A Methodology for Monitoring Concentrations of Complex Waste Solutions

Daniel Griffin Martha Grover

Yoshiaki Kawajiri Ronald Rousseau

Published in Proceedings of the Waste Management Conference, Phoenix, AR, Mar. 2-6, 2014.

Abstract

In this contribution, a methodology for calibrating infrared absorbance readings to solution composition is presented. While this methodology has been specifically developed to track in-solution solute concentration profiles during crystallizations, it can be applied more generally to establish on line solution composition analysis. In particular, the calibration method is expected to be of use in establishing robust *in situ* measurement techniques for characterizing high salt content nuclear waste solutions.

Contents

1 Introduction

Much of the liquid nuclear waste stored at Hanford is saturated with solutes. Handling this waste may therefore result in unwanted crystallization and solid precipitation that disrupts downstream operations. The solutes of concern are predominantly sodium salts with the following anions: NO_3^-, NO_2^-, SO_4^{2-}, CO_3^{2-}, OH^-, $Al(OH)_4^-$, CrO_4^{2-}, and PO_4^{3-} [1-3]. To develop and run waste treatment operations successfully, it is important to establish automated, *in situ* measurement techniques for monitoring the concentration of these solutes in solution [4, 5].

Spectroscopic measurements can be made rapidly and in situ and, accordingly, provide a potential route for establishing automated, real-time monitoring. However, accurately calibrating spectroscopic measurements to give the concentration of multiple solution components can be difficult. Despite challenges associated with multicomponent calibration, Bryan et al. [6] demonstrated the ability to accurately monitor the concentrations of NO_3^-, NO_2^-, SO_4^{2-}, CO_3^{2-}, OH^-, $Al(OH)_4^-$, CrO_4^{2-}, and PO_4^{3-} in solution using a Raman-based monitoring system. Although this system performed well for single phase solutions, the presence of solids significantly degrades the measured Raman intensity. As we are interested in monitoring solutions with potentially high solid contents, an alternative technique was required.

The development of attenuated total reflectance Fourier-transform infrared (ATR-FTIR) technology has made it possible to measure the infrared absorbance spectrum of a solution even when solids are suspended in that solution [7, 8]. Although not all of the anions in nuclear waste absorb infrared light, in this study we have limited our attention to monitoring four anions that are likely targets for fractional crystallization of nuclear waste (NO_3^-, NO_2^-, SO_4^{2-}, CO_3^{2-})—each of these anions strongly absorbs infrared light.

Even for this reduced set of anions, developing a calibration model that accurately predicts the concentration of each anion simultaneously is a challenging task, with difficulties arising to due to overlap of the characteristic IR absorbance peaks and a number of disturbance variables. To overcome these challenges we have constructed a tailored calibration methodology that utilizes a novel regression algorithm, termed robust parameter support vector regression (RPSVR). This paper briefly describes the developed

methodology, examines the accuracy of established calibration models, and compares the performance of RPSVR with more common linear regression algorithms like partial least-squares regression (PLSR).

2 Calibrating Infrared Measurements

2.1 IR-to-Concentration Calibration Models

Infrared absorbance measurements can be used to identify the type and amount of molecules present. However, to derive quantitative concentration information from infrared absorbance spectra, a calibration model is needed.

In many cases a single spectrum feature (typically "peak height" or "peak area") can be correlated with the solute concentration in a straightforward manner. In this case, however, multiple infrared-absorbing anions are present in the solutions of interest and the characteristic infrared absorbance peaks for these anions overlap (Fig. 1).

To capture the intricacies of the convoluted infrared spectrum required to accurately predict the concentration of each anion, a feature vector is required. Correlating this feature vector to composition requires multivariate regression techniques and a large, well-formed training data set. The following sections present the methodology we employed to build accurate and robust calibration models that predict the concentrations of multiple solutes in a complex electrolytic solution.

2.2 Linear Multicomponent IR-to-Concentration Calibration Models: Notation

Before presenting the calibration methodology, some notation needs to be introduced and the calibration model defined.

