previously discussed, one can exploit the fact that the  $\nu_1$  peak of nitrate is composed of five subpeaks, as shown in Table 4.1.

In order to determine if the Gaussian model from Section Section 4.2.3 is appropriate for our system, spectra from the unary system of sodium nitrate in water (Exp. 1 in Table Table 4.2), corresponding to a low (0.7 m) and high (9.5 m) ionic strength were fit and the errors were quantified using Equation 4.4. Since the ions are monovalent, the ionic strength equals the concentration of each individual ion. Figure 4.2 provides a visual representation of the fitting quality for the spectra, which are plotted in the range from  $1020$  to  $1080 \text{ cm}^{-1}$  (*i.e.*, where the  $\nu_1$  nitrate peak is located). The errors range from 3–6% and indicate the Gaussian model fit the data well, with a slight disagreement around the peak maximum. Model comparison using the Akaike information criterion (AIC) was performed for the sodium nitrate system and is shown in Section B.2. Additional spectra from four experiments (Exps. 1, 5, 6 and 7 from Table 2) with  $\text{Na}^+$  concentrations of 1.4 m, 2.2 m, 3.2 m and 3.1 m respectively were fit to further test the Gaussian model. The model fits the data well, as shown in Figure 4.3. A slight disagreement can be observed in the rightmost end of the peak (higher wavenumbers), with errors in the 5–7% range (calculated using Equation 4.4). The subpeak areas are summarized in Table 4.3.

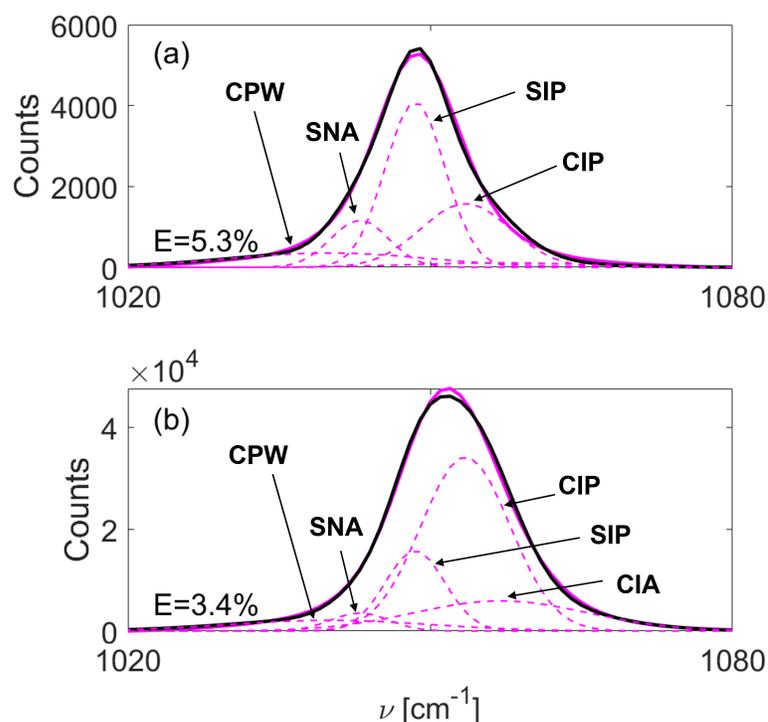


Figure 4.2: The areas of the subpeaks indicate the type of ionic interactions prevailing in each unary system at (a) 0.7 m Na<sup>+</sup> and (b) 9.5 m Na<sup>+</sup>. The measurements (solid black lines) were fit using the Gaussian model from Section 4.2.3 (dashed and solid magenta lines). The errors were calculated using Equation 4.4. The prevailing subpeak changes from SIP to CIP as the concentration increases, corresponding with the formation of contact ion pairs in the presence of greater sodium nitrate concentrations. The CIA subpeak also becomes relevant at high concentration, indicating the presence of complex ion aggregates.

#### 4.4.2 Nitrate Peak Analysis

##### *Influence of Nitrate Counter Ion*

The effect of the counter ion on the  $\nu_1$  nitrate peak shift was studied first, starting with experiments of increasing concentration of NaNO<sub>3</sub>. The system of sodium nitrate in water (Exp. 1) was used as a reference as it was the simplest mixture comprised of both ions of interest (NO<sub>3</sub><sup>-</sup> and Na<sup>+</sup>). The  $\nu_1$  nitrate peak not only shifts, but it also changes shape, as seen in Figure 4.1 (b) and Figure 4.2. These results are in agreement with those of Yu and coworkers [52]. As discussed previously, Figure 4.2 shows the analysis of the subpeaks identified using the Gaussian model. At the lower concentration, the subpeaks for CPW,

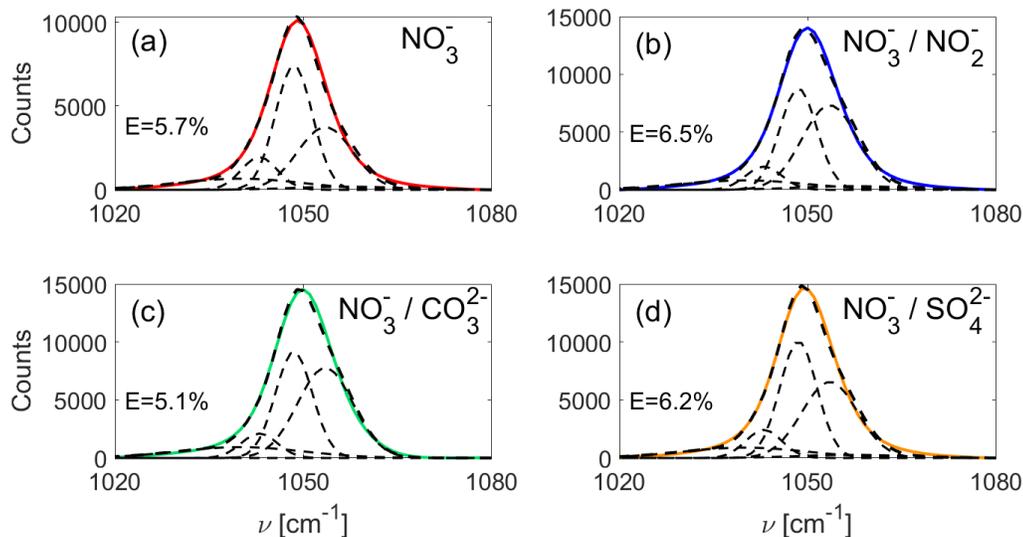


Figure 4.3: The experimentally measured  $\nu_1$  nitrate peaks for (a) unary solution of sodium nitrate (Exp. 1), (b) binary solution of sodium nitrate and nitrite (Exp. 6), (c) binary solution of sodium nitrate and carbonate (Exp. 7) and (d) binary solution of sodium nitrate and sulfate (Exp. 5). The measurements (solid lines) were fit using the Gaussian model from Section 4.2.3 (dashed lines). The errors were calculated using Equation 4.4.

Table 4.3: Table of fitted subpeak areas for spectra in Figure 4.3, using the Gaussian model from Section 4.2.3. The subpeak areas have been normalized by the total area.

Experiment	A <sub>1</sub> (CPW)	A <sub>2</sub> (SNA)	A <sub>3</sub> (SIP)	A <sub>4</sub> (CIP)	A <sub>5</sub> (CIA)
NaNO <sub>3</sub> (Exp. 1)	0.124	0.108	0.421	0.317	0.030
NaNO <sub>3</sub> / NaNO <sub>2</sub> (Exp. 6)	0.107	0.078	0.347	0.433	0.036
NaNO <sub>3</sub> / Na <sub>2</sub> CO <sub>3</sub> (Exp. 7)	0.120	0.079	0.353	0.448	0.000
NaNO <sub>3</sub> / Na <sub>2</sub> SO <sub>4</sub> (Exp. 5)	0.114	0.091	0.387	0.378	0.030

SNA, SIP and CIP are present, with SIP interactions dominating. The subpeak of CIA interactions (located at about  $1057 \text{ cm}^{-1}$ ) is not observable. At the higher concentration, all five subpeaks are present, but the CIP subpeak now dominates the peak, while the SNA and SIP subpeaks have reduced in size. The CIA subpeak is now visible, with a peak width significantly larger than the other subpeaks. Finally, the CPW peak, located at the lowest wavenumber, exhibits a decrease, but is still present even at the high concentration. The changes in subpeak distributions explain the behavior of the overall peak, which exhibits a shift of its maximum and a deformation (see also Figure 4.1 (b)).

While  $\text{Na}^+$  is the most abundant cation in the low-activity supernate waste,  $\text{K}^+$  is also present [29]. In addition, the effect of  $\text{Al}^{3+}$ , a cation with a smaller radius, is studied. Note that the  $\text{Al}^{3+}$  ion is not present in Hanford nuclear waste [63]. Figure 4.4 shows the extent of peak shifting in mixtures of potassium nitrate and aluminium nitrate (Exp. 2 and Exp. 3 in Table 4.2), compared to the sodium nitrate control experiment (Exp. 1). The  $\text{NaNO}_3$  solutions exhibited a significant peak shift, followed by the  $\text{KNO}_3$  system, while the  $(\text{AlNO}_3)_3$  system only exhibited a minor shift. The studies were limited by the solubilities of the salts in water. Vollmar reported an inverse relationship between the size of the atomic radius and extent of hydration [54]. Therefore, larger ions, such as  $\text{Na}^+$  and  $\text{K}^+$  are less hydrated and more susceptible to forming contact ion pairs, resulting in more pronounced blue shift of the  $\nu_1$  nitrate peak. Cations with smaller radii, such as  $\text{Al}^{3+}$ , are more hydrated and less likely to form contact ion pairs. Rudolph studied the solvation shell structure of  $\text{Al}^{3+}$ , concluding that it is strongly hydrated [64].  $\text{Al}^{3+}$  does not induce a significant blue shift of the  $\nu_1$  peak; CIP and CIA interactions, responsible for the observed shift, may be prevented by a stable solvation shell, rendering the  $\text{Al}^{3+}$  ion relatively inert. Figure 4.4 also shows that the ionic strength does not correlate strongly with the extent of the shift, while the concentration of the cation seems to be a better indicator of the shifting behavior of the nitrate peak.

#### *Influence of $\text{Cl}^-$ Monoatomic Anion*

Monoatomic halide anions, such as  $\text{Cl}^-$ , are Raman inactive, hence they do not produce additional peaks in the spectra. Frost and James investigated chloride interactions with nitrate by conducting several experiments (first at constant ionic strength, then at varying concentrations of chloride). They also studied the influence of the halide type by performing experiments with both sodium chloride and iodide, concluding that larger anion size favors the formation of CIP interaction between water and the halides, rather than with

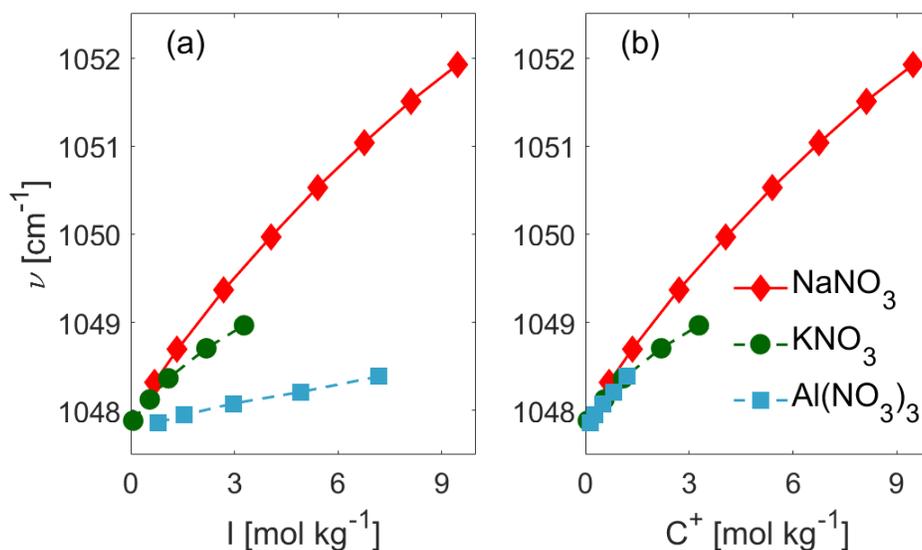


Figure 4.4: The blue shift of the  $\nu_1$  peak in the presence of three different cations as a function of (a) ionic strength and (b) cation concentration.

nitrate [59]. In this work, the effect of  $\text{Cl}^-$  ions was studied by introducing NaCl to a 2 m solution of sodium nitrate in water (Exp. 4 in Table 4.2). The nitrate concentration was kept constant and the blue shift was quantified as a function of the added NaCl salt. The addition of the salt, which contributes to an increase in ionic strength, produced a shift in the nitrate peak, shown in Figure 4.5. The shift was similar to that observed in the control experiment (Exp. 1) initially, but became more prominent in the presence of  $\text{Cl}^-$  ions at greater ionic strengths. This indicates that the anion may also play a role in determining the extent of the shift, especially at higher concentrations, *i.e.* when CIP and CIA are more likely to occur. Figure 4.6 shows the relative areas ( $x_{A,j} = A_j / \sum_j A_j$ ) of each subpeak as a function of  $I$ . The difference between the two systems is visible here too: while both overlap initially, the SIP decreases more significantly at higher concentrations when NaCl is present, while the CIP and CIA increase.

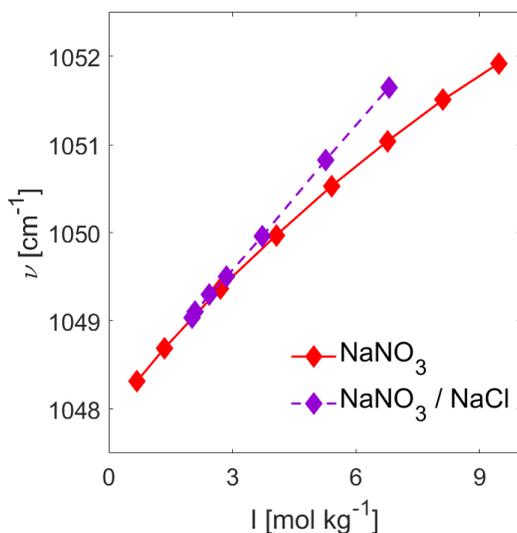


Figure 4.5: The blue shift of the main peak of the nitrate anion as a function of ionic strength,  $I$ , which is equal to cation and anion concentration in these experiments.

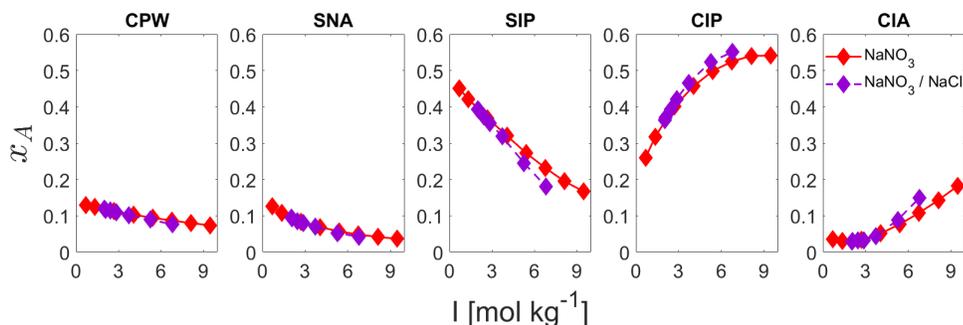


Figure 4.6: The relative fractions,  $x_A$ , of the subpeak areas associated with each type of ionic interaction in Exps. 1 and 4.

### *Influence of Polyatomic Oxyanions*

Four sodium salts are of most interest for Hanford low activity waste:  $\text{Na}_2\text{SO}_4$  due to potential salt accumulation during vitrification [16],  $\text{NaNO}_2$  and  $\text{NaNO}_3$  due to their abundance, and  $\text{Na}_2\text{CO}_3$  due to its abundance and its spectroscopic proximity to the nitrate peak [65, 11, 29, 33]. Additional polyatomic species present in the waste include phosphate, oxalate, aluminate, acetate, chromate, etc [65]. In this section, binary systems with 2 m nitrate plus another sodium salt (nitrite, carbonate, sulfate and phosphate) were studied initially, followed by ternary systems (sodium nitrate and two other salts). Lastly, Exps.

13–15 simulate more complex mixtures by incorporating many of the major and minor salt components that are present in low-activity waste.

Figure 4.7 shows the position of the peak maximum as a function of the total ionic strength,  $I$ , the concentration of anions,  $C^-$ , and the concentration of cations,  $C^+$ , for the binary systems. The shift is present in each binary system; however, the extent of peak shift is difficult to predict based on measurements of the ionic strength and anion concentration. The cation concentration in each system seems to be an important factor, indicating that the interactions between the sodium and nitrate are strongly affecting the extent of peak shifting.

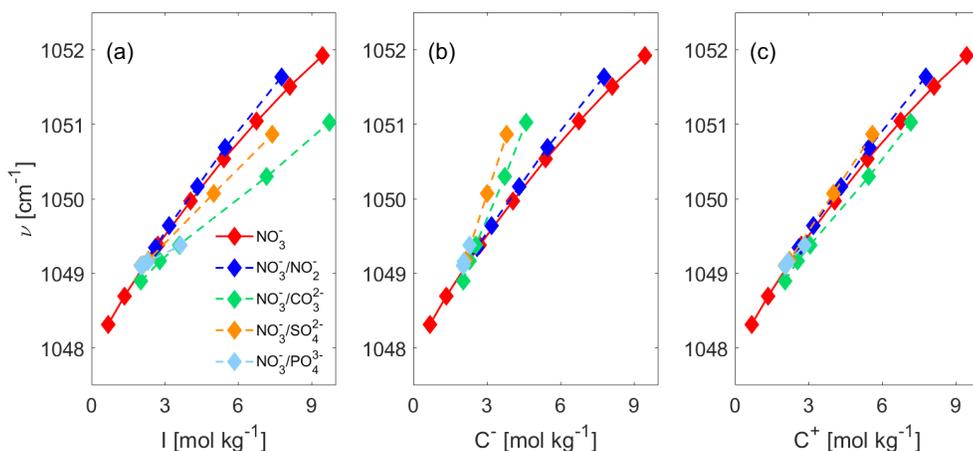


Figure 4.7: The extent of blue shift in the binary systems plotted against (a) the ionic strength,  $I$ , (b) the concentration of anions,  $C^-$ , and (c) the concentration of the cation,  $C^+$ . The sodium nitrate system is included in the comparison, as a reference.

Figure 4.8 shows the relative areas  $x_{A,j}$  for each type of interaction, given as a function of the cation concentration of the solutions. The blue shift of nitrate is due to a reduction of the areas of the subpeaks related to simpler interactions (CPW, SNA, and SIP), while those of the more complex interactions (CIP and CIA) increase.

When comparing the binary systems, minor differences related to SIP, CIP, and CIA interactions can be observed, except for the carbonate system which shows more signifi-

cant deviations in the CIP and CIA plots. The CIA interactions for this system seem to be negligible, except for the last point, while the CIP interactions are more significant compared to the other experiments for the same cation concentration. The explanation for this behavior relates to the main Raman peak of carbonate, which overlaps with that of nitrate [66]. In order to study the blue shifting of nitrate, the spectra were preprocessed to estimate and subtract the carbonate peak's contribution to the region between 1020 and 1080  $\text{cm}^{-1}$ . The preprocessing was performed using reference spectra of the five nitrate subpeaks and carbonate combined with blind source separation techniques, which were developed in our previous work [33]. An example of the subtraction is shown in Figure 4.9. Due to the overlap between nitrate and carbonate, especially with the rightmost subpeak (CIA interactions), the carbonate may interfere with the estimates of the total nitrate and its distribution in Figure 4.8.

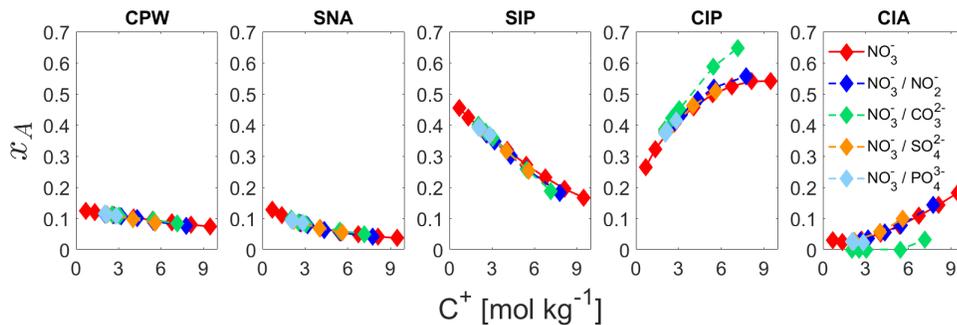


Figure 4.8: The relative areas of the subpeaks,  $x_A$ , indicate the type of ionic interactions prevailing in each binary system.

Finally, we also looked at variations of the total peak area and maximum intensity in the binary systems, which may not be captured by the relative variations of the subpeaks. Figure 4.10 shows the variation of the nitrate peak area and intensity. Overall, each binary system exhibits diminishing values for the areas and heights, with those of the nitrate-nitrite

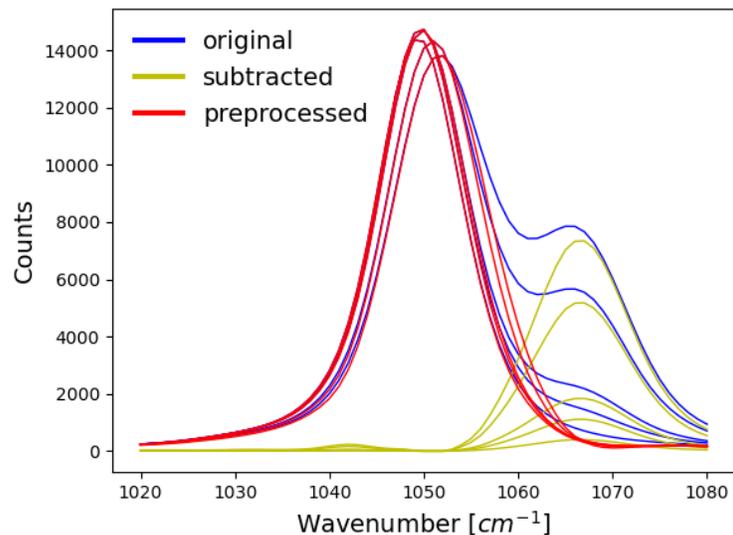


Figure 4.9: Subtraction of the carbonate contribution using blind source separation [33].

system showing the greatest reductions. Considering that the overall subpeak trend involves increase in CIP and CIA interactions at the expense of SIP interactions, the shifting may be accompanied by a decrease in overall peak intensity due to the variations in Raman activity of the different subpeaks. The behavior of the reference system (sodium nitrate in water) also showed a drop in the peak area and intensity, when scaled with its concentration to account for the salt additions to the solution (see Figure 4.11). The decrease in both the area and height indicates that the system behaves nonlinearly. As mentioned previously, the nonlinear behavior associated with a blue shift and drop in intensity may lead to inaccurate estimates of the composition of the waste using linear spectra-to-composition chemometric models.