Let $\mathbf{IR} \in \mathbf{R}^m$ represent the digitized infrared absorbance spectrum:

$$\mathbf{IR} \equiv [a_1 \dots a_l \dots a_m] \tag{1}$$

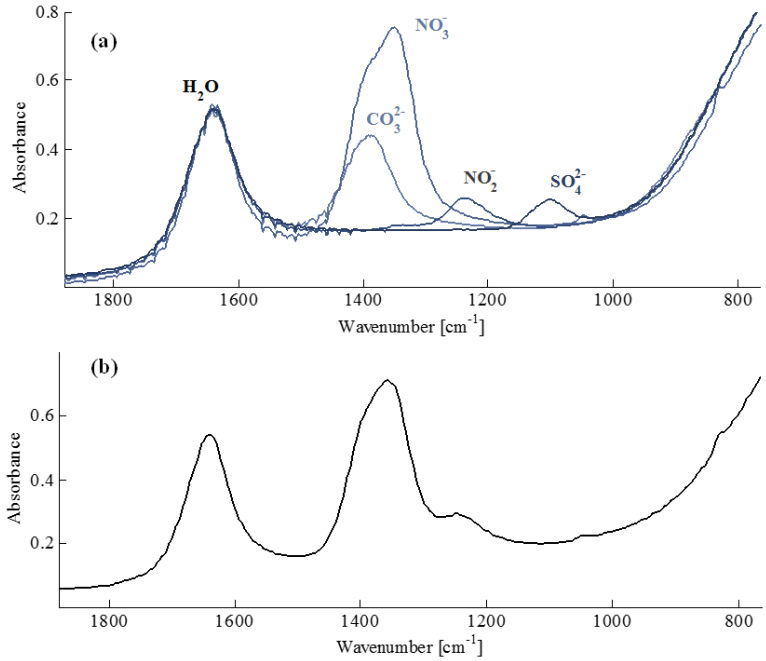

Figure 1: (a) Infrared spectra for single-solute aqueous solutions containing sodium carbonate, sodium nitrate, sodium sulfate, and sodium nitrate; and (b) the infrared spectrum of a complex electrolytic solution containing each anion.

Furthermore let $\mathbf{x} \in \mathbf{R}^q$ represent the input feature vector (constructed from the digitized infrared absorbance spectrum for a select range of frequencies):

$$\mathbf{x}^T \equiv [1 \; \mathbf{IR}_{\text{select}}] \tag{2}$$

Finally, let $\mathbf{y} \in \mathbf{R}^p$ represent the solution composition:

$$\mathbf{y}^T \equiv \; [C_{\text{CO}_3^{2-}} \quad C_{\text{NO}_3^-} \quad C_{\text{SO}_4^{2-}} \quad C_{\text{NO}_2}] \tag{3}$$

In Equation (1) a_l is the measured infrared absorbance (l denotes the frequency). In Equation (2) the constant 1 is added as the first input feature to include a constant term in the calibration model, $\mathbf{IR}$ is the digitized infrared absorbance spectrum for frequencies between 764 cm^{-1} and 1507 cm^{-1} and q is the dimension of $\mathbf{x}$ (201 in this study). Finally, in Equation (3), C_j is the concentration of the of the jth anion and p is the number of anions

monitored (4 here).

Using this notation, a linear calibration model can be defined as follows:

$$\mathbf{g}(\mathbf{x}) \equiv \mathbf{x}^T\mathbf{B} = \hat{\mathbf{y}} \approx \mathbf{y} \tag{4}$$

where $\mathbf{g}$ is the vector-valued calibration model, $\mathbf{B} \in \mathbf{R}^{q \times p}$ contains the model parameters, and $\hat{\mathbf{y}}$ is the model-predicted concentration vector.

2.3 Calibration Methodology

There are two basic steps in building a calibration model: collect data to compose a training data set and use a regression algorithm to fit a calibration model to the training data (Fig. 2).