Next, the blue shift in ternary systems consisting of sodium nitrate plus two other salts was studied (Exps. 9 to 12 in Table 4.2). The nitrate-nitrite pair was studied in three of the experiments, since previous work indicated a correlation between the nitrite addition and

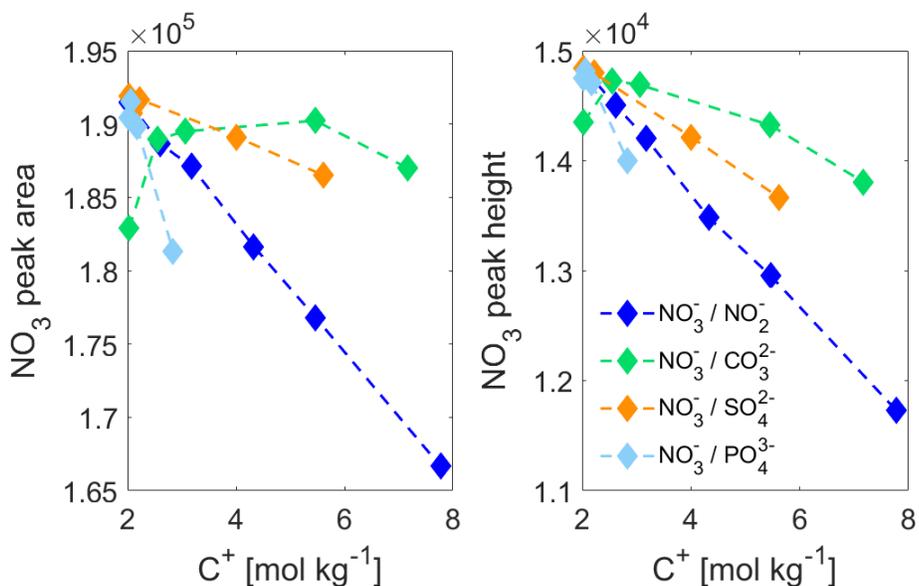


Figure 4.10: Change in (left) total area of the  $\nu_1$  peak of nitrate and (right) maximum height of the nitrate peak as a function of the cation concentration.

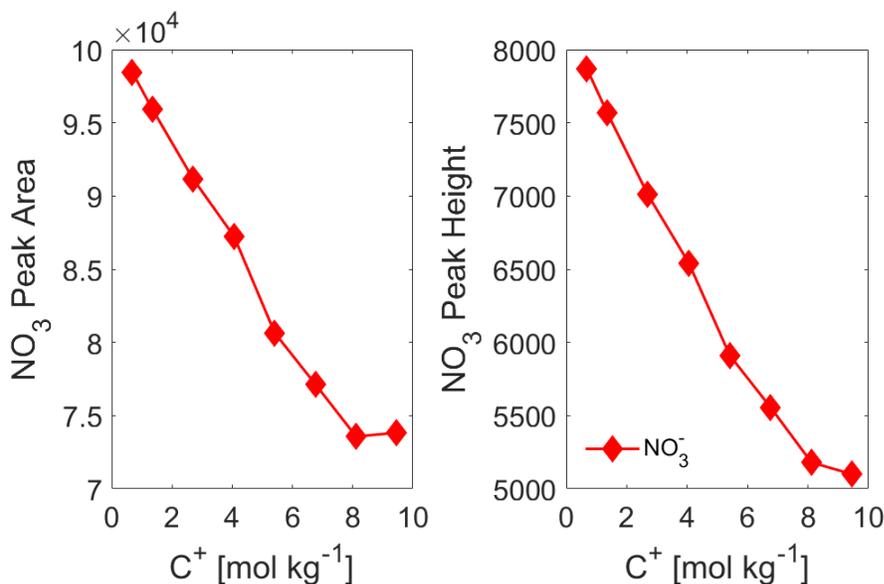


Figure 4.11: Change in (left) total area of the  $\nu_1$  peak of nitrate and (right) maximum height of the nitrate peak as a function of the cation concentration. The values were scaled by the concentration of nitrate at each point, to account for the nitrate salt additions to the solution.

nitrate peak shift [40]. Figure 4.12 shows the extent of the nitrate peak shift. Similarly to the binary systems, the cation concentration provided the best estimate for the extent of peak shifting in all systems. For Exp. 11 (sodium nitrate, sulfate and carbonate), the nitrate shifting behavior was accompanied by a decrease in intensity for not only the nitrate but

also for the sulfate peak. The subpeaks for the ternary system are plotted in Figure 4.13.

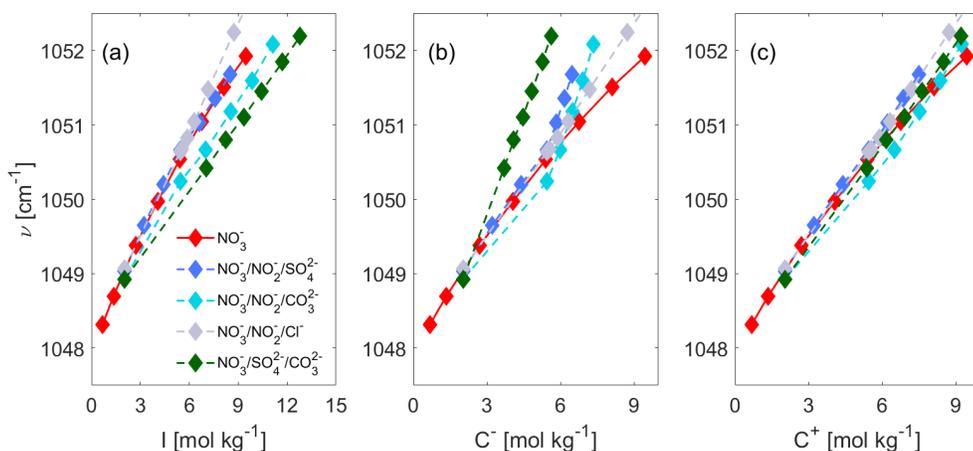


Figure 4.12: The extent of blue shift in the ternary systems plotted against (a) the ionic strength,  $I$ , (b) the concentration of anions,  $C^-$ , and (c) the concentration of the cation,  $C^+$ . The sodium nitrate system is included in the comparison, as a reference.

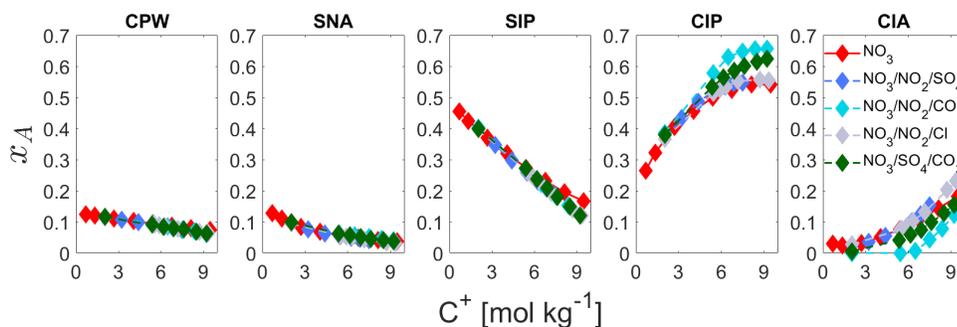


Figure 4.13: The relative areas of the subpeaks,  $x_A$ , indicate the type of ionic interactions prevailing in each ternary system.

Lastly, we analyzed the LAW simulants from our previous work, which included a training data set for a blind source separation - partial least squares regression model consisting of  $\text{NaNO}_3$ ,  $\text{Na}_2\text{SO}_4$ ,  $\text{NaNO}_2$  and  $\text{Na}_2\text{CO}_3$  (Exp. 13), and a test data set for the same model consisting of  $\text{NaNO}_3$ ,  $\text{Na}_2\text{SO}_4$ ,  $\text{NaNO}_2$ ,  $\text{Na}_2\text{CO}_3$ ,  $\text{Na}_3\text{PO}_4$ ,  $\text{NaCH}_3\text{COO}$  and  $\text{Na}_2\text{C}_2\text{O}_4$

(Exp. 14) [33]. In our previous work, shown in Chapter 3, the blind source separation model identified an additional peak, which correlated with a reference spectrum of nitrate, but also exhibited a blue shift [33, 67, 40]. Figure 4.14 shows that a nitrate shift does occur in the LAW training and test simulants. The shift is not as pronounced as that in the sodium nitrate reference system in Exp. 1 (the simulant systems also do not reach the same levels of ionic strength and concentration as the reference system). Similarly to the binary and ternary experiments, the cation concentration seems to predict the extent of peak shifting in the system; however as the concentration increases, the extent of shifting in the complex solutions becomes slightly less pronounced than that in the reference case.

We also analyzed the low-activity waste feed simulants prepared by Savannah River National Laboratory (SRNL) that contain many species ( $\text{Na}^+$ ,  $\text{NO}_3^-$ ,  $\text{NO}_2^-$ ,  $\text{CO}_3^{2-}$ ,  $\text{OH}^-$ ,  $\text{Al}(\text{OH})_4^-$ ,  $\text{K}^+$ ,  $\text{SO}_4^{2-}$ ,  $\text{Cl}^-$ ,  $\text{PO}_4^{3-}$ ,  $\text{CrO}_4^{2-}$ ,  $\text{C}_2\text{O}_4^{2-}$ ,  $\text{F}^-$  and total organic carbon (TOC) species) [29]. The simulants provide the best estimate for the waste at Hanford, with both the chemical (number and concentration of species) and physical properties (density and high pH) matching those of the waste. The extent of the nitrate peak shift for the simulants is also shown in Figure 4.14. The ion concentrations and ionic strength of the solutions were estimated using inductively coupled plasma (ICP) measurements provided by SRNL for some of the species and values from gravimetric analysis for the remaining species [29]. Figure 4.14 shows that while peak shifting can be observed for the low-activity waste samples, the extent of shifting as a function of the cation concentration is not as high as the shift in the reference case. Note that the low-activity waste contains a small amount of  $\text{K}^+$  ions in addition to the large amount of  $\text{Na}^+$  ions [29]. The subpeak areas associated with Figure 4.14 are shown in Figure 4.15.

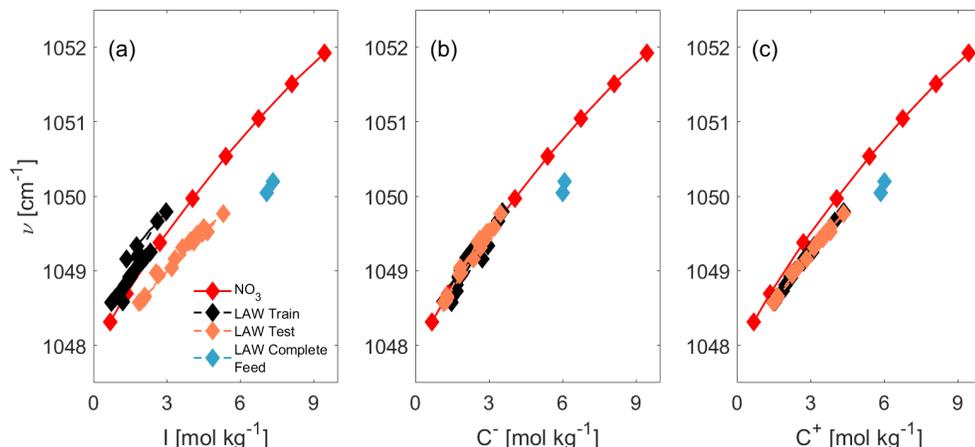


Figure 4.14: The extent of blue shift in the simulants, including LAW train ( $\text{NaNO}_3$ ,  $\text{Na}_2\text{SO}_4$ ,  $\text{NaNO}_2$  and  $\text{Na}_2\text{CO}_3$ ), LAW test ( $\text{NaNO}_3$ ,  $\text{Na}_2\text{SO}_4$ ,  $\text{NaNO}_2$ ,  $\text{Na}_2\text{CO}_3$ ,  $\text{Na}_3\text{PO}_4$ ,  $\text{NaCH}_3\text{COO}$  and  $\text{Na}_2\text{C}_2\text{O}_4$ ) and LAW complete feed ( $\text{Na}^+$ ,  $\text{NO}_3^-$ ,  $\text{NO}_2^-$ ,  $\text{CO}_3^{2-}$ ,  $\text{OH}^-$ ,  $\text{Al}(\text{OH})_4^-$ ,  $\text{K}^+$ ,  $\text{SO}_4^{2-}$ ,  $\text{Cl}^-$ ,  $\text{PO}_4^{3-}$ ,  $\text{CrO}_4^{2-}$ ,  $\text{C}_2\text{O}_4^{2-}$ ,  $\text{F}^-$  and total organic carbon (TOC) species at  $pH > 13$ ) plotted against (a) the ionic strength,  $I$ , (b) the concentration of anions,  $C^-$ , and (c) the concentration of the cation,  $C^+$ . The sodium nitrate system is included in the comparison, as a reference.

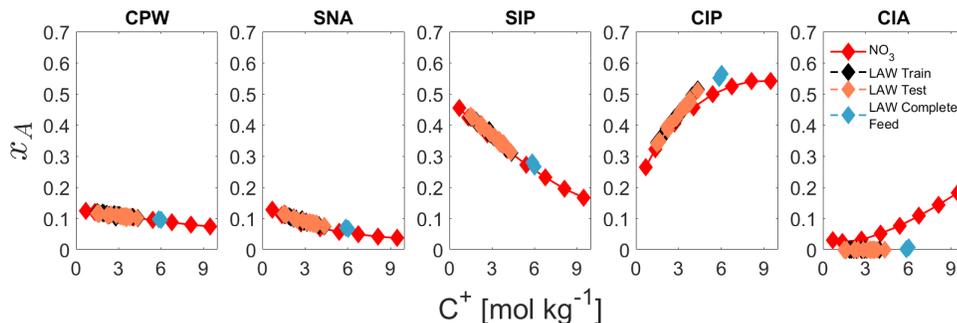


Figure 4.15: The relative areas of the subpeaks,  $x_A$ , indicate the type of ionic interactions prevailing in each LAW system.

## 4.5 Conclusions

Quantification of the composition of multicomponent mixtures is of great importance for the stabilization of low-activity radioactive waste at Hanford, and for other processes that involve complex salt mixtures, such as desalination, mining, and wastewater treatment. Raman spectroscopy is an established technique that can be used to measure the composition of the waste in real time. However, the development of spectra-to-composition models

depends on the complexity of the waste.

In this work, we have studied the phenomena of blue shift of the  $\nu_1$  Raman peak of nitrate in salt solutions of increasing complexity. The extent of peak shift was related to the ion association and subpeaks were assigned for each ion interaction.

Each system we examined exhibited a shift in the  $\nu_1$  peak of nitrate. The shift occurred with increases in nitrate concentration in solutions containing only that species, and the behavior was also observed in mixtures simulating the liquid low-activity waste at Hanford. For all solutions, the cation concentration was found to be a good indicator of the extent of the blue shift, while the shift did not always show a singular dependence on ionic strength.

The observed shift in the  $\nu_1$  peak associated with the nitrate ion is an indicator that the systems used to simulate Hanford solutions cannot be described with a linear chemometrics model if only a single nitrate spectrum is considered. Improvement in a chemometric model can be obtained by reinforcing training data with reference spectra of the nitrate subpeaks, whose behavior corresponds to different ionic interactions occurring in the system.

The present work contributes to bridging the gap between complexities of ion association and the design of training data sets for development of chemometrics models. In particular, the results from this work show that nonlinear phenomena stemming from interactions among various ions in solution can help resolve discrepancies in spectroscopic data and lead to a more efficient design of experiments that target regions where nonlinearities and deviations from ideal behavior become important.

**CHAPTER 5**  
**EXPANSION OF *IN-SITU* MONITORING AND DATA-DRIVEN MODELING**  
**ACROSS MULTIPLE DFLAW SAMPLING LOCATIONS<sup>1</sup>**

**5.1 Introduction**

In order to facilitate direct feed low activity waste (DFLAW) processing, extensive sampling is planned, both to gain process knowledge and to satisfy regulatory requirements [11]. The real-time in-line monitoring (RTIM) program seeks to identify process analytical technology (in-line instruments) to reduce or replace offline sampling. RTIM can reduce the hazards associated with sample collection and improve plant operations. The potential utility of *in-line* infrared and Raman spectroscopy probes in measuring the liquid low-activity waste feed has been the focus of the work presented in Chapters 2 and 3.

Groups at Pacific Northwest National Laboratory and Los Alamos National Laboratory are also working on RTIM instruments, such as Raman spectroscopy and laser-induced breakdown spectroscopy (LIBS) [34]. In order to evaluate the RTIM instruments, simulants that represent various DFLAW sampling locations were developed by Savannah River National Laboratory, using information on the LAW feed, melter feed and effluent management facility [29]. An overview of potential sampling locations for RTIM is shown in Figure 5.1. Note that the program has identified multiple locations for monitoring of the waste, in addition to the LAW feed sampling location that was introduced previously in Chapters 2 and 3.

---

<sup>1</sup>Adapted from "S. Kocavska, S. Crouse, R. Prasad, G. M. Maggioni, M. A. Grover and R. W. Rousseau, "Measuring Liquid Low Activity Waste Simulants with Process Analytical Technology: Real-Time In-Line Monitoring at Hanford," in Proceedings of the Waste Management Symposia, 2022." [68]

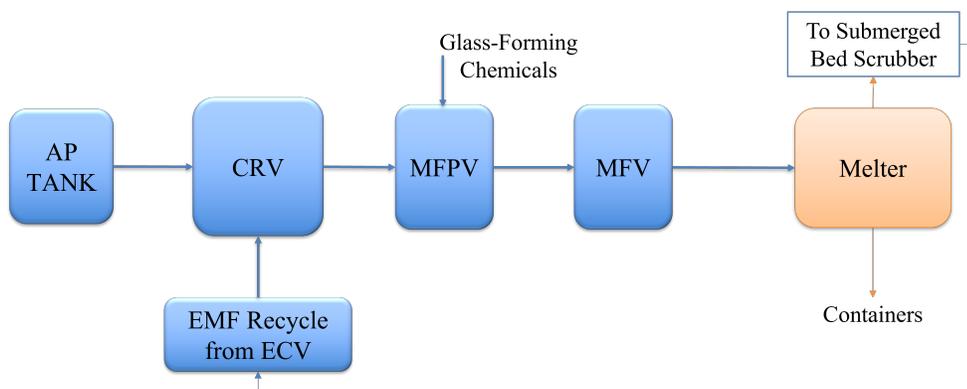


Figure 5.1: Proposed RTIM sampling locations: AP holding tank for low activity waste (LAW) feed, CRV or concentrate receipt vessel, EMF recycle, which contains the effluent management facility recycled waste from the evaporator concentrate vessel or ECV, MFPV or melter feed prep vessel where the glass forming chemicals are added, followed by the MFV (melter feed vessel) [7].

## 5.2 Methods

The IR absorbance and Raman intensity of DFLAW simulants were measured with in-line molecular spectroscopy probes (infrared and Raman spectroscopy, respectively).

### 5.2.1 In-line spectroscopy

In-line infrared (attenuated total reflectance – Fourier transform IR or ATR-FTIR) and Raman spectroscopy were used simultaneously to collect spectra of anion species in the simulated waste. In previous chapters, both techniques were used to collect spectra of liquid simulants. In this contribution, some of the samples include slurry waste, in which case the intensity of the Raman measurements is affected by the presence of solids, while ATR-FTIR spectroscopy can measure the solute absorbance in slurry samples, with negligible interference due to higher solids content. The depth of penetration of the evanescent wave is small for ATR, such that only the signal of the material in direct contact with the probe can be measured, which is mostly the liquid phase [69, 70]. The spectral measurements of the DFLAW simulants using both ATR-FTIR and Raman probes are shown in Figure 5.2.

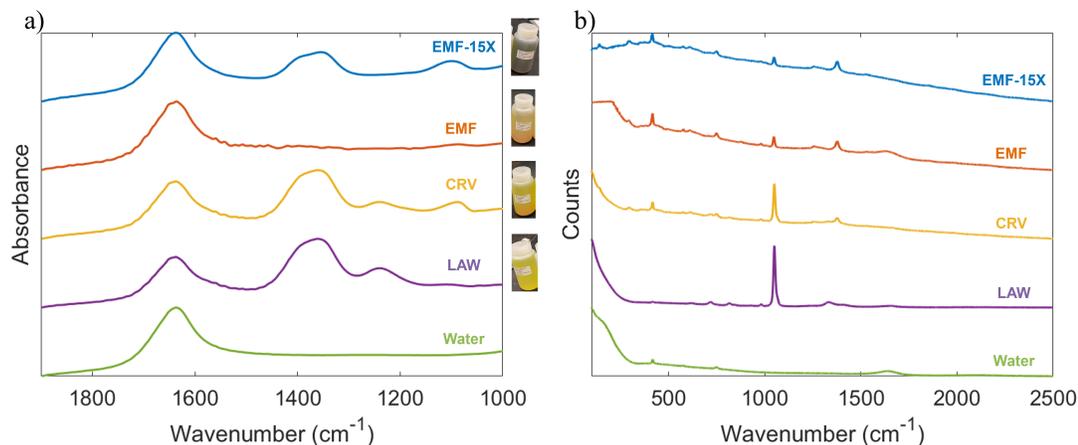


Figure 5.2: Normalized spectra of the DFLAW simulants for a) ATR-FTIR spectroscopy and b) Raman spectroscopy, measured at 20°C for the CRV sample and 25°C for the remaining samples.

### 5.2.2 Blind source separation – partial least squares regression (BSS – PLSR) modeling

The training data set from Chapter 2 was reused for this work (Table 2.1). To generate the test data sets, data points for the LAW and CRV samples were added to the preexisting test data set from Chapter 3 (Table 3.1). In the current work, blind source separation is used as a preprocessing step, to identify the independent components in the mixture spectra, which can then be classified as target or non-target species. Based on this classification, we isolated the signals of target species, so they can be quantified without the interference from the minor (non-target) signals. While in our previous work (discussed in Chapter 3) the contributions of the non-target species were calculated and subtracted from the original spectra (subtraction-based approach), here we also compare a new approach of calculating the contributions of the target species directly from the original spectra (target-based approach). An overview of the steps with an emphasis of the differences between the subtraction-based approach and target-based approach is presented in Figure 5.3.

Once the blind source separation steps are completed, the anion concentrations are calculated using a partial least-squares regression model. The development of the blind source separation – partial least squares regression (BSS – PLSR) framework was described

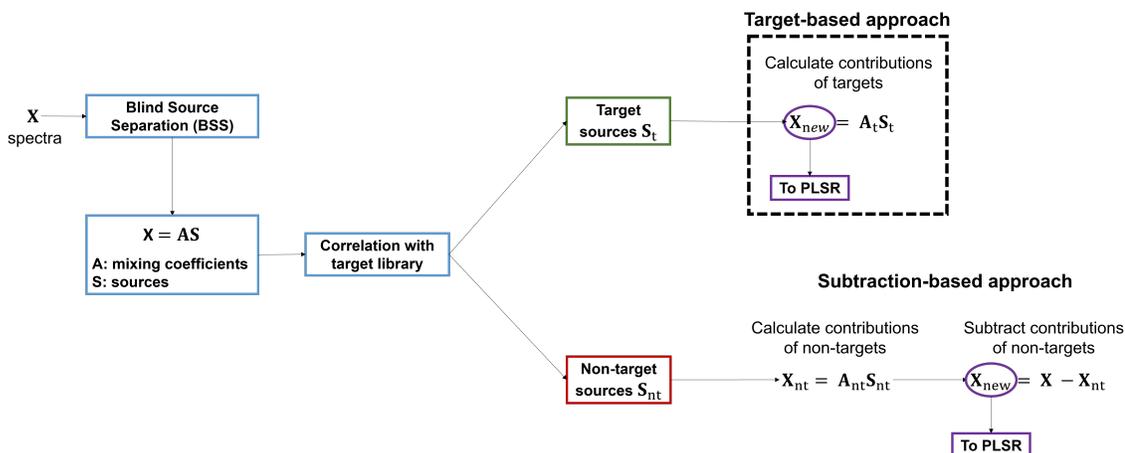


Figure 5.3: Blind source separation (BSS) framework, showcasing the difference between a subtraction-based approach (developed in our previous work in Chapter 3) and a target-based approach introduced in the current work.

in more detail in Chapter 3.

### 5.3 Results and Discussion

The anion concentrations of nitrate, nitrite and sulfate were calculated with a partial least squares regression (PLSR) model, using spectra preprocessed with blind source separation (BSS). The BSS-PLSR approach was compared to quantification of the species using a PLSR model without BSS preprocessing. The results from each model were compared to the anion concentrations quantified with offline ion chromatography (IC) at Savannah River National Laboratory, shown in Table 5.1. Since the IC analysis does not include carbonate and water, the focus in this chapter will be the quantification of the species reported in Table 5.1 (even though the carbonate and water concentrations were predicted by the PLSR and BSS-PLSR models).

#### 5.3.1 Low activity waste (LAW) feed simulant

The LAW feed simulant includes the following molecular species: nitrate, nitrite, carbonate, hydroxide, aluminate, sulfate, phosphate, citrate, formate, acetate, chromate and oxalate. As discussed in Chapter 3, the BSS-PLSR model was developed with a simplified

Table 5.1: Target species concentrations in mol/L, quantified with ion chromatography (IC) at SRNL.

<b>Sample ID</b>	<b>Nitrate</b>	<b>Nitrite</b>	<b>Sulfate</b>
<b>LAW-1</b>	2.07	0.98	0.04
<b>LAW-6</b>	1.82	1.07	0.05
<b>CRV-1</b>	1.46	0.61	0.05

LAW feed simulant, consisting of nitrate, nitrite, sulfate, carbonate and water as the target species and phosphate, acetate and oxalate as the non-target species. The training data for the PLSR model included the following concentration ranges: 0 – 1.32 mol/L for nitrate, 0 – 1.17 mol/L for nitrite, 0 – 0.19 mol/L for sulfate, 0 – 0.57 mol/L for carbonate and 42.5 – 55.5 mol/L for water. While the LAW feed simulant samples are more complex compared to the simulants used to develop the BSS-PLSR framework, since both systems contain the target species (nitrate, nitrite and sulfate), it was assumed the model can quantify their concentrations in the LAW simulants (LAW-1 and LAW-6).

However, the training data set may need to be updated with new measurements to account for the nitrate concentrations in the LAW samples, which are currently out of range from the concentrations used to train the PLSR model. In addition, the LAW simulants have a caustic pH (due to the presence of NaOH), while the pH in the training data for the PLSR model and BSS-PLSR framework was not adjusted. The potential pH effect on the spectral measurements of LAW mixtures needs to be studied further. Finally, since the carbonate concentration was not measured with IC, as shown in Table 5.1, the quantification of carbonate was not included in this work. However, since carbonate was included in the training data, the BSS-PLSR model generated concentration predictions for carbonate as well.

#### *ATR-FTIR Spectroscopy*

ATR-FTIR results using the previously developed BSS-PLSR model (subtraction-based approach used in BSS and first-order derivative spectra used in PLSR) are shown in Fig-

Figure 5.4. Note that Figure 5.4 shows the results for the original test data set from Chapter 3, plus the additional data points for LAW-1 and LAW-6. The nitrate and nitrite concentrations for the additional LAW-1 and LAW-6 samples were outside  $\pm 20\%$  prediction errors, which are based on the uncertainty at the Hanford analytical laboratory [11]. Next, the target-based approach was implemented in BSS-preprocessing, since we can identify the target sources with greater certainty than the non-targets, especially for the more complicated LAW-1 and LAW-6 samples. The results for the target-based approach, while using first-order derivative spectra in PLSR, are shown in Figure 5.5. The target-based approach was also tested with second-order derivative spectra in the PLSR model, which further improved the nitrite predictions, as shown in Figure 5.6. The nitrate prediction errors remained high, perhaps because the nitrate concentrations in the LAW-1 and LAW-6 are outside the range covered in the PLSR training data.

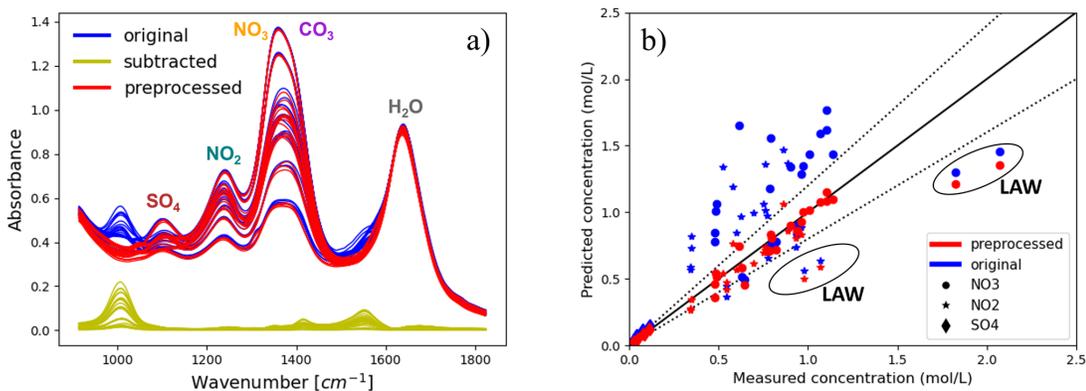


Figure 5.4: ATR-FTIR analysis for LAW feed: a) BSS-preprocessing (using a subtraction-based approach) comparing the original spectra (blue) with the preprocessed spectra (red); b) PLSR parity plot (using first-order derivative spectra) comparing the measured and model-predicted concentrations. The dashed lines indicate the  $\pm 20\%$  error deviations, which are based on the uncertainty at the Hanford analytical laboratory.

### *Raman Spectroscopy*

Raman results using the previously developed BSS-PLSR model (subtraction-based approach used in BSS and first-order derivative spectra used in PLSR) are shown in Fig-

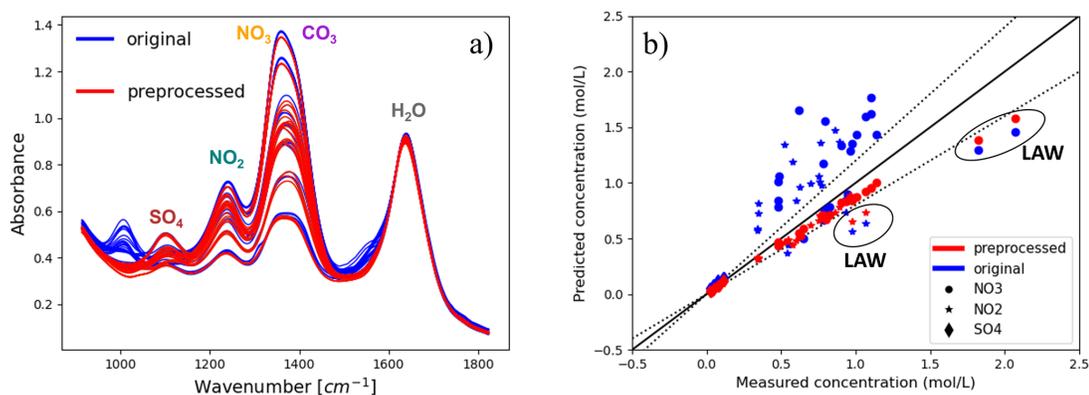


Figure 5.5: ATR-FTIR analysis for LAW feed: a) BSS-preprocessing (using a target-based approach) comparing the original spectra (blue) with the preprocessed spectra (red); b) PLSR parity plot (using first-order derivative spectra) comparing the measured and model-predicted concentrations. The dashed lines indicate the  $\pm 20\%$  error deviations, which are based on the uncertainty at the Hanford analytical laboratory.

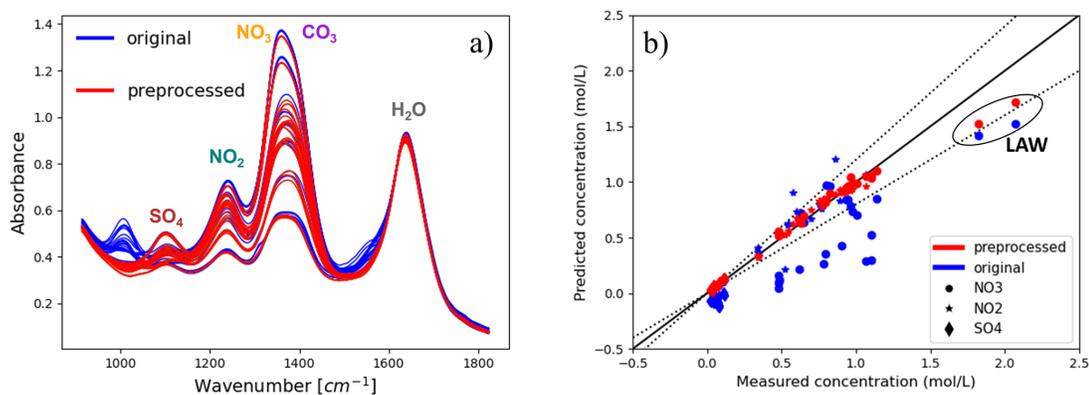


Figure 5.6: ATR-FTIR analysis for LAW feed: a) BSS-preprocessing (using a target-based approach) comparing the original spectra (blue) with the preprocessed spectra (red); b) PLSR parity plot (using second-order derivative spectra) comparing the measured and model-predicted concentrations. The dashed lines indicate the  $\pm 20\%$  error deviations, which are based on the uncertainty at the Hanford analytical laboratory.

ure 5.7. Note that once again Figure 5.7 shows the results for the original test data set from Chapter 3, plus the additional data points for LAW-1 and LAW-6. Similar to the ATR-FTIR results, the LAW nitrate concentration is underestimated. The overall prediction results were improved using the target-based approach with a first-order derivative in PLSR, as shown in Figure 5.8.

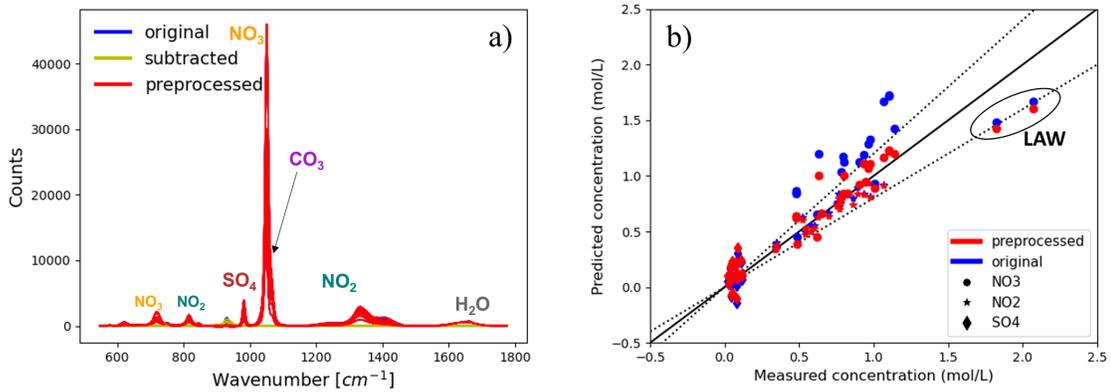


Figure 5.7: Raman analysis for LAW feed: a) BSS-preprocessing (using a subtraction-based approach) comparing the original spectra (blue) with the preprocessed spectra (red); b) PLSR parity plot (using first-order derivative spectra) comparing the measured and model-predicted concentrations. The dashed lines indicate the  $\pm 20\%$  error deviations, which are based on the uncertainty at the Hanford analytical laboratory.

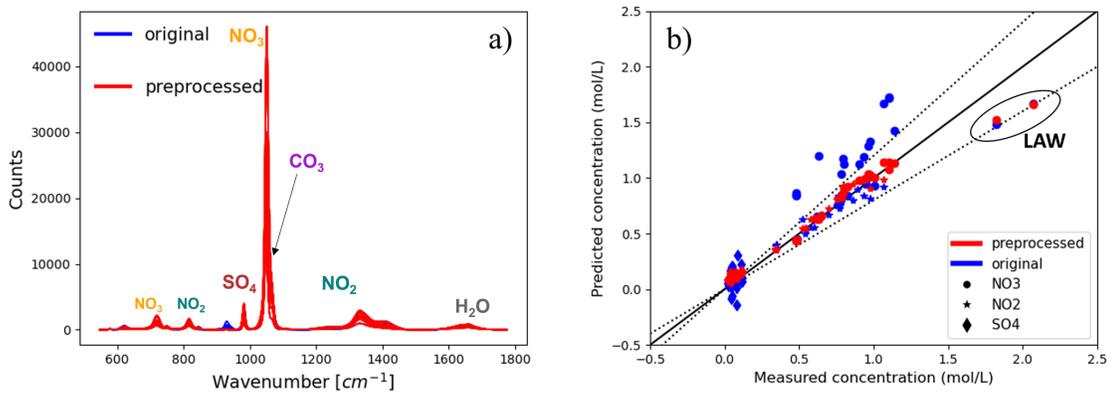


Figure 5.8: Raman analysis for LAW feed: a) BSS-preprocessing (using a target-based approach) comparing the original spectra (blue) with the preprocessed spectra (red); b) PLSR parity plot (using first-order derivative spectra) comparing the measured and model-predicted concentrations. The dashed lines indicate the  $\pm 20\%$  error deviations, which are based on the uncertainty at the Hanford analytical laboratory.

### 5.3.2 Concentrate receipt vessel (CRV) simulant

The CRV simulant combines the LAW feed and EMF concentrate recycle, which is discussed in the next subsection. Despite the presence of solids, the ATR-FTIR probe measures the absorbance of the solution. The BSS-PLSR results using a target-based approach are shown in Figure 5.9. To generate the results, the CRV data points were added to the

model used in the previous section. The results indicate that the concentration predictions for nitrate and nitrite were improved with BSS preprocessing. Note that the analysis of the CRV sample using our previously developed training data is subject to the same limitations as the LAW samples, including any potential pH effects and the extrapolation of the concentration of nitrate (even though the measured IC concentration in the CRV sample is closer to the maximum value in the training set compared to the LAW samples, it remains out of range).

The Raman spectra of the CRV simulant were significantly impacted by the presence of solids, which resulted in a drop of the measured intensity, as shown in Figure 5.10. Therefore, the spectra were not used to quantify the concentrations of the target species.

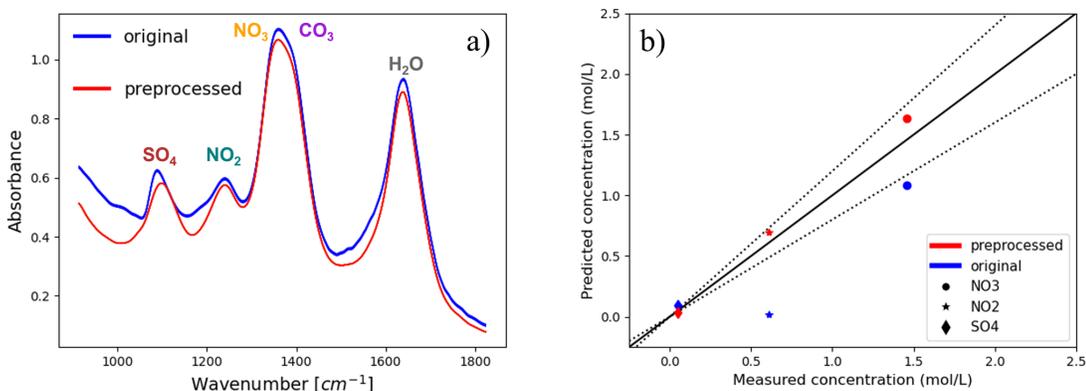


Figure 5.9: ATR-FTIR analysis for CRV: a) BSS-preprocessing (using a target-based approach) comparing the original spectra (blue) with the preprocessed spectra (red); b) PLSR parity plot (using first-order derivative spectra) comparing the measured and model-predicted concentrations. The dashed lines indicate the  $\pm 20\%$  error deviations, which are based on the uncertainty at the Hanford analytical laboratory.

### 5.3.3 Effluent management facility (EMF) simulants

The EMF feed contains nitrate, ammonium, sulfate, chromate, aluminate and phosphate. The EMF concentrate vessel is based on the EMF simulant, assuming a 15-times turndown in the EMF evaporator. While EMF sample spectra were collected for qualitative analyses, the concentrations of the analytes have not been quantified in this work. Figure 2 a) shows

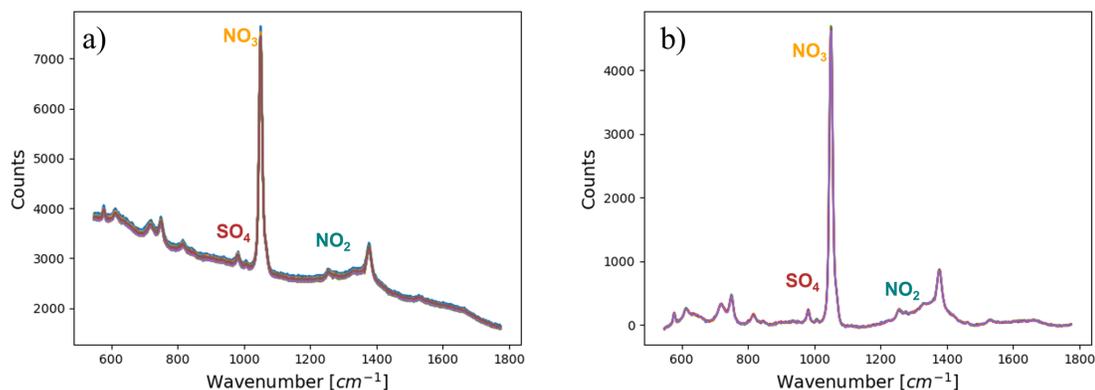


Figure 5.10: Raman spectra for CRV: a) Raw spectra; b) Baseline-corrected spectra.

that only water is visible in the ATR-FTIR spectrum of the diluted EMF sample, while additional peaks are present in the concentrated EMF sample. Figure 5.2 b) shows that the peaks of nitrate and sulfate are still visible in the Raman EMF spectra, despite a decrease in the measured intensity (resulting from the presence of solids in the sample).

## 5.4 Conclusions

In this work, ATR-FTIR and Raman probes were used to measure the spectroscopic signal of simulants associated with the direct feed low activity waste (DFLAW) process, including the low activity waste (LAW) feed, concentrate receipt vessel (CRV) and effluent management facility (EMF) simulants. While the EMF simulant measurements remained qualitative, the spectra measured from LAW and CRV samples were used in a quantitative analysis to predict the concentrations of nitrate, nitrite and sulfate in the simulated waste.

The present work shows that in-line spectroscopy probes, such as infrared and Raman, may be a suitable replacement and/or addition for offline sampling in the DFLAW process, to facilitate continuous waste processing and immobilization. While either probe could be used to measure the spectra of liquid simulants (LAW feed), the ATR-FTIR probe is more suitable for measuring the solution absorbance in the presence of solids (CRV).

The BSS-PLSR model approach developed in our previous work using a simplified

simulant system was of assistance in analyzing the target species (nitrate, nitrite and sulfate) in more complex DFLAW simulants, including the LAW feed and CRV. In the present work, the BSS-PLSR model was improved by incorporating a target-based approach to preprocessing the mixture spectra. The BSS-PLSR model can be further improved by expanding the training range for the nitrate concentration and studying the pH effect on the spectral measurements. Work in this paper shows that the BSS-PLSR model can be incorporated across multiple sampling locations (such as LAW and CRV) to quantify target species, despite changes in the non-target composition of the unknown mixtures.

## CHAPTER 6

# IR SPECTROSCOPY MONITORING COUPLED WITH MACHINE LEARNING FOR DEFENSE WASTE PROCESSING FACILITY (DWPF) APPLICATIONS AT THE SAVANNAH RIVER SITE<sup>1</sup>

### 6.1 Introduction

Unlike the Hanford site which is preparing to begin waste processing, the Savannah River Site (SRS) has been operational since 1996. The site operates the Defense Waste Processing Facility (DWPF), which processes and vitrifies high-level waste into borosilicate glass [71, 72, 73, 74, 75]. Approximately 35 million gallons of high-level radioactive waste remain to be stabilized into a borosilicate glass form.

The DWPF operates multiple batch processes and generates a continuous output. As part of the DWPF, sludge batch waste from the tank farm is fed into the Sludge Receipt and Adjustment Tank (SRAT). The sludge waste contains insoluble solids, including long-lived radionuclides which need to be vitrified. The waste at SRS also contains mercury, which adds further complexity to the treatment process. In addition to the sludge waste, the SRAT batch also receives effluent waste from the Salt Waste Processing Facility (SWPF). Note that at SRS, the low-activity waste is grouted and not vitrified, so there are two facilities that process waste (DWPF for high-level waste vitrification and SWPF for low-level waste grouting).

For DWPF, the key batch processing steps occur in the Sludge Receipt and Adjustment Tank (SRAT) and Slurry Mix Evaporator (SME). A simplified process flowsheet for the DWPF is shown in Figure 6.1. During SRAT processing, nitric and glycolic acid are fed

---

<sup>1</sup>This work was produced by Battelle Savannah River Alliance, LLC under Contract No. 89303321CEM000080 with the U.S. Department of Energy. Publisher acknowledges the U.S. Government license to provide public access under the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

to the batch (following the nitric-glycolic acid flowsheet [72]), to neutralize the alkaline sludge waste and reduce the mercury ( $\text{HgO} \rightarrow \text{Hg}^0$ ). The process is operated at boiling temperatures to strip the mercury. The next batch process occurs in the Slurry Mix Evaporator (SME), where the water is evaporated to concentrate the waste (dewatering) and frit (glass formers) are added to make the melter feed. During batch processing, the heels from previous batches are incorporated in the new batch, which results in variations in the compositions of various species.

The work presented in this chapter is focused on the potential for real-time monitoring of the SRAT and SME cycles using *in-situ* ATR-FTIR spectroscopy. The simulated high-level waste is a challenging system to monitor. The pH of the system varies from 4 to 13 (starting with a caustic solution, the pH decreases as the nitric and glycolic acid are introduced). The SRAT and SME cycles also span changes in temperature, reaching boiling temperatures during mercury stripping. The species concentrations change during the process due to ongoing reactions, such as mercury reduction and nitrite conversion to nitrate. Another complication associated with the sludge batch waste is the high solids content, which reaches 15 wt% (reported as total solids, which do not dissolve at 110°C). Due to the high solids content, the sludge waste is probed with ATR-FTIR spectroscopy, which measures the absorbance of the liquid solution, even in the presence of suspended solids. The work in this chapter is focused on detection and monitoring of the anionic species, discussed in more detail in the next section.