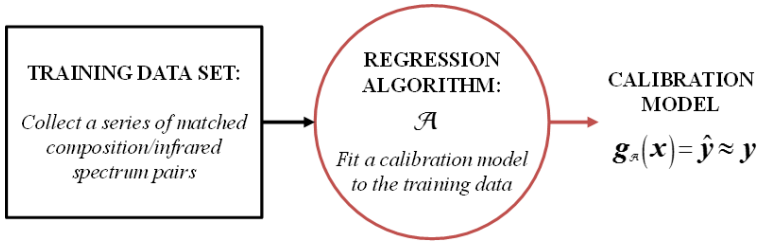

Figure 2: Schematic outline of the basic calibration steps.

2.4 Training Set

To build a training data set, a series of matched concentration/infrared spectrum pairs (Fig. 3) must be collected. It is convenient to represent the training set with matched X and Y matrices:

$$\mathbf{X}^{\text{tr}} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} = \begin{bmatrix} 1 & a_{1,1} & \dots & a_{1,m_{\text{select}}} \\ \vdots & & \ddots & \\ 1 & a_{1,1} & \dots & a_{1,m_{\text{select}}} \end{bmatrix} \in \mathbf{R}^{N \times q}, \tag{5}$$

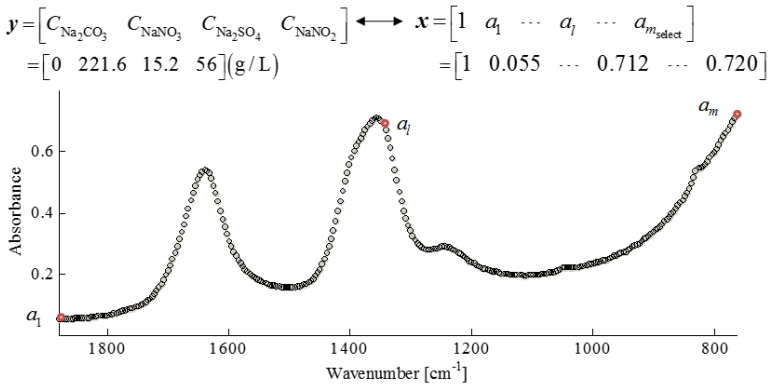

Figure 3: Example IR spectrum solution/composition (x/y) pair in the training set.

$$\mathbf{Y}^{\text{tr}} = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix} = \begin{bmatrix} C_{1,\text{CO}_3^{2-}} & C_{1,\ \text{NO}_3^-} & C_{1,\ \text{SO}_4^{2-}} & C_{1,\text{NO}_2} \\ \vdots & \vdots & \vdots & \vdots \\ C_{N,\text{CO}_3^{2-}} & C_{N,\ \text{NO}_3^-} & C_{N,\ \text{SO}_4^{2-}} & C_{N,\text{NO}_2} \end{bmatrix} \in \mathbf{R}^{N \times p}. \tag{6}$$

In order to obtain an accurate calibration model, $\mathbf{Y}^{\text{tr}}$ must be carefully constructed. In particular, the effects of each solute on the infrared spectrum should be decoupled and the training set should be unbiased. This can be accomplished with an orthogonal and balanced experimental design [9, 10]. In the present work, we built the training set using a 3^{4-1} fractional factorial design using the concentration levels shown in Table 2. These concentration levels are chosen to be representative of high salt content nuclear waste solutions [11-13], however it should be noted that the training set solutions contain only the target solutes. To extend the presented calibration methodology to monitor actual nuclear waste, the training solutions should be modified to include the additional components of nuclear waste and the concentration levels should be adjusted to reflect those observed for the different classes of nuclear waste to be monitored.

2.5 Multivariate Regression

For an unbiased training set the relationship between $\mathbf{X}^{\text{tr}}$ and $\mathbf{Y}^{\text{tr}}$ should be representative of the general relationship between the solution composition and the infrared absorbance spectrum. Therefore we would expect the solution to the set of equations

Table 1: Training set design.