## 6.2 Materials and Methods

### 6.2.1 Overview of the waste

In preparation for processing of Sludge Batch (SB) 10, SRNL has performed flowsheet testing with nonradioactive simulants to understand the chemistry for the SB 10 campaign [75]. The physical and chemical composition of a representative sludge simulant is shown in Table 6.1. In addition to the ionic species, the simulant also contains the following solids:

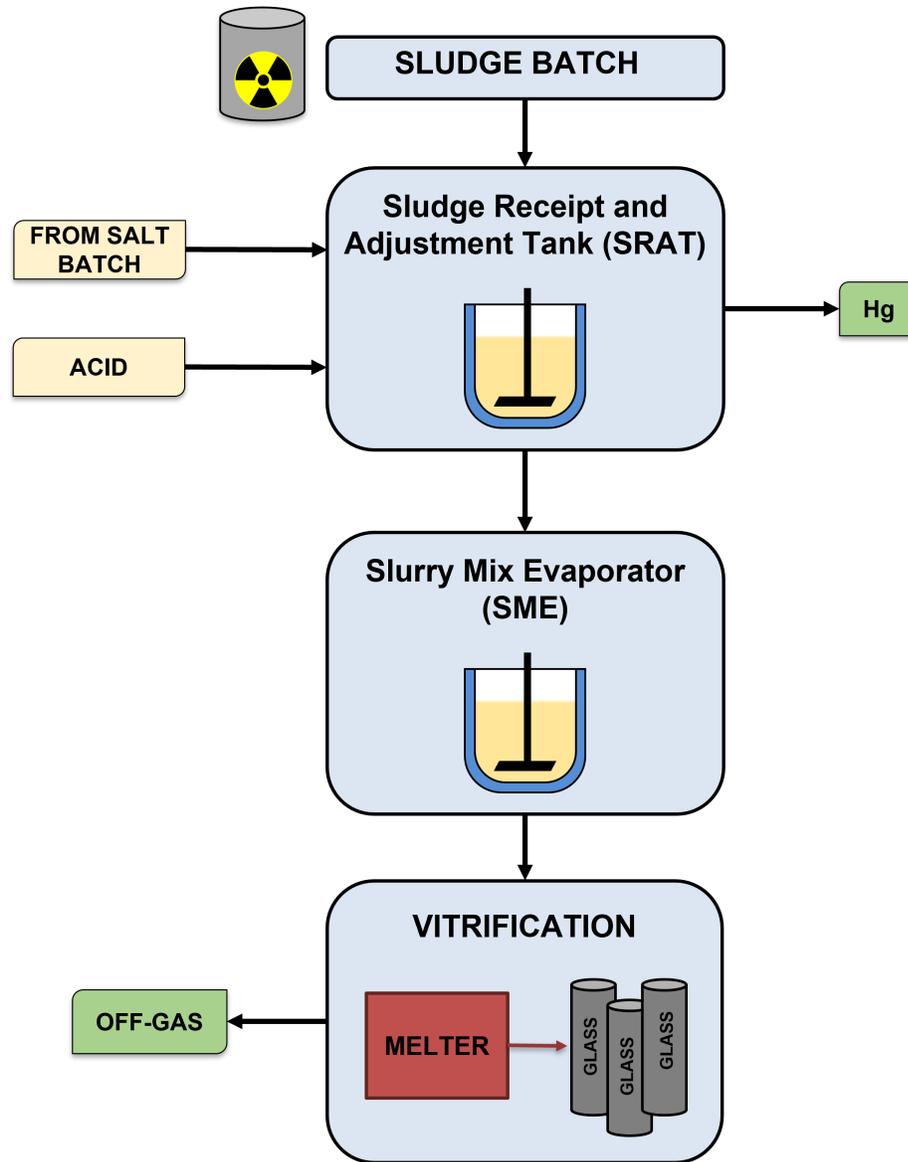


Figure 6.1: Simplified flowsheet for Direct Waste Processing Facility (DWPF) processing [71].

Al, Ba, Ca, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, S, Si, Zn, and Zr [75].

During SRAT and SME processing, the simulated waste from Table 6.1 is used as the starting point. The simulant which represents the final point (*i.e.* the SRAT/SME product slurry) is shown in Table 6.2, focusing on the anionic species. The target anion species in the SRAT and SME waste include nitrate ( $\text{NO}_3^-$ ), nitrite ( $\text{NO}_2^-$ ) and glycolate ( $\text{HOCH}_2\text{CO}_2^-$ ), whose concentrations are expected to change during processing. Other

Table 6.1: Tank 40 sludge waste nonradioactive simulant composition, reported by Woodham and co-workers [75]. The wt% total solids (TS) represents the solids that do not dissolve after heating to 110°C, while the wt% insoluble solids (IS) represents the difference between the measured total solids and dissolved solids.

<b>Tank 40</b>		
<b>Parameter</b>	<b>Target</b>	<b>Measured</b>
<b>wt% Insoluble Solids (IS)</b>	9.45	8.39
<b>wt% Total Solids (TS)</b>	14.49	13.90
<b>Slurry Density (g/mL)</b>	-	1.1102
<b>Supernatant Density (g/mL)</b>	-	1.0477
<b>Na<sup>+</sup> ([M] supernatant)</b>	9.48E-01	1.01
<b>HCO<sub>2</sub><sup>-</sup> ([M] supernatant)</b>	-	<2.22E-03
<b>Cl<sup>-</sup> ([M] supernatant)</b>	1.06E-03	<2.82E-03
<b>NO<sub>2</sub><sup>-</sup> ([M] supernatant)</b>	2.30E-01	1.94E-01
<b>NO<sub>3</sub><sup>-</sup> ([M] supernatant)</b>	1.21E-01	1.19E-01
<b>SO<sub>4</sub><sup>2-</sup> ([M] supernatant)</b>	1.64E-02	1.37E-02
<b>C<sub>2</sub>O<sub>4</sub><sup>2-</sup> ([M] supernatant)</b>	1.04E-02	7.00E-03
<b>PO<sub>4</sub><sup>3-</sup> ([M] supernatant)</b>	3.69E-04	<1.05E-03
<b>OH<sup>-</sup> ([M] supernatant)</b>	3.73E-01	2.19E-01
<b>CO<sub>3</sub><sup>2-</sup> ([M] supernatant)</b>	5.66E-02	3.61E-02

species of interest include formate and oxalate, which may be produced as side products. The key anion conversions that may occur during the SRAT and SME processes include nitrite destruction, nitrite-to-nitrate conversion, glycolate destruction, glycolate-to-formate conversion and glycolate-to-oxalate conversion [75]. Detailed information on the expected anion conversions and limiting factors is provided by Woodham and coworkers [75]. Typically, the anion concentrations are quantified with IC which is associated with approximately 10% uncertainty and long waiting times on the order of days. The goal of this work is to show a proof of concept regarding measurement of target species in real time using ATR-FTIR spectroscopy.

### 6.2.2 Design of Experiments

In order to evaluate if ATR-FTIR spectroscopy is suitable for measurement of the sludge waste, an experiment simulating the SRAT and SME cycles was performed in a 2-L Mettler Toledo Reaction Calorimeter (RC1) vessel equipped with temperature, pH and ATR-FTIR

Table 6.2: Ion composition in the nonradioactive simulated Tank 40-8 SRAT/SME product slurry, reported by Woodham and co-workers [75].

Tank 40-8	
Parameter	[M] slurry
$\text{HCO}_2^-$	1.7E-02
$\text{Cl}^-$	3.70E-03
$\text{NO}_2^-$	<2.48E-03
$\text{NO}_3^-$	1.07E+00
$\text{PO}_4^{3-}$	<1.20E-02
$\text{SO}_4^{2-}$	1.75E-02
$\text{C}_2\text{O}_4^{2-}$	9.06E-01
$\text{HOCH}_2\text{CO}_2^-$	9.06E-01
$\text{CO}_3^{2-}$	<6.66E-03
$\text{NH}_4^+$	<2.77E-03

probes. The main SRAT and SME steps are summarized in Figure 6.2. The anion concentrations measured offline with IC are summarized in Table 6.3. The RC1 experimental set-up was used to collect the training data, which were designed based on the anion concentrations in the sludge simulant (Table 6.1) and SRAT/SME product (Table 6.2). The concentrations associated with the training data are shown in Table 6.4.

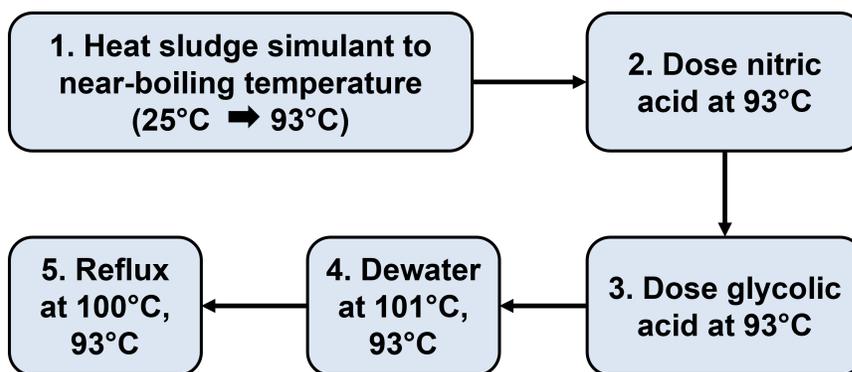


Figure 6.2: Steps for the SRAT/SME cycle to be monitored with ATR-FTIR spectroscopy [75].

### 6.3 Results and Discussion

The first step to evaluating the feasibility of ATR-FTIR for monitoring of the slurry waste was to collect reference spectra measurements. While nitrate, nitrite and glycolate rep-

Table 6.3: Anion composition measured by IC for a nonradioactive lab run of SRAT/SME in the RC1 experimental set-up. The samples match the steps in Figure 6.2.

Sample	Time point	$\text{NO}_3^-$ [M]	$\text{NO}_2^-$ [M]	$\text{HOCH}_2\text{CO}_2^-$ [M]
1	Sludge at 93°C	0.16	0.18	0.00
2	Post nitric acid addition	0.79	0.15	0.00
3	Post glycolic acid addition	0.79	0.03	0.59
4	Post dewater	1.28	0.00	0.82
5	Post reflux	1.43	0.00	0.86

Table 6.4: Training data set for the target anion species and solvent (water).

Sample	$\text{NO}_3^-$ [M]	$\text{NO}_2^-$ [M]	$\text{HOCH}_2\text{CO}_2^-$ [M]	$\text{H}_2\text{O}$ [M]
1	0.00	0.00	0.00	55.49
2	0.95	0.00	0.00	51.03
3	0.00	0.28	0.00	54.42
4	0.00	0.00	0.79	51.17
5	0.14	0.04	0.81	50.28
6	0.88	0.26	0.75	46.28
7	0.95	0.05	0.00	50.82
8	0.16	0.30	0.00	53.60

resent the target species for the slurry, in this step, the data collection also included the remaining anion species. The reference data collected at room temperature are shown in Figure 6.3. While all the species have characteristic IR peaks, the target species seem to be highly IR-active, while some of the non-target species have weak IR signals, such as oxalate and formate. The spectra shown in Figure 6.3 were collected at concentrations similar to those expected in the simulants. In cases where the concentrations in the simulants did not produce any signal (oxalate, formate, phosphate, sulfate), the concentrations were increased until visible peaks were observed. Since these species do not exhibit high concentrations in the waste, it is possible that they will not interfere with the signals of the targets significantly. However, many of the non-target signals overlap with target peaks, especially with glycolate and nitrate. In addition, since the species concentrations vary during the SRAT/SME cycles (due to the batch heels and ongoing reactions and conversions), it is important to collect all pertinent information of the system, in case the signals of the

non-target species are higher than expected and require preprocessing to be removed.

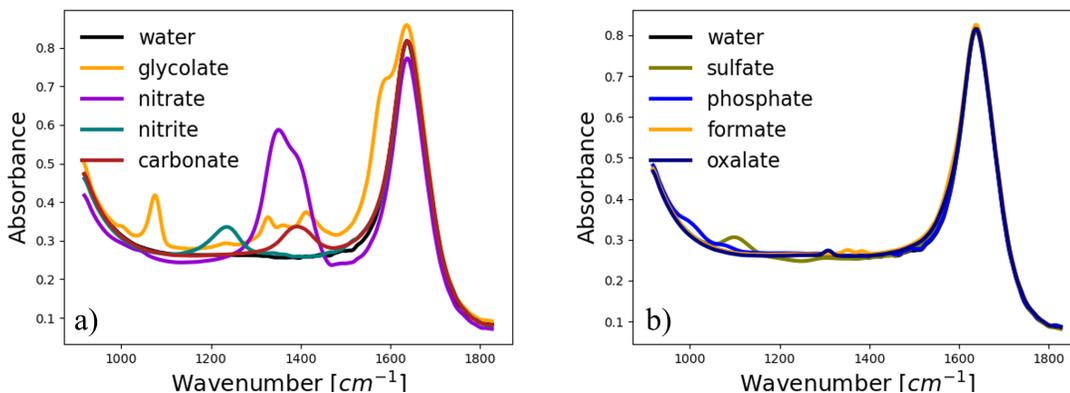


Figure 6.3: Reference spectra of anion species found in the DWPF waste during SRAT/SME cycling. The spectra were collected at room temperature, with concentrations similar to those expected in the simulants.

In the previous chapters, we worked with isothermal systems. However, the temperature during DWPF processing reaches boiling temperature and the key measurements during real testing are performed near boiling temperatures (approximately 93°C). The reference spectra of nitrate indicate that the peak shape is temperature-dependent, shown in Figure 6.4. Since the key measurements during DWPF occur at a constant temperature of 93°C, the training data for the system were collected at near-boiling temperatures, to remove any spectral changes associated with temperature changes. The training data for the system are shown in Figure 6.5 (a). The test data, shown in Figure 6.5 (b), were collected during one full run of the SRAT/SME simulation. Additional data will be collected in the future and incorporated in the models.

Similar to our previous work from Chapter 5, the blind source separation model was "target-based" (refer to Figure 5.3). The reasoning to apply target-based BSS is our lack of knowledge of the non-target species in the system. While we have an idea of the species and levels of concentrations of the non-target components, their chemistry has not been fully elucidated. There also may be additional non-target species that need to be studied in detail, such as MnO or HgO. However, we do have a database of the target species (nitrate,

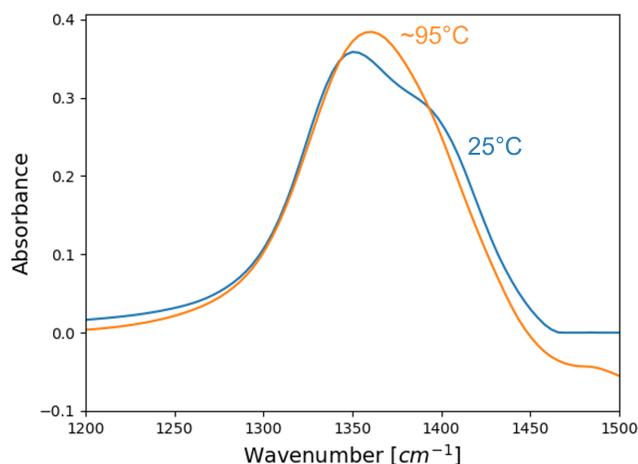


Figure 6.4: Comparison of the nitrate spectra at different temperatures.

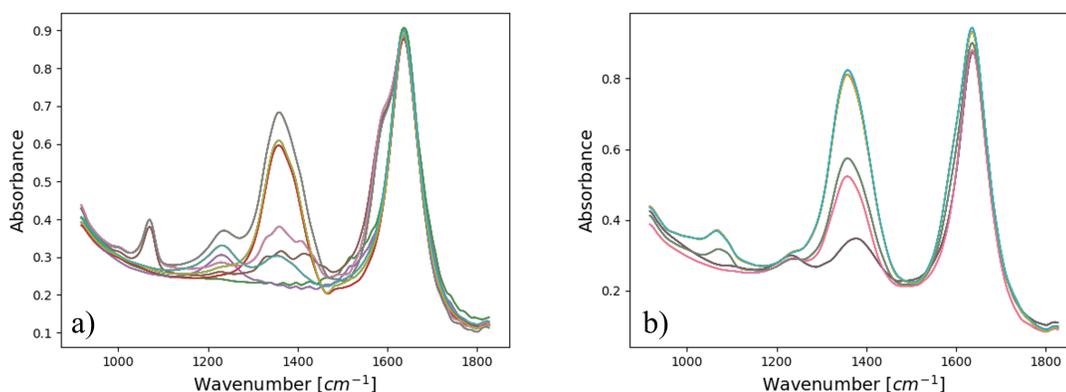


Figure 6.5: (a) Training data associated with Table 6.4; (b) Test data associated with Table 6.3.

nitrite and glycolate), therefore estimating the target contributions directly from the sources which correlated with the target library should introduce less uncertainty to the process.

The corresponding BSS sources which correlated with the target library are shown in Figure 6.6. The results indicate that the target sources were identified with high certainty, considering the values for the correlation coefficients.

The BSS model also identified a source for the non-target species, whose contribution was not included in the BSS-preprocessed data (see Figure 6.7). The non-target source is active in the 1400-1600  $\text{cm}^{-1}$  region, and also has a raised baseline in the 1600-1800  $\text{cm}^{-1}$

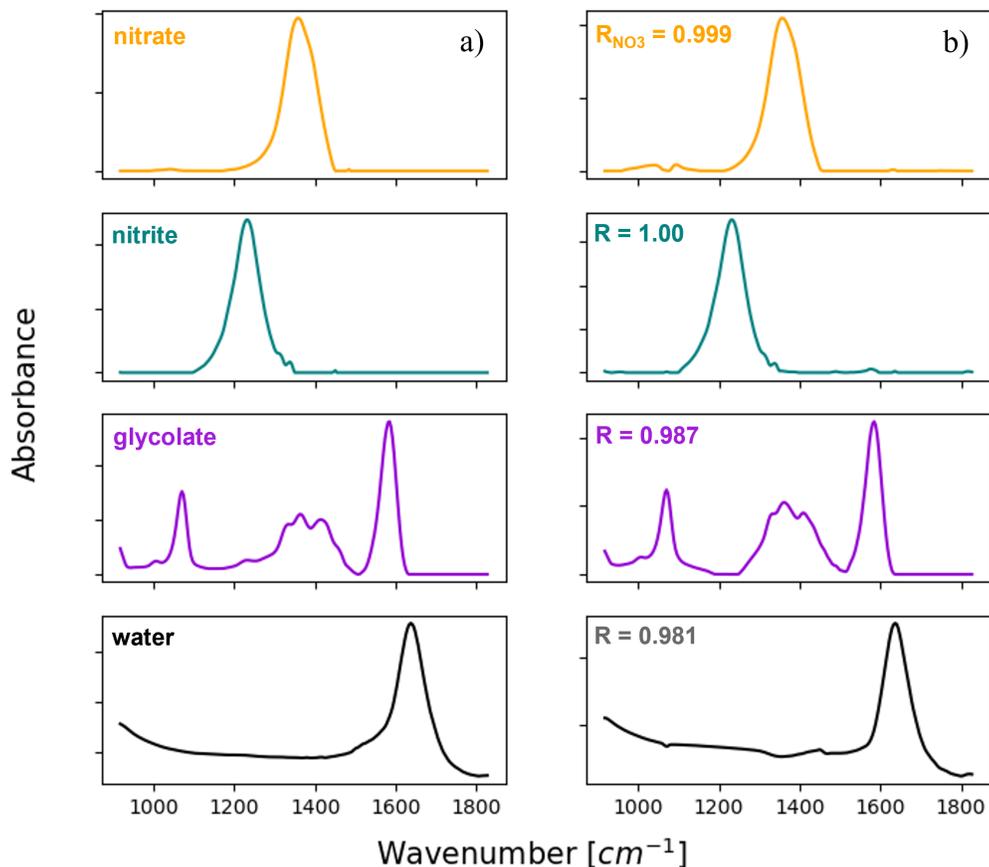


Figure 6.6: (a) Library of reference spectra of the target species; (b) BSS sources which correlated with the references.

region and a small peak around  $1100\text{cm}^{-1}$ . Based on the location of the peaks, the source may include information from oxalate and sulfate.

Comparing the BSS-preprocessed spectra to the original data, shown in Figure 6.8 (a), indicates that the BSS-preprocessing did not remove any significant peaks from the data. The most apparent subtraction occurs in the nitrate/carbonate region, for one of the lower concentration spectra. The parity plots comparing the concentration predictions for both models are shown in Figure 6.8 (b). Both models seem to have predictions comparable to the IC-measured concentrations. As discussed previously, the non-target species do not have a major impact on the spectral measurements. However, the predictions of the species

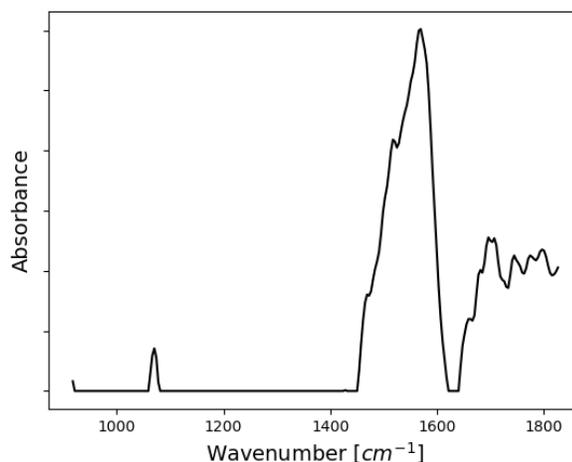


Figure 6.7: BSS source which did not correlate with any of the reference spectra in the target library.

can be further improved, especially for the high concentrations. Upon closer inspection, we learned that the concentrated samples (post dewatering in the SRAT/SME cycle) have higher concentrations than the upper range in the PLS training model, which needs to be expanded. Our understanding of the kinetics in the system is also limited and more work needs to be performed to map out possible reactions and conversions, along with their effect on the spectral measurements. The future work is discussed in more detail in Chapter 7.

The predicted concentrations associated with each model are summarized in Table 6.5 and Table 6.6. Another depiction of the results is shown in Figure 6.9. The figure shows the trends for the concentrations of nitrate, nitrite and glycolate, with an overall increase in the nitrate and glycolate and a decrease in the nitrite. This work shows the potential for real-time IR monitoring to map the reactions in the process.

## 6.4 Conclusions

The work in this chapter serves as a proof of concept on the use of ATR-FTIR spectroscopy for in-line monitoring of slurry samples. The solid content for the SRAT/SME test run reached 15 wt% (reported as total solids, which do not dissolve at 110°C). This work in

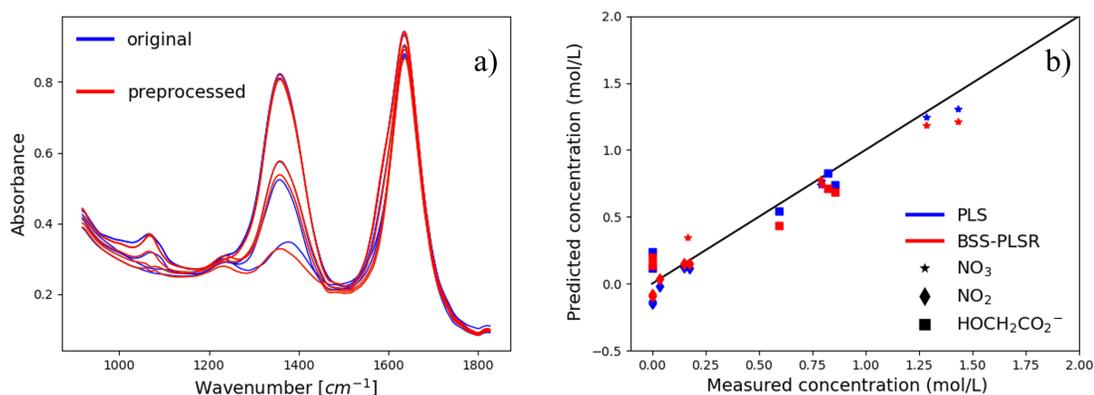


Figure 6.8: (a) Overlay of the spectra before and after BSS-preprocessing; (b) Parity plot comparing the concentration predictions for a PLSR model (blue) and BSS-PLSR model (red).

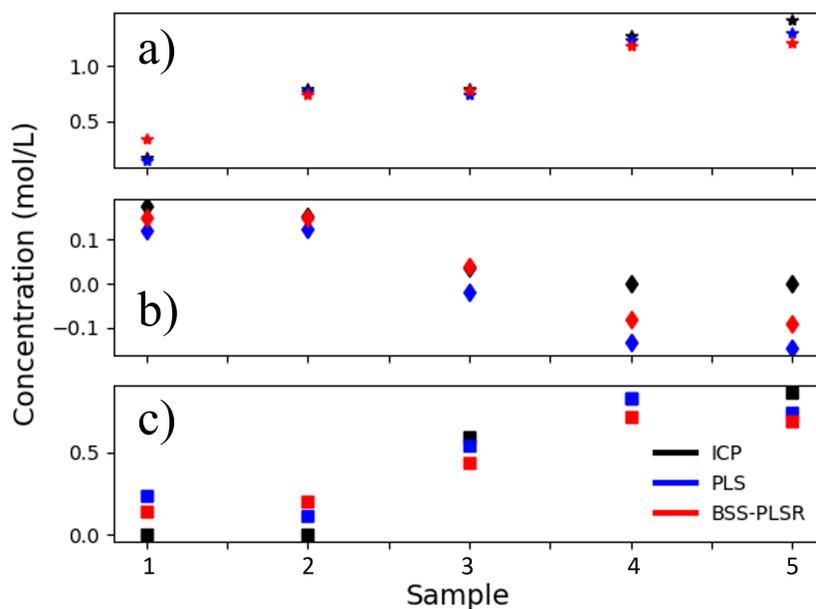


Figure 6.9: (a) Nitrate concentrations over time; (b) Nitrite concentrations over time; (c) Glycolate concentrations over time.

this chapter is also an example of a successful implementation of ATR-FTIR monitoring at near-boiling temperatures. As a result, the PLSR and BSS-PLSR models were built at near-boiling temperatures (93°C). The concentrations obtained with the constant-temperature model were comparable to the matching concentrations from IC measurements. However,

Table 6.5: Table of the concentration predictions using PLSR, matching the sample numbers for the IC measurements from Table 6.3.

Sample	$\text{NO}_3^-$ [M]	$\text{NO}_2^-$ [M]	$\text{HOCH}_2\text{CO}_2^-$ [M]	$\text{H}_2\text{O}$ [M]
1	0.14	0.12	0.24	53.09
2	0.77	0.12	0.12	50.76
3	0.74	-0.02	0.54	49.12
4	1.25	-0.13	0.83	45.63
5	1.31	-0.15	0.74	45.87

Table 6.6: Table of the concentration predictions using BSS-PLSR, matching the sample numbers for the IC measurements from Table 6.3.

Sample	$\text{NO}_3^-$ [M]	$\text{NO}_2^-$ [M]	$\text{HOCH}_2\text{CO}_2^-$ [M]	$\text{H}_2\text{O}$ [M]
1	0.34	0.15	0.14	52.54
2	0.75	0.15	0.20	50.31
3	0.78	0.04	0.44	49.31
4	1.19	-0.08	0.71	46.33
5	1.21	-0.09	0.68	46.39

future monitoring may need to incorporate the temperature as a variable, to adjust for the temperature change during the process, if the data are to be collected continuously during the SRAT/SME cycles [17].

The concentrations of the target species (nitrate, nitrite and glycolate) were predicted by both PLSR and BSS-PLSR. The non-target species do not seem to have a major effect on the measurements, however detailed analysis of their limits of detection are needed to proceed with the modeling of the cycle.

Overall, the results from this chapter are promising and show that ATR-FTIR can measure the change in concentration in the target species as the process is ongoing. This capability would make the process more efficient and reduce the need for offline sampling, while clearly mapping out when reactions are completed.

## CHAPTER 7

### CONCLUSIONS AND FUTURE DIRECTIONS

#### 7.1 Conclusions

This thesis combines recent advances in machine learning with real-time data collection using *in-situ* spectroscopy technology, to address the challenge associated with nuclear waste remediation. While progress has been made towards the clean-up of nuclear waste, the two sites discussed in this work contain a significant portion of the remaining nuclear waste in the United States. The Hanford site in Washington State and the Savannah River Site in South Carolina store a total of 91 million gallons of radioactive waste. The nuclear waste poses an environmental threat to the surrounding areas, and storage tanks have leaked radioactive waste in the past. In order to address this challenge, the waste will be processed into a glass or grout form for stabilization. There are many unknowns associated with the stabilization processes, which ultimately affect the stability of the final glass product. This thesis aims to remove the barrier in real-time quantification of nonradioactive components in the liquid low activity waste at Hanford, which pass through the vitrification plant during the Direct Feed Low-Activity Waste (DFLAW) process. DFLAW is scheduled to begin processing waste into glass in 2023. On the other hand, the Defense Waste Processing Facility (DWPF) at the Savannah River Site has been processing high-level waste for many years. This thesis also addresses the use of *in-situ* technology for monitoring of DWPF waste, allowing for elucidation of the chemistry of the waste and faster sample turnaround times using real-time information.

This thesis started with the design of a simplified mixture system that simulates the process-relevant species in DFLAW waste. The goal of the initial work in Chapter 2 was to analyze the applicability of Raman and ATR-FTIR spectroscopy probes for nuclear

waste monitoring. The nonradioactive, but process-relevant species of particular interest for this project include nitrate, nitrite and sulfate. In this work, we developed a general framework to study subsets of well-defined mixture systems, outlined in Figure 7.1. Various linear multivariate models were applied to extract the concentration of the species from the spectroscopy data, including Classical Least-Squares Regression (CLSR), Inverse Least-Squares Regression (ILSR), Partial Least-Squares Regression (PLSR) and Principal Component Regression (PCR). We discovered that when one has a relatively well-defined mixture with only a few components, the linear models are successful at predicting the concentrations of the species for both IR and Raman spectroscopy.

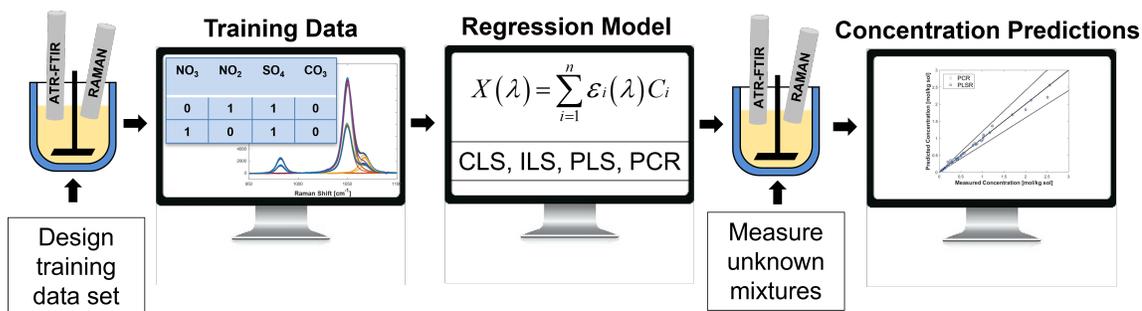


Figure 7.1: Approach for a well-defined multicomponent mixture system, relying on design of experiments and chemometric modeling.

However, the story changes as more species are introduced to the mixtures and the collection of calibration data to retrain the models becomes the bottleneck in the process. Depending upon the number of constituents, the size of the calibration set can be quite large. In some cases, the mixtures may contain numerous species, but only a small subset is central to the process for which quantification is being undertaken. For example, the nuclear waste at the Hanford site contains a large number of radioactive and non-radioactive species, which complicates remediation efforts. However, only the concentrations of a few target species (such as nitrate, nitrite and sulfate) may need to be quantified in real-time to facilitate operation of the clean-up process. In Chapter 3, we introduced a preprocessing procedure that reduces the need for extensive model calibration. The preprocessing framework uses blind source separation (BSS) to identify the independent components in

the mixture, which is followed by a correlation to classify them as either target species (part of the critical quality attributes that need to be measured during waste processing) or non-target species. The classification is used to preprocess the original mixture data: the signals of the target components are retained, while those of the non-target components are removed. Since the preprocessed spectra only contain the target components, the spectra-to-concentration regression model can be trained with a smaller calibration set. The approach was tested for Raman and infrared spectroscopy using simulated and experimental data sets based on simulant mixtures of the nuclear waste, summarized in Figure 7.2.

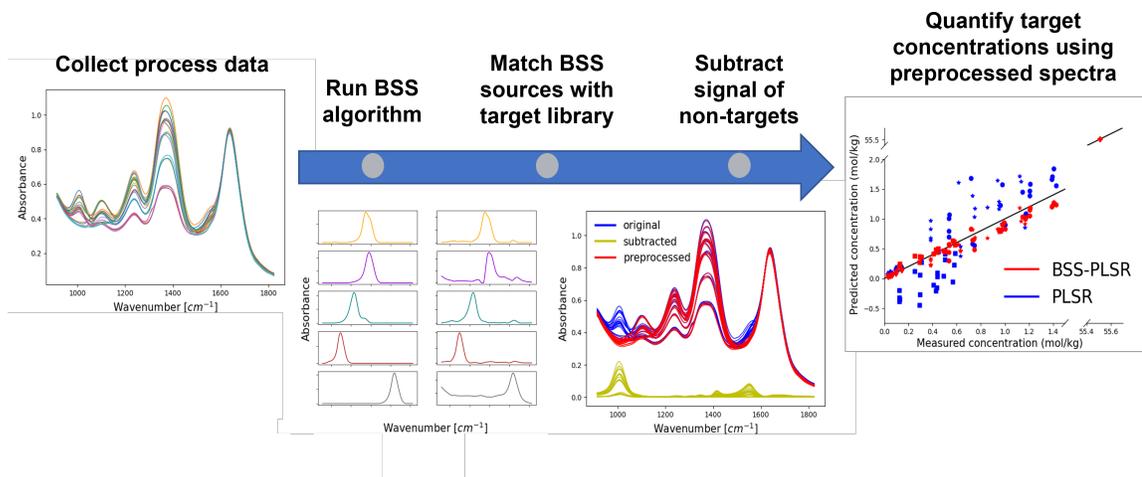


Figure 7.2: (a) BSS-PLSR framework applied to Hanford waste as part of DFLAW.

While the BSS-PLSR approach was successful at simplifying the spectroscopy data before we reached the quantification step, challenges related to the linearity of the data arose. In particular, the BSS model (which is linear) identified a separate source for a blue shift of nitrate, indicating the presence of nonlinear phenomena in the mixtures.

Chapter 4 explores the nonlinear behavior associated with the nitrate anion,  $\text{NO}_3^-$ , which is present in abundance in low activity nuclear waste. We examine changes in the main Raman peak of the nitrate anion in the presence of other ions. A wide range of concentrations are covered, including those expected during direct-feed low-activity waste (DFLAW) processing at the Hanford site. The experiments showed that the ions interact and associate to form ion pairs, which results in a blue shift (*i.e.* a shift towards higher

wavenumbers) in the main Raman peak of nitrate. The results indicate that cation concentration is a better predictor of the peak shift, compared to ionic strength, both for binary and multicomponent mixtures. These findings have direct implications on the development of spectra-to-composition models for the DFLAW system, since they show deviations from the linearity assumptions used in common chemometric models. An overview of the work from Chapter 4 is shown in Figure 7.3.

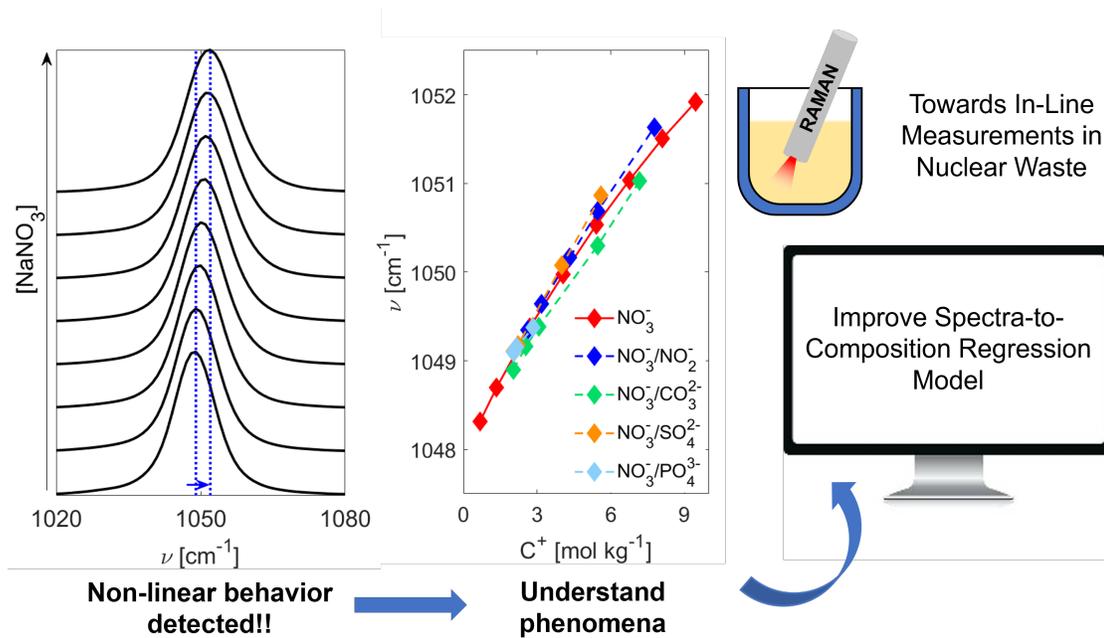


Figure 7.3: Overview of the non-linear phenomena detected in low-activity waste simulants and next steps for improvements in the models

The work up until Chapter 4 was centered around the low-activity waste feed for Direct Feed-Low Activity Waste (DFLAW) processing. The next two chapters diverged from this narrative to explore additional applications for *in-situ* monitoring. First, Chapter 5 extended beyond the feed, and the pre-developed modeling techniques were tested with waste of varying compositions. The sampling location included vessels from the DFLAW process, including the Concentrate Receipt Vessel (CRV), where the LAW feed meets the effluent recycle. More complex LAW feed samples that represent the real simulated feed with all the major and minor components were also quantified, taking the *in-situ* monitoring

to the next level. The location of the sampling in DFLAW is shown in Figure 7.4.

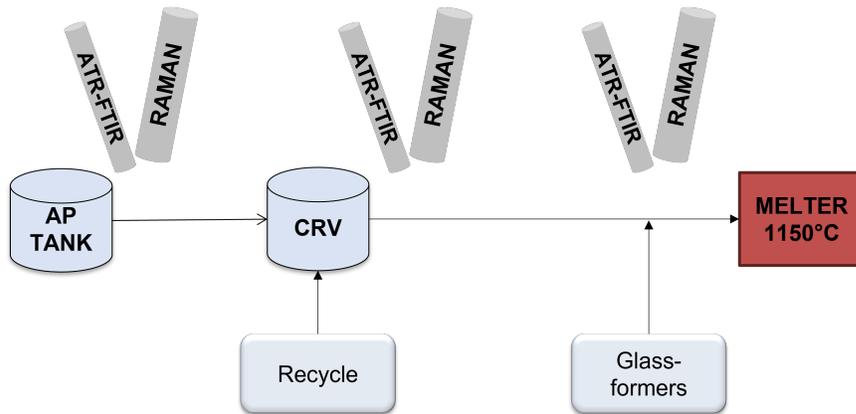


Figure 7.4: DFLAW process section that contains multiple potential sampling locations.

The scenarios from Chapter 5 indicated that the previously developed tools can be effective in measuring additional wastes, but the training data for the target species should still span the range of the expected concentrations. This work also showed that the ATR-FTIR and Raman performance diverges when it comes to analyzing slurry samples, since the Raman signal is diminished from the presence of suspended solids, while the ATR-FTIR measurements were not affected. Chapter 6 explores potential applications for *in-situ* ATR-FTIR probe deployment in the Savannah River Site (SRS) Defense Waste Processing Facility (DWPF). ATR-FTIR could be useful in elucidating the chemistry of the DWPF, as well as tracking reactions in real time and reducing the processing lag time associated with grab sampling, as outlined in Figure 7.5. As part of this work, the IR probe was used to track the concentrations of glycolate, nitrate and nitrite during a simulated run of the Sludge Receipt and Adjustment Tanks (SRAT) and Slurry Mix Evaporator (SME) tank. This work has the unique application to track the extent of reactions in the tanks, during the oxidation and reduction steps. Previous applications of the IR and Raman probe monitoring at Hanford only involved monitoring of preexisting tank waste feeds. Even with a limited amount of data points, the blind source separation and partial least-squares algorithms was able to aid in the concentration calculations. The overall concentrations trends were consistent

with the IC measurements.

This thesis was focused on developing data modeling frameworks that can improve the flow of experimental design, data collection and data processing, putting an emphasis on improving the robustness of data-driven models. The various applications of the modeling tools developed in Chapters 2 and 3 showcase their versatility in analyzing multidimensional data and classifying the components based on process relevance.

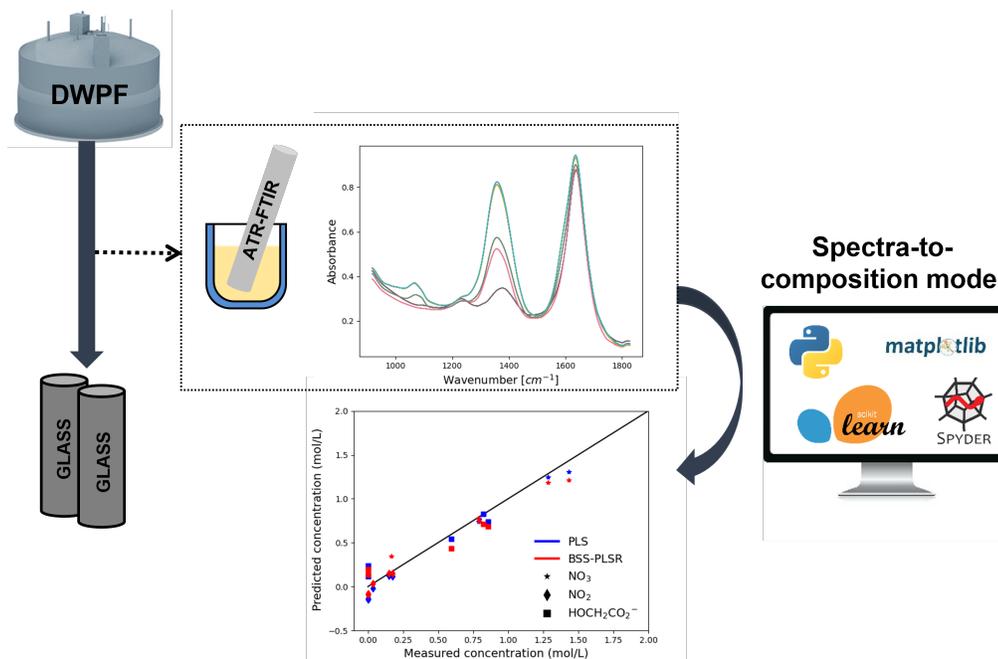


Figure 7.5: Application of real-time monitoring and data-driven modeling for the DWPF process at SRS.

## 7.2 Future Directions

There are a few venues that could be explored as a continuation of the work presented in this thesis.

### 7.2.1 Future work with Blind Source Separation (BSS)

#### *Uncertainty Quantification*

The BSS-PLSR framework developed in Chapter 3 of the thesis proved useful for preprocessing of liquid waste mixtures. However, it is important to understand the uncertainty associated with the BSS-preprocessing and how it varies for the subtraction-based versus the target-based implementation of the framework. In order to have a comparison study, the confidence intervals for both BSS and PLSR should be quantified. The uncertainty should also be quantified for PLSR models which have a full training data set and PLSR models which omit the non-target species in the training data. These analyses would provide insight on the framework, and provide a means for comparison of the sources of error for each method. The final step in the study should be an uncertainty analysis of the IC data, which have served as the reference data for species measurements in this work, even though the Hanford and SRNL sites have cited uncertainty levels of up to approximately 10% for the SRNL lab and 20% for the Hanford lab [75, 11].

#### *Application of BSS to slurry samples in the Melter Feed Prep Vessel (MFPV)*

The MFPV is the point of contact for the glass-forming chemicals (GFCs) and the low-activity waste feed. However, preliminary work has shown that some of the GFC components may dissolve in the feed. In order to eliminate any grab samples at this point, the ATR-FTIR could be used to verify the composition of the feed before and after the GFCs have been added. The ATR-FTIR information can be used to close the mass balance around this vessel and ensure that the concentrations of the target species are within the acceptance criteria and can be fed into the melter. This application also calls for introducing an elemental probe, such as LIBS, to quantify the elemental composition of the glass formers to ensure that the glass formulation would pass the acceptance criteria.

During the measurements the data would need to be evaluated and a processing ap-

proach would need to be selected accordingly. One approach could involve understanding which species are dissolving and quantifying their composition in the liquid phase with ATR-FTIR. For this purpose, BSS could be used to remove the signal of the species otherwise present in the liquid waste, especially if there is overlap.

However, if the interest lays in measuring the liquid waste species, despite the dissolution of any glass formers, then the opposite approach could be applied, where BSS is used to remove the signals of the GFC species and the quantification of the process-relevant species (such as nitrate) can proceed without the need to develop a new training data set that incorporates the dissolving glass formers.

### 7.2.2 Understanding the nonlinear phenomena in the spectra and its implication on waste quantification

*Threshold for nonlinear phenomena during partial least-squares regression (PLSR) - Exploring Cover's theorem for applications in nuclear waste monitoring*

Cover's theorem states that: "A complex pattern-classification problem, cast in a high-dimensional space nonlinearly, is more likely to be linearly separable than in a low-dimensional space, provided that the space is not densely populated [76]." A study that takes a closer look at the behaviour of the nonlinearity as a function of the number of principal components (or latent variables should be conducted), using the data from Chapter 4, which showcase different levels of peak shifting.

### 7.2.3 Using ATR-FTIR to elucidate the chemistry during Defense Waste Processing at SRS

The high-level waste at SRS goes through many reaction steps before it reaches the melter. Since each sludge batch campaign is unique due to the varying composition of the waste, the lab needs methods for understanding the chemistry during processing. The process is especially complex since it involves mercury reduction and stripping. The first step towards

tracking multiple species in real-time using ATR-FTIR spectroscopy involves developing a database of the IR-active species from the waste. Next, their limits of detection need to be quantified using IUPAC standards [77]. Understanding which species will be detected during the process is important for model development and training. The number of species that can be detected during the process will also influence the complexity of the spectra-to-composition model. For example, the data in Chapter 6 did not show any prominent peaks other than nitrate, nitrite and glycolate, however this hypothesis can only be confirmed by calculating the detection limits. As shown in Figure 7.6, the mixture spectra during SRAT/SME cycling change with the extent of the oxidation and reduction reactions, as well as dewatering and refluxing. Instead of only measuring a few points during the process, the ATR-FTIR probe can provide continuous process information that should be utilized.

Finally, if the limits of detection analysis confirms that most of the analytes are not measurable in the waste, easy-to-use PLSR model development in the Mettler Toledo software environment should be considered. However, since some of these species can appear in the waste even if they were not anticipated, it may also be beneficial to have a more complex algorithm, such as BSS-PLSR as an option. Another idea includes the development of a graphical user interface (GUI) that would have the capability to display the species concentrations from the spectroscopy data.

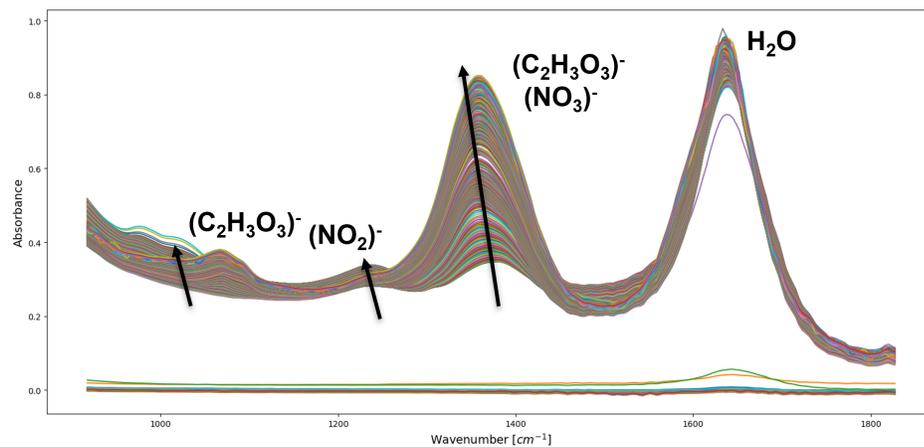


Figure 7.6: Simulated high-level waste spectra during SRAT and SME cycling.