Factor (solute)	Level 1 [g-solute/100g-H_2O])	Level 2 [g-solute/100g-H_2O]	Level 3 [g-solute/100g-H_2O]
Na_2CO_3	0	6.50	13.00
$NaNO_3$	0	27.25	55.50
Na_2SO_4	0	1.85	3.70
$NaNO_2$	0	3.50	7.00

$\mathbf{X}^{\text{tr}}\mathbf{B} = \mathbf{Y}^{\text{tr}}$ to give a linear calibration model. Of course, measurement error and small nonlinearities make this set of equations over-determined and impossible to solve exactly. Consequently, regression techniques must be used to find the best model for which $\mathbf{X}^{\text{tr}}\mathbf{B} \approx \mathbf{Y}^{\text{tr}}$. To accurately predict the concentration of multiple solution components from a convoluted spectrum, a high-dimensional input feature vector must be used; however the features used are often collinear and therefore specific multivariate regression techniques are required to identify a calibration model that accurately correlates these features vectors to the concentration. Most commonly, collinear variables are handled with projection-based regression methods. Of these, partial least-squares regression has been extensively applied to develop multicomponent Raman-to-concentration calibration models [14-18] and single component IR-to-concentration calibration models [19-22].

While projection-based methods successfully navigate collinear variables, these methods do not explicitly take into account errors in the variables. In exploring the use of ATR-FTIR to monitor the concentration of multiple solution components, we noted a number of disturbance variables (temperature, infrared probe fiber optic cable alignment, liquid/probe window contact, and background infrared absorbance) that affected the infrared absorbance and can cause significant variance in the concentration prediction accuracy. To build a robust calibration model that was insensitive to these disturbance variables, we developed a customized regression algorithm. This algorithm, which we term robust parameter support vector regression (RPSVR), blends ideas used in robust estimation [23] and support vector regression.

3 Results

3.1 Calibration Model using RPSVR

To estimate the accuracy of established calibration models, the model-predicted concentrations were compared against the gravimetrically measured concentration for a test data set. The test set used in this study consisted of 310 spectra recorded for 11 solutions not used in the training set. For a calibration model built with RPSVR, the concentration prediction accuracy is shown visually in Fig. 4. In addition the concentration prediction