# **Appendices**

**APPENDIX A**  
**PREDICTION ERRORS FOR BSS-PLSR**

**A.1 Prediction errors for Raman simulated data**

Mixture Index	PLSR only										BSS-PLSR				
	$\text{NO}_3^-$ (%)	$\text{NO}_2^-$ (%)	$\text{SO}_4^{2-}$ (%)	$\text{CO}_3^{2-}$ (%)	$\text{H}_2\text{O}$ (%)	$\text{NO}_3^-$ (%)	$\text{NO}_2^-$ (%)	$\text{SO}_4^{2-}$ (%)	$\text{CO}_3^{2-}$ (%)	$\text{H}_2\text{O}$ (%)	$\text{NO}_3^-$ (%)	$\text{NO}_2^-$ (%)	$\text{SO}_4^{2-}$ (%)	$\text{CO}_3^{2-}$ (%)	$\text{H}_2\text{O}$ (%)
1	1.7	6.8	7.2	20.2	0.0	1.9	1.2	4.8	22.3	0.0	1.2	4.8	22.3	0.0	
2	12.7	8.2	0.4	4.2	0.0	0.2	3.4	5.4	21.8	0.0	3.4	5.4	21.8	0.0	
3	38.4	13.9	49.1	4.2	0.0	0.1	3.1	2.3	15.6	0.0	3.1	2.3	15.6	0.0	
4	7.3	4.5	94.7	23.4	0.0	1.0	2.7	20.4	49.9	0.0	2.7	20.4	49.9	0.0	
5	35.3	8.5	127.1	35.9	0.0	4.3	4.1	28.4	66.6	0.0	4.1	28.4	66.6	0.0	
6	26.6	39.0	100.5	34.2	0.0	1.2	10.3	14.1	66.0	0.0	10.3	14.1	66.0	0.0	
7	29.4	39.5	115.7	5.2	0.0	4.5	11.3	23.6	38.3	0.0	11.3	23.6	38.3	0.0	
8	15.2	16.7	29.9	9.4	0.0	2.1	1.7	1.6	4.0	0.0	1.7	1.6	4.0	0.0	
9	1.3	13.1	5.9	6.7	0.0	1.1	6.0	2.7	5.8	0.0	6.0	2.7	5.8	0.0	
10	39.6	9.4	24.0	72.2	0.0	1.6	0.3	1.4	17.5	0.0	0.3	1.4	17.5	0.0	
11	11.3	1.3	79.7	66.0	0.0	0.1	3.4	6.4	9.8	0.0	3.4	6.4	9.8	0.0	
12	32.3	7.7	99.2	2.7	0.0	2.5	5.0	5.3	14.3	0.0	5.0	5.3	14.3	0.0	
13	18.9	5.6	176.0	7.8	0.0	1.7	3.4	18.5	14.5	0.0	3.4	18.5	14.5	0.0	
14	14.6	7.2	42.3	1.1	0.0	2.8	2.1	3.5	7.8	0.0	2.1	3.5	7.8	0.0	
15	0.8	0.8	53.0	22.0	0.0	4.0	0.6	10.4	1.4	0.0	0.6	10.4	1.4	0.0	
16	27.2	7.7	94.1	45.7	0.0	7.0	3.8	0.2	24.2	0.0	3.8	0.2	24.2	0.0	
17	17.8	4.8	25.6	4.8	0.0	4.2	3.3	2.9	15.3	0.0	3.3	2.9	15.3	0.0	
18	16.6	5.8	178.5	1.4	0.0	4.2	1.2	22.9	7.1	0.0	1.2	22.9	7.1	0.0	
19	29.5	3.7	79.8	26.2	0.0	1.0	1.3	3.0	17.0	0.0	1.3	3.0	17.0	0.0	
20	0.2	0.0	11.2	14.7	0.0	1.0	2.7	12.9	17.2	0.0	2.7	12.9	17.2	0.0	
21	31.6	11.2	100.7	0.5	0.0	3.0	2.5	5.8	14.0	0.0	2.5	5.8	14.0	0.0	
Average	<b>19.4</b>	<b>10.3</b>	<b>71.2</b>	<b>19.4</b>	<b>0.0</b>	<b>2.4</b>	<b>3.5</b>	<b>9.4</b>	<b>21.4</b>	<b>0.0</b>	<b>3.5</b>	<b>9.4</b>	<b>21.4</b>	<b>0.0</b>	

Table A.1: Absolute values of the percent errors for the simulated Raman data set using a 0th derivative in the PLSR model. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

Mixture Index	PLSR only							BSS-PLSR							
	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)
1	5.1	5.8	9.3	12.2	0.0	4.0	4.0	5.1	11.3	0.0	4.0	4.0	5.1	11.3	0.0
2	1.9	8.0	20.8	22.0	0.0	5.7	2.9	7.0	19.5	0.0	2.9	2.9	7.0	19.5	0.0
3	9.9	16.0	58.0	46.6	0.0	3.2	3.7	9.4	35.4	0.0	3.2	3.7	9.4	35.4	0.0
4	10.9	3.2	3.2	136.7	0.0	8.2	1.5	46.5	159.4	0.0	8.2	1.5	46.5	159.4	0.0
5	20.4	5.9	43.7	129.7	0.0	8.6	5.2	28.6	157.8	0.0	8.6	5.2	28.6	157.8	0.0
6	8.3	13.9	80.2	87.8	0.0	1.2	0.5	1.6	118.2	0.0	1.2	0.5	1.6	118.2	0.0
7	7.1	9.7	76.7	99.6	0.0	0.2	4.6	6.5	128.9	0.0	0.2	4.6	6.5	128.9	0.0
8	5.9	9.7	40.8	22.5	0.0	10.3	1.3	8.6	12.0	0.0	10.3	1.3	8.6	12.0	0.0
9	0.6	5.5	3.1	8.7	0.0	1.5	6.6	4.9	9.2	0.0	1.5	6.6	4.9	9.2	0.0
10	13.5	11.1	32.2	8.0	0.0	1.4	2.2	6.5	2.2	0.0	1.4	2.2	6.5	2.2	0.0
11	3.7	5.5	46.0	0.7	0.0	0.2	0.0	2.3	17.7	0.0	0.2	0.0	2.3	17.7	0.0
12	4.2	16.0	69.0	24.9	0.0	7.9	4.8	13.3	12.9	0.0	7.9	4.8	13.3	12.9	0.0
13	2.4	7.5	84.2	21.4	0.0	4.5	0.3	63.1	15.2	0.0	4.5	0.3	63.1	15.2	0.0
14	3.5	3.5	40.2	18.5	0.0	2.6	5.2	1.2	10.6	0.0	2.6	5.2	1.2	10.6	0.0
15	5.0	4.1	14.2	17.8	0.0	6.1	1.7	17.8	25.3	0.0	6.1	1.7	17.8	25.3	0.0
16	3.6	15.3	68.7	7.5	0.0	8.5	4.5	12.2	15.5	0.0	8.5	4.5	12.2	15.5	0.0
17	10.4	7.6	20.0	28.9	0.0	3.0	2.7	17.1	26.4	0.0	3.0	2.7	17.1	26.4	0.0
18	6.0	9.0	73.7	8.3	0.0	1.3	2.1	65.5	0.3	0.0	1.3	2.1	65.5	0.3	0.0
19	10.6	7.4	51.4	31.2	0.0	0.5	0.6	8.1	18.6	0.0	0.5	0.6	8.1	18.6	0.0
20	4.0	6.1	42.7	10.5	0.0	3.8	5.1	57.0	9.8	0.0	3.8	5.1	57.0	9.8	0.0
21	6.0	15.9	69.5	37.3	0.0	6.5	0.1	9.4	22.1	0.0	6.5	0.1	9.4	22.1	0.0
Average	<b>6.8</b>	<b>8.9</b>	<b>45.1</b>	<b>37.2</b>	<b>0.0</b>	<b>4.2</b>	<b>2.8</b>	<b>18.6</b>	<b>39.4</b>	<b>0.0</b>	<b>4.2</b>	<b>2.8</b>	<b>18.6</b>	<b>39.4</b>	<b>0.0</b>

Table A.2: Absolute values of the percent errors for the simulated Raman data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 7. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

Mixture Index	PLSR only										BSS-PLSR				
	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)
1	10.5	5.2	8.8	2.5	0.0	1.4	4.5	16.2	12.7	0.0					
2	25.0	12.7	94.2	111.1	0.0	0.4	1.6	9.6	2.1	0.0					
3	118.6	36.0	258.5	344.1	0.0	2.3	2.0	13.7	0.3	0.0					
4	77.6	43.3	122.3	220.0	0.0	8.9	7.8	22.2	34.8	0.0					
5	160.6	53.5	340.1	654.3	0.0	21.4	9.4	59.4	72.2	0.0					
6	92.6	80.6	392.3	598.3	0.0	1.5	9.7	67.4	13.8	0.0					
7	99.0	78.1	397.4	542.4	0.0	5.6	6.9	58.3	14.2	0.0					
8	50.6	36.6	137.2	188.2	0.0	4.8	3.0	6.2	27.4	0.0					
9	24.1	6.5	2.0	44.2	0.0	6.5	2.0	7.7	42.1	0.0					
10	109.3	19.8	213.1	315.3	0.0	1.7	10.2	10.7	33.3	0.0					
11	99.6	42.0	101.4	179.7	0.0	4.7	1.3	0.3	18.5	0.0					
12	116.7	33.0	387.4	329.8	0.0	1.3	3.4	37.8	3.6	0.0					
13	68.5	18.9	661.9	151.8	0.0	0.3	5.0	98.5	18.2	0.0					
14	59.1	26.1	201.1	219.9	0.0	0.1	1.3	12.3	0.8	0.0					
15	22.2	13.0	43.2	90.9	0.0	1.6	2.0	31.2	0.7	0.0					
16	107.6	33.8	389.9	567.1	0.0	10.3	0.7	25.4	74.0	0.0					
17	52.3	6.0	229.3	142.1	0.0	0.9	4.3	52.9	14.9	0.0					
18	69.7	18.8	689.0	237.4	0.0	1.5	2.8	23.6	15.1	0.0					
19	119.3	29.3	228.8	284.9	0.0	2.1	2.0	16.3	12.0	0.0					
20	5.8	6.1	15.7	7.6	0.0	2.9	2.4	66.5	8.6	0.0					
21	121.2	47.0	390.8	454.0	0.0	0.5	1.6	13.6	25.9	0.0					
Average	<b>76.7</b>	<b>30.8</b>	<b>252.6</b>	<b>270.7</b>	<b>0.0</b>	<b>3.8</b>	<b>4.0</b>	<b>30.9</b>	<b>21.2</b>	<b>0.0</b>					

Table A.3: Absolute values of the percent errors for the simulated Raman data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 27. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

## **A.2 Prediction errors for ATR-FTIR simulated data**

Mixture Index	PLSR only										BSS-PLSR				
	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)
1	1.0	8.3	9.3	2.5	0.0	4.6	6.1	9.1	7.5	0.0				0.0	
2	10.1	6.5	8.6	38.8	0.0	0.7	6.9	11.4	3.5	0.0				0.0	
3	24.0	36.3	183.2	49.9	0.0	9.2	3.0	11.0	25.2	0.0				0.0	
4	108.2	129.2	551.4	94.1	0.0	2.2	30.8	27.8	74.1	0.0				0.0	
5	91.4	121.7	544.0	13.2	0.0	21.6	20.9	25.7	98.2	0.0				0.0	
6	81.1	98.7	502.3	45.3	0.0	2.2	26.7	24.8	83.7	0.0				0.0	
7	82.2	91.4	538.6	55.2	0.0	2.1	18.7	30.5	67.0	0.0				0.0	
8	39.8	30.8	169.2	27.4	0.0	12.5	1.1	4.2	4.8	0.0				0.0	
9	1.6	9.3	11.3	12.5	0.0	5.2	10.3	13.4	11.2	0.0				0.0	
10	52.1	9.5	8.5	115.6	0.0	0.1	17.2	9.7	32.2	0.0				0.0	
11	134.3	120.6	506.2	44.6	0.0	7.9	10.3	17.6	8.2	0.0				0.0	
12	41.5	45.1	382.3	38.2	0.0	12.6	0.4	22.0	17.1	0.0				0.0	
13	22.0	28.0	661.8	25.4	0.0	3.9	6.0	73.0	6.9	0.0				0.0	
14	17.7	42.8	191.0	33.3	0.0	3.8	12.7	16.7	10.8	0.0				0.0	
15	41.6	41.1	321.5	28.0	0.0	5.0	10.2	36.4	21.3	0.0				0.0	
16	40.3	46.1	386.5	72.2	0.0	11.5	0.8	21.7	36.8	0.0				0.0	
17	21.4	0.9	5.1	35.9	0.0	1.1	7.6	31.8	7.4	0.0				0.0	
18	17.0	30.8	637.1	36.8	0.0	0.0	7.6	60.5	6.4	0.0				0.0	
19	63.3	63.3	391.6	15.4	0.0	0.3	6.1	19.8	10.9	0.0				0.0	
20	2.5	7.3	15.3	3.1	0.0	3.2	2.3	54.5	5.2	0.0				0.0	
21	27.1	81.9	387.7	65.8	0.0	1.4	7.8	17.3	11.2	0.0				0.0	
Average	<b>43.8</b>	<b>50.0</b>	<b>305.4</b>	<b>40.6</b>	<b>0.0</b>	<b>5.3</b>	<b>10.2</b>	<b>25.7</b>	<b>26.2</b>	<b>0.0</b>				<b>0.0</b>	

Table A.4: Absolute values of the percent errors for the simulated IR data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 5. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

### **A.3 Prediction errors for Raman experimental data**

Mixture Index	PLSR only										BSS-PLSR				
	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)
1	0.9	2.4	11.2	28.9	0.0	7.3	2.6	20.8	5.2	0.0				0.0	
2	47.5	0.6	74.3	203.6	0.0	4.0	3.4	11.1	19.1	0.0				0.0	
3	46.7	4.0	32.8	239.8	0.0	15.9	1.9	90.1	6.4	0.0				0.0	
4	2.0	10.3	257.5	64.9	0.0	14.1	0.1	107.4	92.3	0.0				0.0	
5	134.9	31.0	19.2	955.7	0.0	0.2	9.3	68.6	26.4	0.0				0.0	
6	130.2	16.5	4.2	879.3	0.0	8.5	0.2	34.0	158.8	0.0				0.0	
7	124.8	15.8	47.3	889.7	0.0	11.3	0.6	76.0	95.5	0.0				0.0	
8	58.4	1.5	32.0	292.8	0.0	19.5	5.3	71.8	14.3	0.0				0.0	
9	3.4	1.8	2.6	9.6	0.0	0.9	1.5	0.3	33.1	0.0				0.0	
10	168.5	4.0	231.5	861.1	0.0	6.1	5.6	25.4	150.7	0.0				0.0	
11	9.9	24.9	241.8	91.9	0.0	14.6	10.7	80.7	35.5	0.0				0.0	
12	107.6	5.3	18.9	531.3	0.0	15.2	8.8	196.1	62.8	0.0				0.0	
13	64.1	7.2	158.8	286.4	0.0	11.7	0.1	343.0	43.8	0.0				0.0	
14	50.7	5.2	9.1	376.8	0.0	14.3	3.7	134.8	48.9	0.0				0.0	
15	6.3	5.0	174.0	37.9	0.0	15.6	2.8	144.3	33.1	0.0				0.0	
16	104.5	11.6	37.9	1018.8	0.0	18.6	2.3	179.5	94.5	0.0				0.0	
17	80.3	6.7	244.7	254.3	0.0	0.4	10.2	18.2	48.4	0.0				0.0	
18	65.0	1.8	135.5	345.5	0.0	11.2	5.4	349.4	23.1	0.0				0.0	
19	85.2	16.6	60.1	376.6	0.0	11.6	4.4	110.0	30.1	0.0				0.0	
20	0.2	4.6	36.3	6.6	0.0	4.8	4.8	3.1	8.8	0.0				0.0	
21	104.8	16.6	46.2	680.4	0.0	16.1	5.6	153.7	68.1	0.0				0.0	
Average	<b>66.5</b>	<b>9.2</b>	<b>89.3</b>	<b>401.5</b>	<b>0.0</b>	<b>10.6</b>	<b>4.3</b>	<b>105.6</b>	<b>52.3</b>	<b>0.0</b>				<b>0.0</b>	

Table A.5: Absolute values of the percent errors for the experimental Raman data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 7. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

#### **A.4 Prediction errors for ATR-FTIR experimental data**

Mixture Index	PLSR only										BSS-PLSR				
	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)	NO <sub>3</sub> <sup>-</sup> (%)	NO <sub>2</sub> <sup>-</sup> (%)	SO <sub>4</sub> <sup>2-</sup> (%)	CO <sub>3</sub> <sup>2-</sup> (%)	H <sub>2</sub> O (%)
1	7.1	5.4	5.4	17.6	0.0	12.9	13.9	2.9	7.5	0.0	13.9	13.9	2.9	7.5	0.0
2	13.9	19.1	12.7	37.3	0.0	14.1	18.4	1.5	9.8	0.0	18.4	18.4	1.5	9.8	0.0
3	26.7	36.4	38.4	91.9	0.0	9.8	16.7	13.2	5.7	0.0	16.7	16.7	13.2	5.7	0.0
4	99.9	151.5	65.1	263.2	0.0	2.8	13.7	3.6	72.3	0.0	13.7	13.7	3.6	72.3	0.0
5	89.3	117.1	86.3	358.4	0.0	7.5	18.8	4.1	58.5	0.0	18.8	18.8	4.1	58.5	0.0
6	48.6	67.0	115.6	409.1	0.0	17.2	39.1	2.6	66.2	0.0	39.1	39.1	2.6	66.2	0.0
7	29.4	78.4	123.7	368.2	0.0	43.7	41.7	7.0	89.1	0.0	41.7	41.7	7.0	89.1	0.0
8	27.6	40.1	37.4	126.4	0.0	17.4	20.4	5.6	1.1	0.0	20.4	20.4	5.6	1.1	0.0
9	25.6	14.8	10.0	12.3	0.0	34.7	27.5	1.0	18.7	0.0	27.5	27.5	1.0	18.7	0.0
10	19.9	39.4	21.6	50.3	0.0	11.4	23.8	10.4	49.5	0.0	23.8	23.8	10.4	49.5	0.0
11	133.1	162.4	58.5	190.2	0.0	15.5	6.6	9.9	24.1	0.0	6.6	6.6	9.9	24.1	0.0
12	19.8	53.6	80.2	110.6	0.0	13.3	5.7	12.8	12.0	0.0	5.7	5.7	12.8	12.0	0.0
13	15.2	29.2	179.3	60.9	0.0	12.8	13.0	60.0	6.7	0.0	13.0	13.0	60.0	6.7	0.0
14	9.1	38.0	39.4	77.8	0.0	13.3	13.1	6.2	7.7	0.0	13.1	13.1	6.2	7.7	0.0
15	31.3	57.3	30.4	80.6	0.0	1.7	7.9	3.4	20.9	0.0	7.9	7.9	3.4	20.9	0.0
16	30.9	43.5	77.8	253.6	0.0	9.1	15.3	11.7	7.9	0.0	15.3	15.3	11.7	7.9	0.0
17	9.2	26.2	37.0	28.1	0.0	5.5	21.2	9.2	3.6	0.0	21.2	21.2	9.2	3.6	0.0
18	22.3	22.3	155.8	92.2	0.0	4.7	18.3	34.1	0.2	0.0	18.3	18.3	34.1	0.2	0.0
19	61.6	75.5	66.3	135.7	0.0	7.7	11.6	17.1	4.9	0.0	11.6	11.6	17.1	4.9	0.0
20	9.9	7.6	27.0	0.0	0.0	17.0	15.0	14.4	12.0	0.0	15.0	15.0	14.4	12.0	0.0
21	22.6	90.0	70.6	149.2	0.0	12.1	12.5	3.7	13.2	0.0	12.5	12.5	3.7	13.2	0.0
Average	<b>35.9</b>	<b>55.9</b>	<b>63.7</b>	<b>138.8</b>	<b>0.0</b>	<b>13.5</b>	<b>17.8</b>	<b>11.2</b>	<b>23.4</b>	<b>0.0</b>	<b>17.8</b>	<b>17.8</b>	<b>11.2</b>	<b>23.4</b>	<b>0.0</b>

Table A.6: Absolute values of the percent errors for the experimental IR data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 5. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

**APPENDIX B**  
**EXPERIMENTAL DATA FOR THE NITRATE SHIFT SYSTEM**

**B.1 Concentrations and spectra**

The data files used in the analysis have been uploaded on GitHub:

<https://github.com/skocevska/Nitrate-Shift-Data>.

In addition, the molal concentrations for each experiment are shown in Table B.1 – Table B.15. The spectra for each experiment are shown in Figure B.1 – Figure B.5.

Table B.1: Molal concentrations for Exp. 1 from Table 4.2.

<b>Exp. 1</b>				
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	0.68	0.68	0.68	0.68
<b>2</b>	1.35	1.35	1.35	1.35
<b>3</b>	2.70	2.70	2.70	2.70
<b>4</b>	4.06	4.06	4.06	4.06
<b>5</b>	5.41	5.41	5.41	5.41
<b>6</b>	6.76	6.76	6.76	6.76
<b>7</b>	8.11	8.11	8.11	8.11
<b>8</b>	9.47	9.46	9.47	9.46

Table B.2: Molal concentrations for Exp. 2 Table 4.2.

<b>Exp. 2</b>				
<b>Addition</b>	<b>KNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	0.07	0.07	0.07	0.07
<b>2</b>	0.55	0.55	0.55	0.55
<b>3</b>	1.09	1.09	1.09	1.09
<b>4</b>	2.19	2.19	2.19	2.19
<b>5</b>	3.28	3.28	3.28	3.28

Table B.3: Molal concentrations for Exp. 3 from Table 4.2.

<b>Exp. 3</b>				
<b>Addition</b>	<b>(AlNO<sub>3</sub>)<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	0.13	0.13	0.40	0.79
<b>2</b>	0.26	0.26	0.78	1.55
<b>3</b>	0.50	0.50	1.49	2.97
<b>4</b>	0.82	0.82	2.46	4.92
<b>5</b>	1.20	1.20	3.59	7.19

Table B.4: Molal concentrations for Exp. 4 from Table 4.2.

<b>Exp. 4</b>					
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>NaCl</b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	2.02	0.00	2.02	2.02	2.02
<b>2</b>	2.02	0.08	2.10	2.10	2.10
<b>3</b>	2.02	0.42	2.44	2.44	2.44
<b>4</b>	2.02	0.85	2.86	2.86	2.86
<b>5</b>	2.02	1.70	3.72	3.72	3.72
<b>6</b>	2.02	3.24	5.26	5.26	5.26
<b>7</b>	2.02	4.78	6.80	6.80	6.80

Table B.5: Molal concentrations for Exp. 5 from Table 4.2.

<b>Exp. 5</b>					
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>2</sub>SO<sub>4</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	2.03	0.00	2.03	2.03	2.03
<b>2</b>	2.03	0.03	2.09	2.06	2.12
<b>3</b>	2.03	0.10	2.22	2.13	2.32
<b>4</b>	2.03	0.99	4.01	3.02	5.00
<b>5</b>	2.03	1.79	5.61	3.82	7.40

Table B.6: Molal concentrations for Exp. 6 from Table 4.2.

<b>Exp. 6</b>					
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>NaNO<sub>2</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	2.03	0.00	2.03	2.03	2.03
<b>2</b>	2.03	0.57	2.61	2.61	2.61
<b>3</b>	2.03	1.15	3.18	3.18	3.18
<b>4</b>	2.03	2.29	4.33	4.33	4.33
<b>5</b>	2.03	3.44	5.47	5.47	5.47
<b>6</b>	2.03	5.73	7.77	7.77	7.77

Table B.7: Molal concentrations for Exp. 7 from Table 4.2.