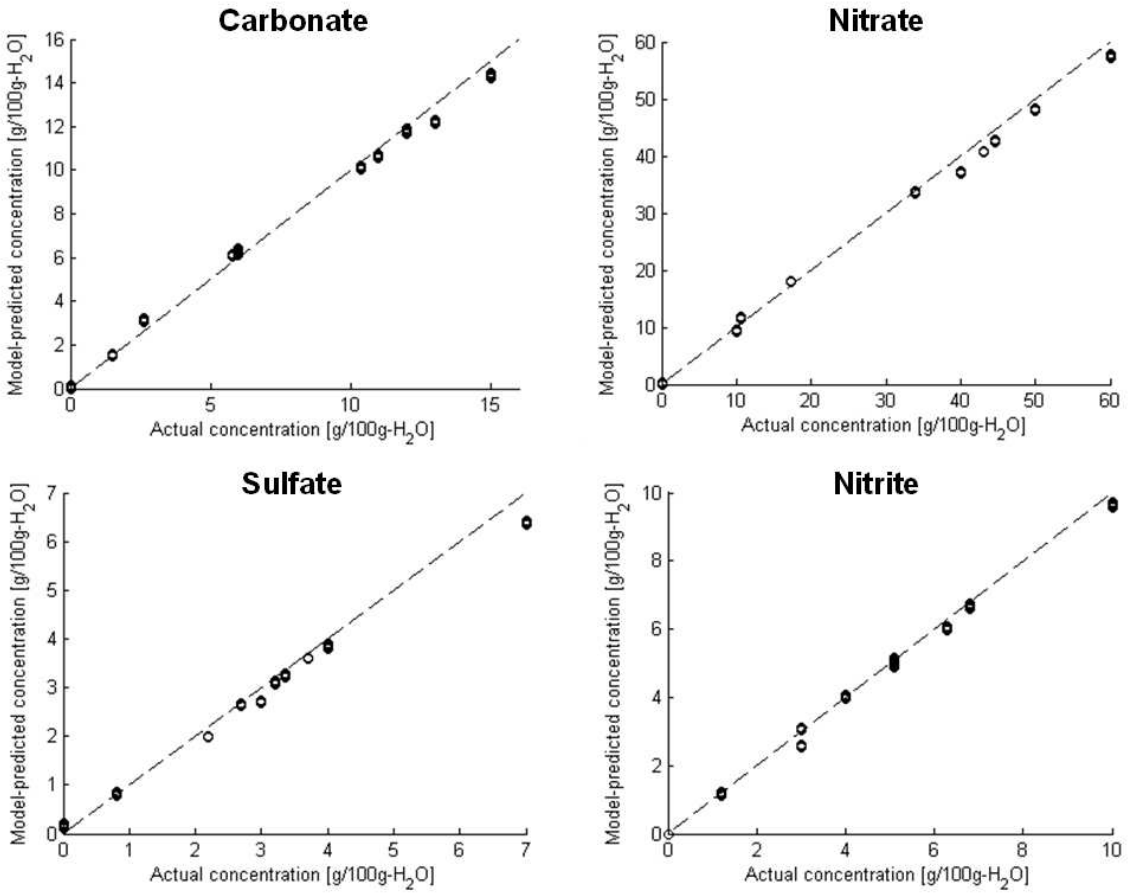

Figure 4: Model-predicted concentrations vs. actual concentrations across the test set for a calibration model constructed using RPSVR.

errors, $\epsilon = C_{\text{predicted}} - C_{\text{actual}}$ are characterized by the average (Eqn. 7), standard deviation (Eqn. 8) and relative error (Eqn. 10). The results are shown in Table 2.

$$\bar{\epsilon} \equiv \Big(\sum_{i=1}^{N_{\text{test}}} \epsilon_i \Big) / N_{\text{test}} \tag{7}$$

$$\sigma_\epsilon \equiv \frac{1}{N_{\text{test}} - 1} \left(\sum_{i=1}^{N_{\text{test}}} \left(\epsilon_i - \bar{\epsilon} \right)^2 \right)^2 \tag{8}$$

$$\bar{\epsilon}_{\text{rel}} \equiv \Big(\sum_{i=1}^{N_{\text{test}}} \left| \epsilon_i \right| \Big) / \Big(\sum_{i=1}^{N_{\text{test}}} C_{\text{actual},\, i} \Big) \times 100\% \tag{9}$$

Table 2: Concentration prediction errors observed over the test set using a calibration model built with RPSVR.

anion	Average Error, $\bar{\epsilon}$ [g-solute/100g-H_2O])	Error Standard Deviation, σ_ϵ [g-solute/100g-H_2O]	Relative Error, $\bar{\epsilon}_{\text{rel}}$ [$\%$]
CO_3^{2-}	-0.129	0.384	4.42
NO_3^-	-1.089	1.299	4.71
SO_42-	-0.112	0.211	6.33
NO_2^-	-0.139	0.163	3.95

These results demonstrate the ability to form an accurate multi-solute IR-to-concentration calibration model using the discussed training data set (large, orthogonal and balanced) and regression algorithm (RPSVR).

3.2 Comparison against calibration models build with more common regression techniques

In addition to examining the accuracy of a calibration model built with RPSVR, a number of additional calibration models were constructed using more established multivariate linear regression techniques, including multiple least-squares regression (MLR), principle

component regression (PCR) [24, 25], partial least-squares regression (PLSR) [26, 27], ridge regression (RR) [28], dead-zone linear regression (DZLR) [23], and support vector regression (SVR) [29].

Each regression algorithm was then compared based on the performance of the calibration model built using the measure of calibration model accuracy given by Equation (11) as the performance metric.

$$SEP(g_{\mathcal{A}})_{\text{combined}} \equiv \frac{1}{N_{\text{test}}} \left\{ \left(\sum_{i=1}^{N_{\text{test}}} \left(g_{\mathcal{A},j}(\mathbf{x}_i^{\text{test}} - y_{i,j}^{\text{test}} \right)^2 \right)^{(1/2)} / \bar{C}_j \right\} \tag{10}$$

where $SEP(g_A)_{\text{combined}}$ is the weighted standard error of prediction over the test set for the calibration model $g_{\mathcal{A},j}$ built using the learning algorithm $\mathcal{A}$, and $\bar{C}_j$ is the average concentration of the jth species in the training set.