<b>Exp. 7</b>					
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>2</sub>CO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	2.03	0.00	2.03	2.03	2.03
<b>2</b>	2.03	0.26	2.55	2.29	2.80
<b>3</b>	2.03	0.52	3.06	2.55	3.58
<b>4</b>	2.03	1.71	5.46	3.74	7.17
<b>5</b>	2.03	2.57	7.17	4.60	9.74

Table B.8: Molal concentrations for Exp. 8 from Table 4.2.

<b>Exp. 8</b>					
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>3</sub>PO<sub>4</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	2.03	0.00	2.03	2.03	2.03
<b>2</b>	2.03	0.01	2.06	2.04	2.10
<b>3</b>	2.03	0.05	2.17	2.07	2.31
<b>4</b>	2.03	0.27	2.84	2.30	3.65

Table B.9: Molal concentrations for Exp. 9 from Table 4.2.

<b>Exp. 9</b>						
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>NaNO<sub>2</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>2</sub>SO<sub>4</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	2.03	0.00	0.00	2.03	2.03	2.03
<b>2</b>	2.03	1.19	0.00	3.22	3.22	3.22
<b>3</b>	2.03	2.37	0.00	4.40	4.40	4.40
<b>4</b>	2.03	3.42	0.00	5.45	5.45	5.45
<b>5</b>	2.03	3.42	0.38	6.22	5.84	6.60
<b>6</b>	2.03	3.42	0.70	6.86	6.16	7.56
<b>7</b>	2.03	3.42	1.02	7.50	6.48	8.52

Table B.10: Molal concentrations for Exp. 10 from Table 4.2.

<b>Exp. 10</b>						
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>NaNO<sub>2</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>2</sub>CO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	2.03	0.00	0.00	2.03	2.03	2.03
<b>2</b>	2.03	3.43	0.00	5.46	5.46	5.46
<b>3</b>	2.03	3.43	0.52	6.49	5.98	7.01
<b>4</b>	2.03	3.43	1.03	7.52	6.49	8.55
<b>5</b>	2.03	3.43	1.46	8.38	6.92	9.84
<b>6</b>	2.03	3.43	1.89	9.24	7.35	11.13

Table B.11: Molal concentrations for Exp. 11 from Table 4.2.

<b>Exp. 11</b>						
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>2</sub>SO<sub>4</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>2</sub>CO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	2.03	0.00	0.00	2.03	2.03	2.03
<b>2</b>	2.03	1.67	0.00	5.37	3.70	7.03
<b>3</b>	2.03	1.67	0.40	6.17	4.10	8.24
<b>4</b>	2.03	1.67	0.77	6.90	4.47	9.34
<b>5</b>	2.03	1.67	1.13	7.63	4.83	10.42
<b>6</b>	2.03	1.67	1.56	8.49	5.26	11.72
<b>7</b>	2.03	1.67	1.92	9.21	5.62	12.80

Table B.12: Molal concentrations for Exp. 12 from Table 4.2.

<b>Exp. 12</b>						
<b>Addition</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>NaNO<sub>2</sub></b> [mol kg <sup>-1</sup> ]	<b>NaCl</b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]
<b>1</b>	2.03	0.00	0.00	2.03	2.03	2.03
<b>2</b>	2.03	3.43	0.00	5.46	5.46	5.46
<b>3</b>	2.03	3.43	0.08	5.54	5.54	5.54
<b>4</b>	2.03	3.43	0.43	5.89	5.89	5.89
<b>5</b>	2.03	3.43	0.86	6.32	6.32	6.32
<b>6</b>	2.03	3.43	1.71	7.18	7.18	7.17
<b>7</b>	2.03	3.43	3.27	8.73	8.73	8.73
<b>8</b>	2.03	3.43	3.89	9.35	9.35	9.35

Table B.13: Molal concentrations for Exp. 13 from Table 4.2.

Exp. 13							
Add.	NaNO <sub>3</sub> [mol kg <sup>-1</sup> ]	NaNO <sub>2</sub> [mol kg <sup>-1</sup> ]	Na <sub>2</sub> SO <sub>4</sub> [mol kg <sup>-1</sup> ]	Na <sub>2</sub> CO <sub>3</sub> [mol kg <sup>-1</sup> ]	C <sup>+</sup> [mol kg <sup>-1</sup> ]	C <sup>-</sup> [mol kg <sup>-1</sup> ]	I [mol kg <sup>-1</sup> ]
1	1.49	1.26	0.19	0.61	4.35	3.55	2.98
2	1.50	1.26	0.19	0.00	3.14	2.95	1.76
3	1.50	1.27	0.00	0.61	3.99	3.38	2.60
4	1.48	0.00	0.00	0.00	1.48	1.48	0.74
5	1.48	0.00	0.19	0.00	1.86	1.67	1.12
6	1.48	0.00	0.19	0.61	3.07	2.28	2.33
7	1.48	0.00	0.00	0.61	2.70	2.09	1.96
8	1.47	1.25	0.00	0.00	2.72	2.72	1.36
9	1.49	0.63	0.10	0.31	2.92	2.52	1.87
10	0.75	0.00	0.10	0.31	1.56	1.15	1.19
11	0.75	1.27	0.10	0.31	2.83	2.42	1.82
12	0.74	0.62	0.00	0.31	1.98	1.67	1.30
13	0.74	0.62	0.19	0.31	2.36	1.86	1.68
14	0.74	0.63	0.10	0.00	1.57	1.47	0.88
15	0.74	0.63	0.10	0.61	2.79	2.08	2.11
16	0.75	0.63	0.10	0.31	2.19	1.78	1.50
17	0.74	0.63	0.10	0.31	2.19	1.78	1.50
18	0.75	0.63	0.09	0.31	2.18	1.78	1.49
19	0.75	0.63	0.10	0.31	2.19	1.78	1.50
20	0.75	0.63	0.10	0.31	2.19	1.79	1.50
21	0.75	0.63	0.10	0.31	2.19	1.79	1.50
22	0.75	0.63	0.10	0.31	2.19	1.78	1.50
23	0.75	0.63	0.10	0.31	2.19	1.78	1.50
24	0.75	0.63	0.10	0.31	2.19	1.78	1.50
25	0.75	0.63	0.09	0.30	2.18	1.78	1.49
26	0.75	0.63	0.09	0.31	2.18	1.78	1.49
27	0.75	0.63	0.10	0.31	2.19	1.79	1.50

Table B.14: Molal concentrations for Exp. 14 from Table 4.2.

<b>Exp. 14</b>											
<b>Add.</b>	<b>NaNO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>NaNO<sub>2</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>2</sub>SO<sub>4</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>2</sub>CO<sub>3</sub></b> [mol kg <sup>-1</sup> ]	<b>Na<sub>3</sub>PO<sub>4</sub></b> [mol kg <sup>-1</sup> ]	<b>NaCH<sub>3</sub>COO</b> [mol kg <sup>-1</sup> ]	<b>Na<sub>2</sub>C<sub>2</sub>O<sub>4</sub></b> [mol kg <sup>-1</sup> ]	<b>C<sup>+</sup></b> [mol kg <sup>-1</sup> ]	<b>C<sup>-</sup></b> [mol kg <sup>-1</sup> ]	<b>I</b> [mol kg <sup>-1</sup> ]	
1	1.17	0.97	0.14	0.50	0.00	0.00	0.02	3.49	2.82	4.16	
2	1.17	0.97	0.14	0.50	0.00	0.08	0.02	3.55	2.88	4.22	
3	1.14	0.95	0.14	0.49	0.07	0.19	0.02	3.80	3.01	4.66	
4	0.54	0.39	0.05	0.12	0.07	0.00	0.00	1.48	1.17	1.86	
5	0.54	0.38	0.05	0.12	0.07	0.07	0.00	1.55	1.24	1.93	
6	0.54	0.38	0.05	0.12	0.07	0.07	0.05	1.65	1.29	2.08	
7	0.54	0.38	0.05	0.13	0.07	0.07	0.05	1.65	1.29	2.09	
8	0.96	0.85	0.14	0.41	0.07	0.07	0.05	3.30	2.56	4.11	
9	0.75	0.63	0.10	0.31	0.00	0.00	0.03	2.24	1.81	2.67	
10	0.73	0.63	0.10	0.25	0.00	0.14	0.00	2.21	1.86	2.56	
11	0.72	0.61	0.09	0.30	0.12	0.00	0.00	2.47	1.84	3.21	
12	1.39	1.12	0.10	0.57	0.10	0.19	0.01	4.34	3.47	5.30	
13	1.21	0.95	0.03	0.58	0.05	0.10	0.01	3.64	2.92	4.41	
14	1.43	0.75	0.10	0.44	0.05	0.10	0.00	3.50	2.86	4.19	
15	1.21	0.75	0.06	0.30	0.05	0.00	0.00	2.83	2.37	3.34	
16	1.41	1.13	0.10	0.29	0.10	0.19	0.01	3.82	3.23	4.51	
17	0.99	1.16	0.06	0.59	0.00	0.10	0.01	3.58	2.92	4.24	
18	1.20	0.95	0.03	0.44	0.05	0.10	0.01	3.35	2.78	3.97	
19	0.98	0.94	0.10	0.43	0.10	0.10	0.00	3.37	2.64	4.19	
20	1.00	1.17	0.03	0.45	0.00	0.00	0.01	3.15	2.66	3.63	
21	1.39	0.73	0.10	0.43	0.10	0.19	0.00	3.65	2.93	4.46	

Table B.15: Molal concentrations for Exp. 15 from Table 4.2.

Component	Exp. 15	
	LAW 1 [mol kg <sup>-1</sup> ]	LAW 6 [mol kg <sup>-1</sup> ]
Na <sup>+</sup>	5.790	5.640
NO <sub>3</sub> <sup>-</sup>	2.480	2.180
NO <sub>2</sub> <sup>-</sup>	1.170	1.280
CO <sub>3</sub> <sup>2-</sup>	0.564	0.471
OH <sup>-</sup>	1.391	1.590
Al(OH) <sub>4</sub> <sup>-</sup>	0.140	0.148
K <sup>+</sup>	0.234	0.214
SO <sub>4</sub> <sup>2-</sup>	0.048	0.055
Cl <sup>-</sup>	0.085	0.076
PO <sub>4</sub> <sup>3-</sup>	0.019	0.018
CrO <sub>4</sub> <sup>2-</sup>	0.007	0.008
C <sub>2</sub> O <sub>4</sub> <sup>2-</sup>	0.005	0.004
F <sup>-</sup>	0.025	0.024
TOC	0.147	0.147

## B.2 Testing Model Fit with the Akaike information criterion (AIC)

We evaluated the model performance based on the number of subpeaks used. The error was measured using Eq. Equation 4.4. In addition, the Akaike information criterion (AIC) was calculated:

$$AIC = n \log(SSE/n) + 2P \quad (\text{B.1})$$

where  $n$  represents the number of independent measurements (61 wavenumbers for each spectrum),  $SSE$  is the sum squared error and  $P$  represents the number of model parameters

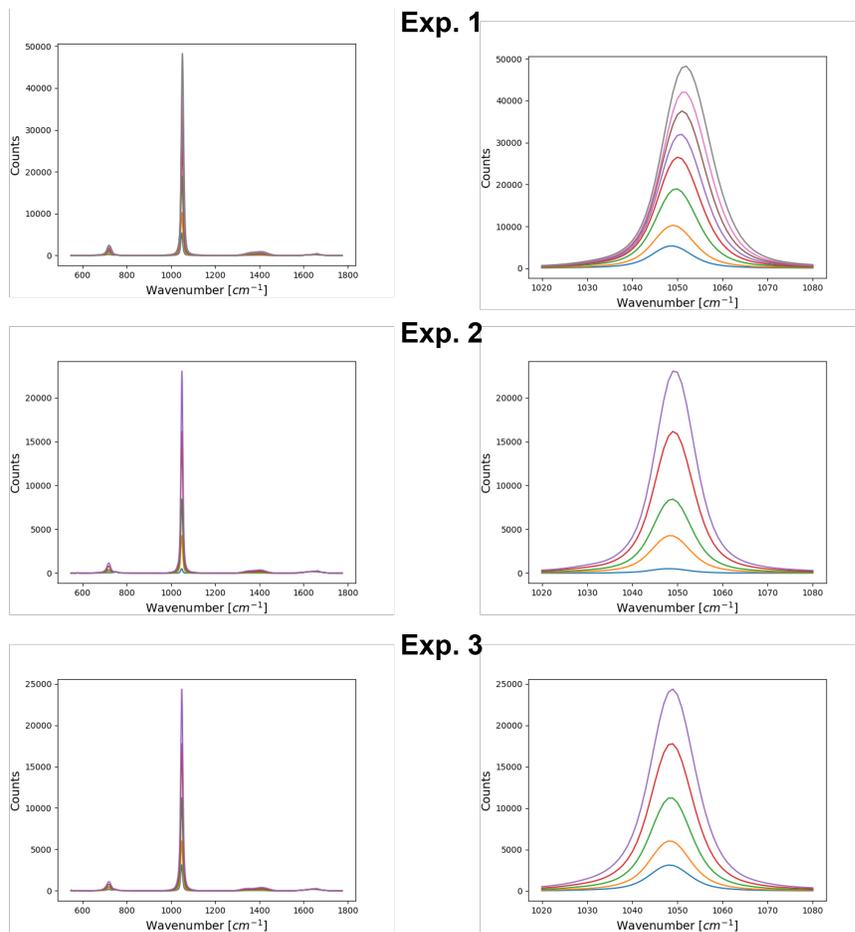
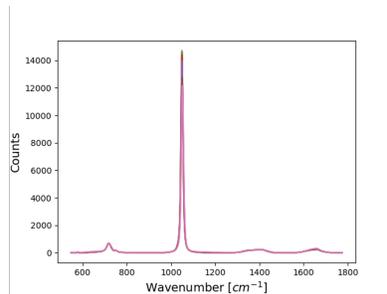
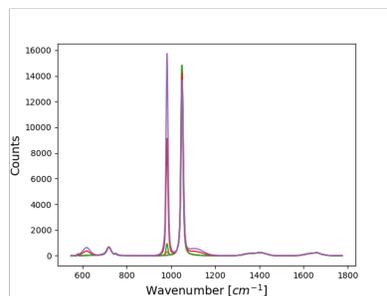
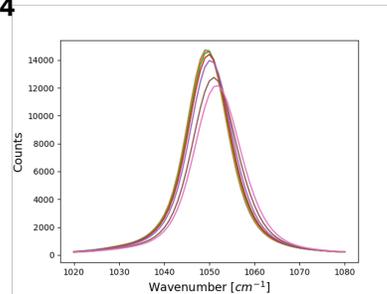


Figure B.1: Spectra for Exps. 1 – 3 in Table 4.2.

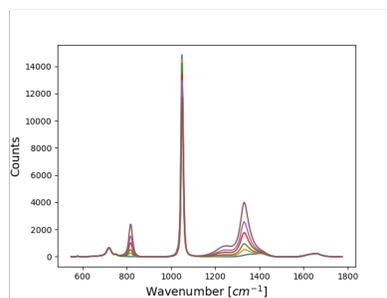
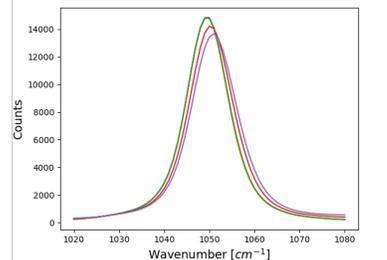
(subpeaks). The results are shown in Figures Figure B.6 and Figure B.7.



**Exp. 4**



**Exp. 5**



**Exp. 6**

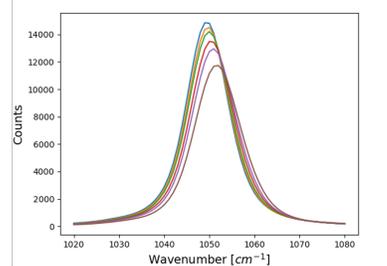
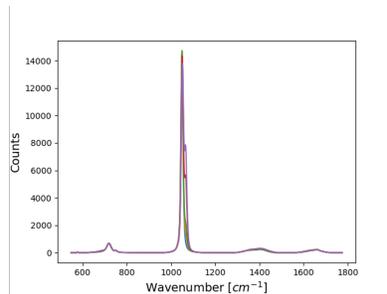
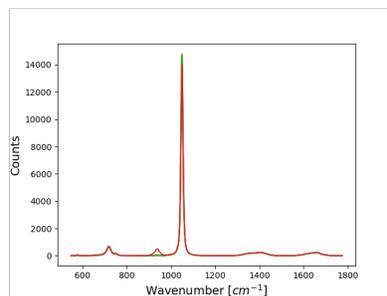
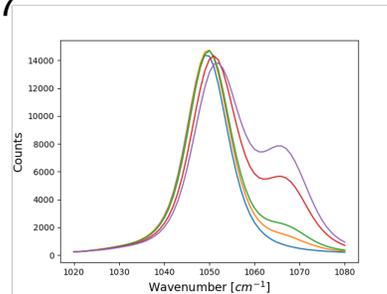


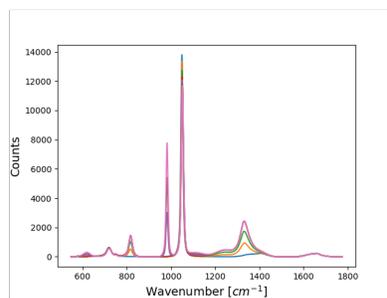
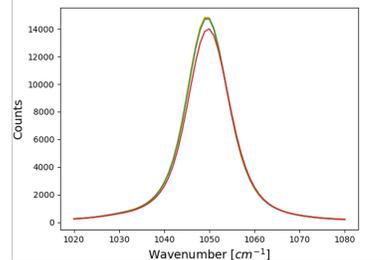
Figure B.2: Spectra for Exps. 4 – 6 in Table 4.2.



**Exp. 7**



**Exp. 8**



**Exp. 9**

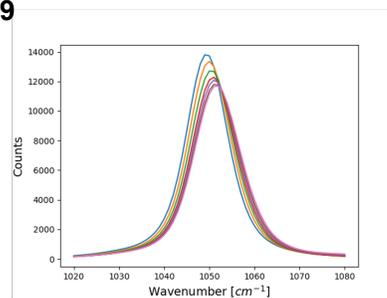


Figure B.3: Spectra for Exps. 7 – 9 in Table 4.2.

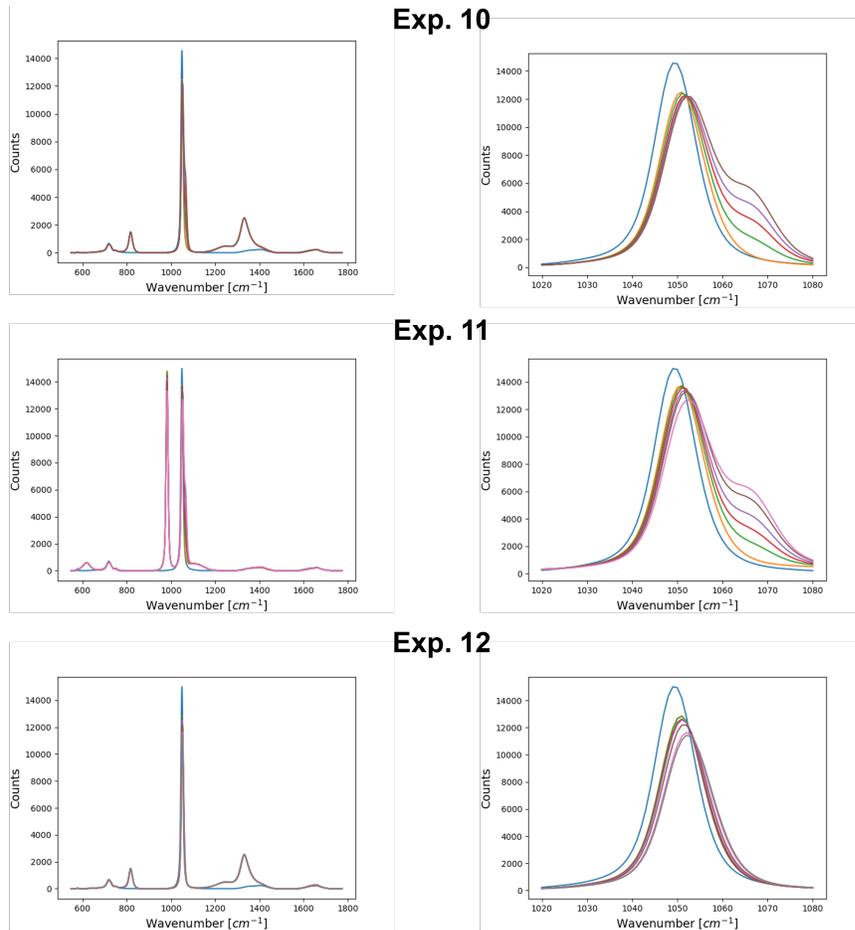


Figure B.4: Spectra for Exps. 10 – 12 in Table 4.2.

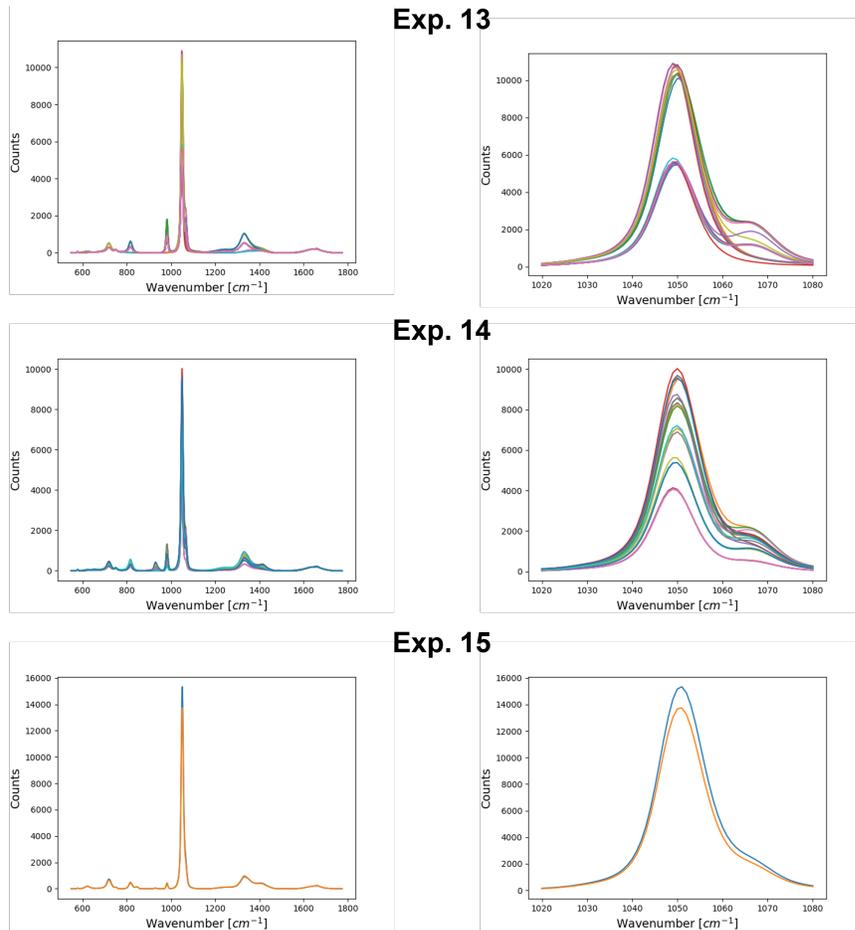


Figure B.5: Spectra for Exps. 13 – 15 in Table 4.2.

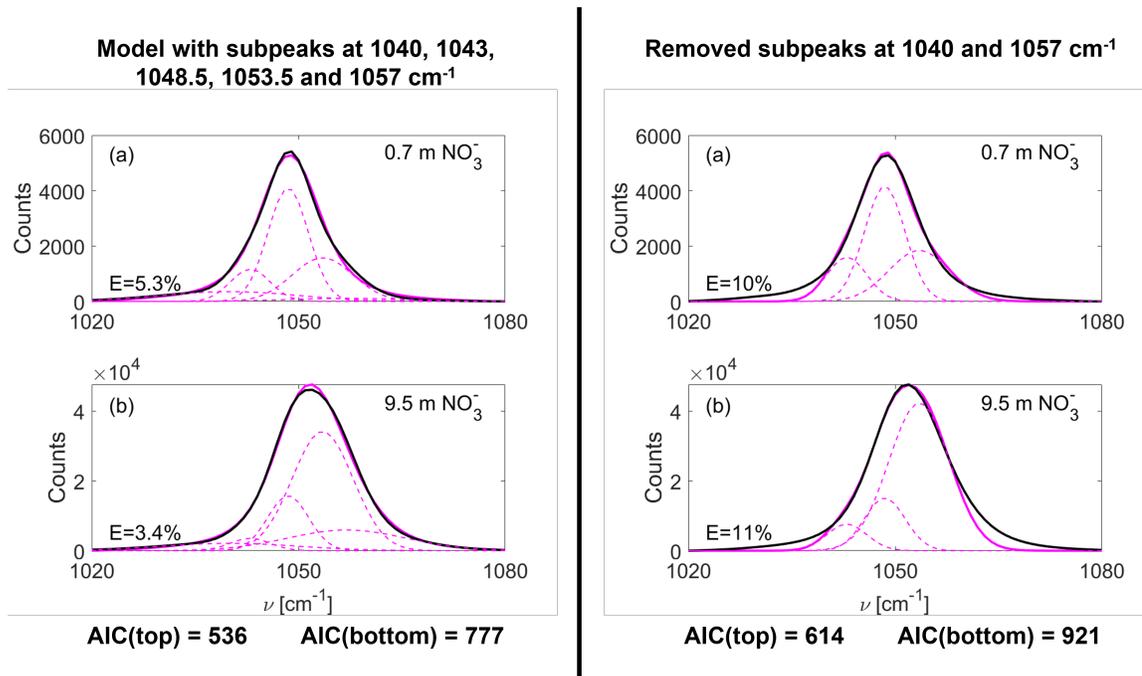


Figure B.6: Model fit for a low and high concentration of sodium nitrate from Exp. 1 in Table 4.2. A comparison was made between a model with five and three subpeaks.

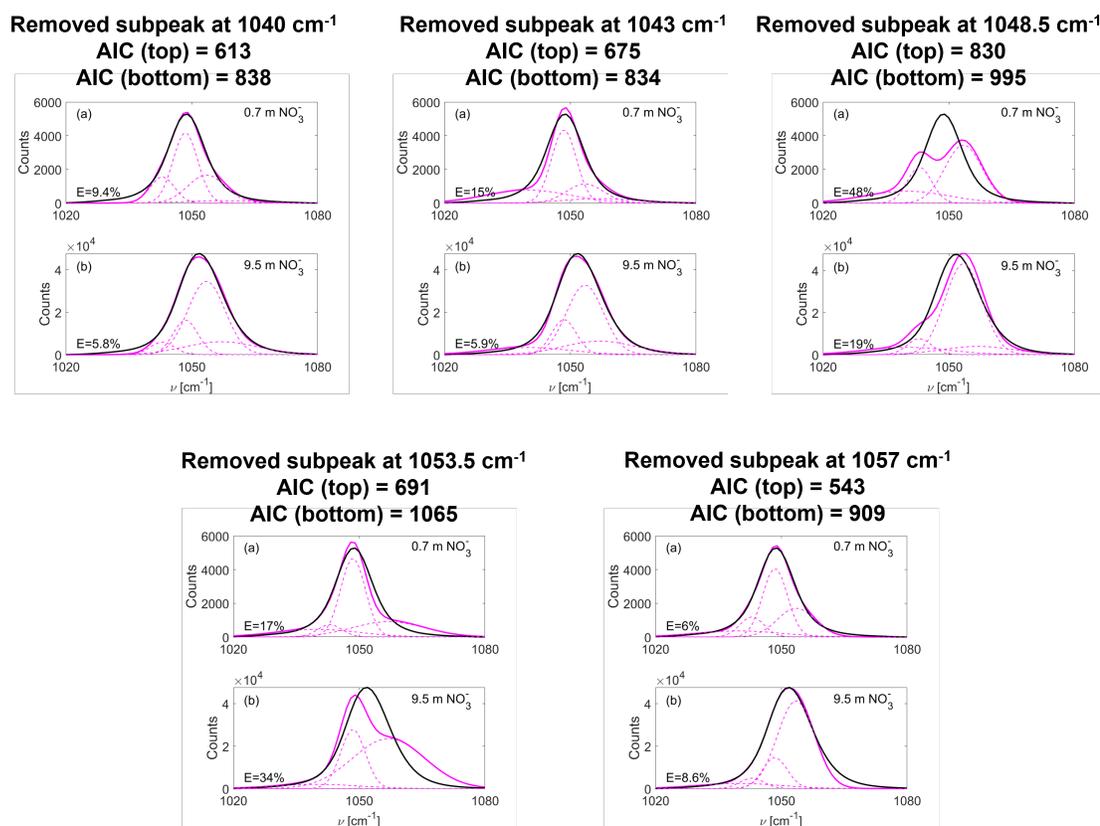


Figure B.7: Model fit for a low and high concentration of sodium nitrate from Exp. 1 in Table 4.2. The effect of removing one subpeak at a time was evaluated.

## REFERENCES

- [1] L. L. Simon *et al.*, “Assessment of Recent Process Analytical Technology (PAT) Trends: A Multiauthor Review,” *Organic Process Research and Development*, vol. 19, no. 1, pp. 3–62, 2015.
- [2] United States Government Accountability Office, “Program-Wide Strategy and Better Reporting Needed to Address Growing Environmental Cleanup Liability,” Tech. Rep., 2019.
- [3] W. R. Wilmarth *et al.*, “Review: Waste-Pretreatment Technologies for Remediation of Legacy Defense Nuclear Wastes,” *Solvent Extraction and Ion Exchange*, vol. 29, no. 1, pp. 1–48, 2011.
- [4] U.S. Department of Energy, “2019 Hanford Lifecycle Scope, Schedule and Cost Report,” Tech. Rep., 2019.
- [5] R. A. Peterson *et al.*, “Review of the Scientific Understanding of Radioactive Waste at the U.S. DOE Hanford Site,” *Environmental Science and Technology*, vol. 52, no. 2, pp. 381–396, 2018.
- [6] M. Spears, J. Eacker, M. Sturges, and B. M. Mauss, “Retrieval and Treatment of Hanford Tank Waste,” in *Waste Management Symposia*, 2006.
- [7] M. E. Stone, “Evaluation of a Material Balance Only and Material Balance with Real-Time In-line Monitoring Approaches for DFLAW Processing,” Savannah River National Laboratory, Tech. Rep., 2019.
- [8] M. E. Stone, C. C. Diprete, M. E. Farrar, A. M. Howe, F. R. Miera, and M. R. Poirier, “WTP Real-Time , In-Line Monitoring Program Task 2: Determine the Technical Basis for Process Control and Task 5: Process Control Challenges,” Savannah River National Laboratory, Tech. Rep., 2017.
- [9] M. R. Poirier, A. M. Howe, F. R. Miera, M. E. Stone, and C. C. Diprete, “WTP Real-Time In-Line Monitoring Program Task 1: LAW and EMF Analytes and Properties – Functional Requirements,” Savannah River National Laboratory, Tech. Rep., 2018.
- [10] M. E. Farrar, M. R. Poirier, C. C. Diprete, and M. E. Stone, “WTP Real-Time In-Line Monitoring Program: Task 3: LAW and EMF Flow Sheet Sampling Points,” Savannah River National Laboratory, Tech. Rep., 2018.
- [11] M. R. Poirier, A. M. Howe, F. R. Miera, M. E. Stone, C. C. DiPrete, and M. E. Farrar, “WTP Real-Time In-Line Monitoring Program Tasks 4 and 6: Data Quality and

- Management and Preliminary Analysis Plan,” Savannah River National Laboratory, Tech. Rep., 2017.
- [12] P. J. Brulotte and K. C. Christensen, “Tri-party agreement databases, access mechanism and procedures,” Oct. 1993.
- [13] T. T. TRAN, “Best basis inventory maintenance tool (bbim) database description and user guide,” 2000.
- [14] G. J. Lumetta, J. C. Braley, J. M. Peterson, S. A. Bryan, and T. G. Levitskaia, “Separating and Stabilizing Phosphate from High-Level Radioactive Waste: Process Development and Spectroscopic Monitoring,” *Environmental Science and Technology*, vol. 46, no. 11, pp. 6190–6197, 2012.
- [15] S. Kocevskaja, R. W. Rousseau, M. A. Grover, and G. M. Maggioni, “Multivariate Linear Models for In Situ Monitoring of Low-Activity Nuclear Waste Using Infrared and Raman Spectroscopy,” in *Waste Management Symposia*, 2020.
- [16] J. D. Vienna, D.-S. Kim, I. S. Muller, G. F. Piepel, and A. A. Kruger, “Toward Understanding the Effect of Low-Activity Waste Glass Composition on Sulfur Solubility,” *Journal of the American Ceramic Society*, vol. 97, no. 10, pp. 3135–3142, 2014.
- [17] D. J. Griffin, M. A. Grover, Y. Kawajiri, and R. W. Rousseau, “Robust Multicomponent IR-to-Concentration Model Regression,” *Chemical Engineering Science*, vol. 116, pp. 77–90, 2014.
- [18] J. D. Roberts and M. C. Caserio, *Basic principles of organic chemistry*. WA Benjamin, Inc., 1977.
- [19] Z. Li, M. J. Deen, S. Kumar, and P. R. Selvaganapathy, “Raman spectroscopy for in-line water quality monitoring — Instrumentation and potential,” *Sensors (Switzerland)*, vol. 14, no. 9, pp. 17 275–17 303, 2014.
- [20] E. V. Thomas and D. M. Haaland, “Comparison of multivariate calibration methods for quantitative spectral analysis,” *Analytical Chemistry*, vol. 62, no. 10, pp. 1091–1099, 1990.
- [21] J. N. Miller and J. C. Miller, *Statistics and Chemometrics for Analytical Chemistry*. 2010, ISBN: 9780273730422.
- [22] T. O’Haver, *Pragmatic Introduction to Signal Processing*. 2021.
- [23] I. T. Jolliffe, “Principal Component Analysis. Second Edition,” *Springer Series in Statistics*, vol. 98, p. 487, 2002. eprint: arXiv:1011.1669v3.

- [24] D. M. Haaland and E. V. Thomas, "Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information," *Analytical Chemistry*, vol. 60, no. 11, pp. 1193–1202, 1988.
- [25] P. Geladi and B. R. Kowalski, "Partial Least-Squares Regression - A Tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986. eprint: arXiv:1011.1669v3.
- [26] S. Wold, M. Sjöström, and L. Eriksson, "PLS-Regression: a Basic Tool of Chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [27] Kevin Dunn, "PID: Process Improvement using Data," no. Release 474-e7a3, pp. 323–414, 2019.
- [28] R. L. Russell, P. P. Schonewill, and C. A. Burns, "Simulant Development for LAWPS Testing," Tech. Rep., 2017.
- [29] M. J. Siegfried and M. E. Stone, "LAW Simulant Recipes for Evaluation of Real-Time, In-Line Monitoring Instruments," Savannah River National Laboratory, Tech. Rep., 2020.
- [30] G. Hanrahan and K. Lu, "Application of Factorial and Response Surface Methodology in Modern Experimental Design and Optimization," *Critical Reviews in Analytical Chemistry*, vol. 36, no. 3-4, pp. 141–151, 2006.
- [31] T. Rakić, I. Kasagić-Vujanović, M. Jovanović, B. Jančić-Stojanović, and D. Ivanović, "Comparison of Full Factorial Design, Central Composite Design, and Box-Behnken Design in Chromatographic Method Development for the Determination of Fluconazole and Its Impurities," *Analytical Letters*, vol. 47, no. 8, pp. 1334–1347, 2014.
- [32] V. Mazet, C. Carteret, D. Brie, J. Idier, and B. Humbert, "Background Removal from Spectra by Designing and Minimising a Non-Quadratic Cost Function," *Chemometrics and Intelligent Laboratory Systems*, vol. 76, no. 2, pp. 121–133, 2005.
- [33] S. Kocevská, G. M. Maggioni, R. W. Rousseau, and M. A. Grover, "Spectroscopic Quantification of Target Species in a Complex Mixture Using Blind Source Separation and Partial Least-Squares Regression: A Case Study on Hanford Waste," *Industrial & Engineering Chemistry Research*, vol. 60, no. 27, pp. 9885–9896, 2021.
- [34] A. M. Lines *et al.*, "Online, Real-Time Analysis of Highly Complex Processing Streams: Quantification of Analytes in Hanford Tank Sample," *Industrial and Engineering Chemistry Research*, vol. 58, no. 47, pp. 21 194–21 200, 2019.

- [35] S. Bryan, T. Levitskaia, and S. Schlahta, "Raman based Process Monitor for Continuous Real-Time Analysis Of High Level Radioactive Waste Components," in *Waste Management Symposia*, 2008.
- [36] Y. B. Monakhova and D. N. Rutledge, "Independent Components Analysis (ICA) at the "Cocktail-Party" in Analytical Chemistry," *Talanta*, vol. 208, no. October 2019, p. 120 451, 2019.
- [37] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
- [38] Y. B. Monakhova and S. P. Mushtakova, "Multicomponent Quantitative Spectroscopic Analysis without Reference Substances based on ICA Modelling," *Analytical and Bioanalytical Chemistry*, vol. 409, pp. 3319–3327, 2017.
- [39] Y. B. Monakhova, S. A. Astakhov, A. Kraskov, and S. P. Mushtakova, "Independent Components in Spectroscopic Analysis of Complex Mixtures," *Chemometrics and Intelligent Laboratory Systems*, vol. 103, pp. 108–115, 2010.
- [40] G. M. Maggioni, S. Kocevská, M. A. Grover, and R. W. Rousseau, "Analysis of Multicomponent Ionic Mixtures Using Blind Source Separation : A Processing Case Study," *Industrial & Engineering Chemistry Research*, vol. 58, pp. 22 640–22 651, 2019.
- [41] S. Gajaraj, C. Fan, M. Lin, and Z. Hu, "Quantitative Detection of Nitrate in Water and Wastewater by Surface-Enhanced Raman Spectroscopy," *Environmental Monitoring and Assessment*, vol. 185, pp. 5673–5681, 2013.
- [42] A. Ianoul, T. Coleman, and S. A. Asher, "UV Resonance Raman Spectroscopic Detection of Nitrate and Nitrite in Wastewater Treatment Processes," *Analytical Chemistry*, vol. 74, no. 6, pp. 1458–1461, 2002.
- [43] H. Shi and P. Yu, "Comparison of Grating-based Near-Infrared (NIR) and Fourier Transform Mid-Infrared (ATR-FT/MIR) Spectroscopy based on Spectral Preprocessing and Wavelength Selection for the Determination of Crude Protein and Moisture Content in Wheat," *Food Control*, vol. 82, pp. 57–65, 2017.
- [44] J. Lee, J. Flores-Cerrillo, J. Wang, and Q. P. He, "Consistency-Enhanced Evolution for Variable Selection Can Identify Key Chemical Information from Spectroscopic Data," *Industrial and Engineering Chemistry Research*, vol. 59, no. 8, pp. 3446–3457, 2020.
- [45] E. A. Kyser, P. E. O'Rourke, and R. Lascola, "A Piecewise Local Partial Least Squares (PLS) Method for the Quantitative Analysis of Plutonium Nitrate Solu-

- tions,” *Applied Spectroscopy*, Vol. 71, Issue 12, pp. 2579-2594, vol. 71, no. 12, pp. 2579–2594, 2017.
- [46] A. De Juan, J. Jaumot, and R. Tauler, “Multivariate Curve Resolution (MCR). Solving the Mixture Analysis Problem,” *Analytical Methods*, vol. 6, no. 14, pp. 4964–4976, 2014.
- [47] C. H. Camp, “PyMCR: A Python Library for Multivariate Curve Resolution Analysis With Alternating Regression (MCR-AR),” *Journal of Research of the National Institute of Standards and Technology*, vol. 124, no. 124018, pp. 1–10, 2019.
- [48] “Corruption of the Pearson correlation Coefficient by Measurement Error and its Estimation, Bias, and Correction under Different Error Models,” *Scientific Reports*, vol. 10, no. 1, pp. 1–19, 2020.
- [49] A. Asfaram, M. Ghaedi, G. R. Ghezelbash, and F. Pepe, “Application of Experimental Design and Derivative Spectrophotometry Methods in Optimization and Analysis of Biosorption of Binary Mixtures of Basic Dyes from Aqueous Solutions,” *Ecotoxicology and Environmental Safety*, vol. 139, no. January, pp. 219–227, 2017.
- [50] J. Moros, S. Garrigues, and M. de la Guardia, “Vibrational Spectroscopy Provides a Green Tool for Multi-Component Analysis,” *TrAC - Trends in Analytical Chemistry*, vol. 29, no. 7, pp. 578–591, 2010.
- [51] A. G. Miller and J. A. Macklin, “Matrix Effects on the Raman Analytical Lines of Oxyanions,” *Analytical Chemistry*, vol. 52, no. 6, pp. 807–812, 1980.
- [52] J. Y. Yu, Y. Zhang, S. H. Tan, Y. Liu, and Y. H. Zhang, “Observation on the Ion Association Equilibria in NaNO<sub>3</sub> Droplets using Micro-Raman Spectroscopy,” *Journal of Physical Chemistry B*, vol. 116, no. 41, pp. 12 581–12 589, 2012.
- [53] S. Kocevskaja, G. M. Maggioni, S. H. Crouse, R. Prasad, R. W. Rousseau, and M. A. Grover, “Effect of ion interactions on the Raman spectrum of NO<sub>3</sub><sup>−</sup>: Toward monitoring of low-activity nuclear waste at Hanford,” *Chemical Engineering Research and Design*, vol. 181, no. 1, pp. 173–194, 2022.
- [54] P. M. Vollmar, “Ionic Interactions in Aqueous Solution: A Raman Spectral Study,” *The Journal of Chemical Physics*, vol. 39, no. 9, pp. 2236–2248, 1963.
- [55] L. V. Volod’ko and L. T. Huoah, “The Vibrational Spectra of Aqueous Nitrate Solutions,” *Journal of Applied Spectroscopy*, vol. 9, no. 4, pp. 1100–1104, 1968.
- [56] Y. Marcus and G. Hefter, “Ion Pairing,” *Chemical Reviews*, vol. 106, no. 11, pp. 4585–4621, 2006.

- [57] R. L. Frost and D. W. James, "Ion–Ion–Solvent Interactions in Solution. Part 3.—Aqueous Solutions of Sodium Nitrate," *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, vol. 78, no. 1, pp. 3223–3234, 1982.
- [58] D. C. Harris and M. D. Bertolucci, *Symmetry and Spectroscopy: An Introduction to Vibrational and Electronic Spectroscopy*. 1989, ISBN: 048666144X.
- [59] R. L. Frost and D. W. James, "Ion–Ion–Solvent Interactions in Solution. Part 5.—Influence of Added Halide, Change in Temperature and Solvent Deuteration on Ion Association in Aqueous Solutions of Nitrate Salts," *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, vol. 78, no. 11, pp. 3249–3261, 1982.
- [60] W. W. Rudolph, D. Fischer, and G. Irmer, "Hydration and Ion-Pair Formation of NaNO<sub>3</sub>(aq): A Vibrational Spectroscopic and Density Functional Theory Study," *Applied Spectroscopy*, vol. 75, no. 4, pp. 395–411, 2021.
- [61] D. Watanabe and H. O. Hamaguchi, "Ion Association Dynamics in Aqueous Solutions of Sulfate Salts as Studied by Raman Band Shape Analysis," *Journal of Chemical Physics*, vol. 123, no. 3, 2005.
- [62] I. T. Rusli, G. L. Schrader, and M. A. Larson, "Raman Spectroscopic Study of NaNO<sub>3</sub> Solution System - Solute Clustering in Supersaturated Solutions," *Journal of Crystal Growth*, vol. 97, no. 2, pp. 345–351, 1989.
- [63] C. T. Johnston *et al.*, "Raman study of aluminum speciation in simulated alkaline nuclear waste," *Environmental Science and Technology*, vol. 36, no. 11, pp. 2451–2458, 2002.
- [64] W. W. Rudolph, R. Mason, and C. C. Pye, "Aluminium(III) Hydration in Aqueous Solution. A Raman Spectroscopic Investigation and An Ab Initio Molecular Orbital Study of Aluminium(III) Water Clusters," *Physical Chemistry Chemical Physics*, vol. 2, no. 22, pp. 5030–5040, 2000.
- [65] H. A. Colburn and R. A. Peterson, "A history of Hanford tank waste, implications for waste treatment, and disposal," *Environmental Progress and Sustainable Energy*, vol. 40, no. 1, 2021.
- [66] P. Tse, J. Shafer, S. A. Bryan, and A. M. Lines, "Quantification of Raman-Interfering Polyoxoanions for Process Analysis: Comparison of Different Chemometric Models and a Demonstration on Real Hanford Waste," *Environmental Science & Technology*, vol. 55, no. 19, pp. 12 943–12 950, 2021.

- [67] S. Kocevskaja, G. M. Maggioni, T. Jones, M. Grover, and R. Rousseau, "Species Interactions in Liquid Low-Activity Waste Simulant Mixtures: Effect on the Raman Signal of Nitrate," in *Waste Management Symposia*, 2021.
- [68] S. Kocevskaja, S. H. Crouse, R. Prasad, G. M. Maggioni, M. A. Grover, and R. W. Rousseau, "Measuring Liquid Low Activity Waste Simulants with Process Analytical Technology: Real-Time In-Line Monitoring at Hanford," in *Waste Management Symposia*, 2022.
- [69] L. X. Yu *et al.*, "Applications of Process Analytical Technology to Crystallization Processes," *Advanced Drug Delivery Reviews*, vol. 56, no. 3, pp. 349–369, 2004.
- [70] T. Togkalidou, M. Fujiwara, S. Patel, and R. D. Braatz, "Solute Concentration Prediction using Chemometrics and ATR-FTIR Spectroscopy," *Journal of Crystal Growth*, vol. 231, no. 4, pp. 534–543, 2001.
- [71] J. Ray, B. Culbertson, S. Marra, and M. Plodinec, "DWPF Glass Product Control Program (U)," Tech. Rep., 2018.
- [72] D. P. Lambert, M. S. Williams, C. H. Brandenburg, and J. D. Newell, "Sludge Batch 9 Simulant Runs Using the Nitric-Glycolic Acid Flowsheet," Savannah River National Laboratory, Tech. Rep. November, 2016.
- [73] D. P. Lambert, A. D. Nikolov, D. T. Wasan, M. S. Williams, A. M. Howe, and W. H. Woodham, "A novel defoamer for processing nuclear waste: Testing and performance," *Environmental Progress and Sustainable Energy*, vol. 40, no. 4, pp. 1–13, 2021.
- [74] D. P. Lambert and A. M. Howe, "Antifoam Development for Eliminating Flammability Hazards and Decreasing Cycle Time in the Defense Waste Processing Facility," Savannah River National Laboratory, Tech. Rep. May, 2021.
- [75] A. M. Howe and M. J. Siegfried, "Sludge Batch 10 Flowsheet Testing with Non-radioactive Simulants," Savannah River National Laboratory, Tech. Rep. September, 2021.
- [76] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Transactions on Electronic Computers*, vol. EC-14, no. 3, pp. 326–334, Jun. 1965.
- [77] G. L. Long and J. D. Winefordner, "Limit of Detection: A Closer Look at the IUPAC Definition," *Analytical Chemistry*, vol. 55, no. 7, 712A–724A, 1983.