The comparison strategy is shown pictorially in Fig. 5 and the results are shown in Fig. 6.

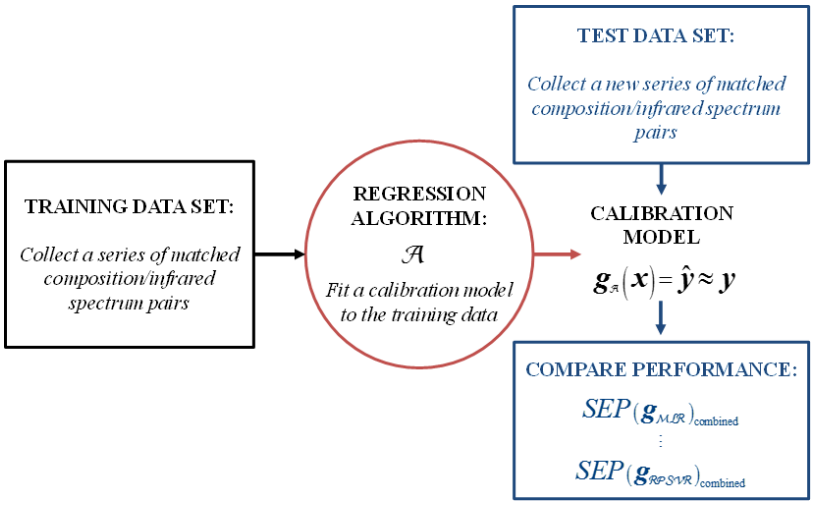

Figure 5: Schematic outline of the strategy used to compare learning algorithms.

Despite the complexity of the electrolytic solutions probed and the corresponding infrared spectrum, the results in Fig. 6 show that the calibration models built with any of the regression algorithms including the MLRthe simplest regression technique appliedpro-

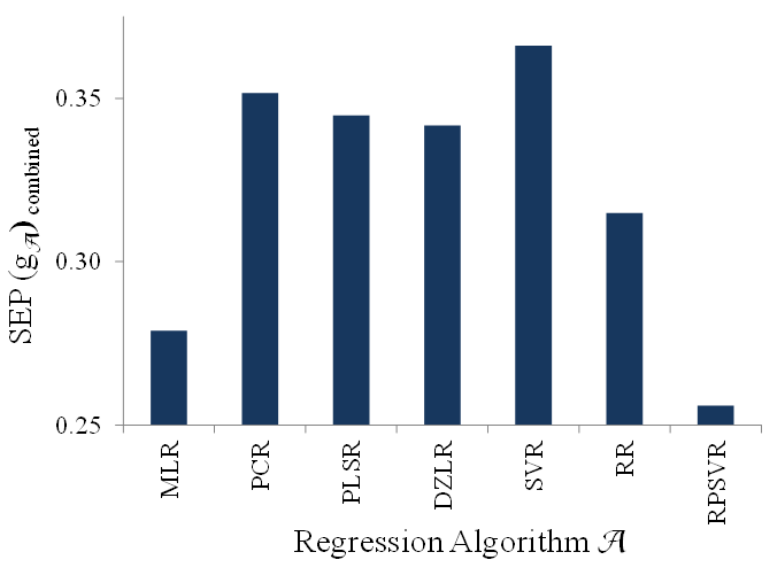

Figure 6: Comparison of the combined SEP for calibration models produced using different regression algorithms.

duced calibration models that performed reasonably well. We believe this is attributable to the fact the training set is large, orthogonal and balanced. In addition, the results shown in Fig. 6 indicate that the newly developed robust parameter support vector regression (RPSVR) performed the best of the examined algorithms.

Conclusions

In this work we have presented a tailored calibration methodology and demonstrated the ability to calibrate complex infrared absorbance spectra to give the in-solution concentration of four solutes simultaneously. In addition, we have shown that calibration models produced with a novel regression algorithm, termed robust parameter support vector regression, outperform models built with more standard regression techniques in the case examined.

References

1. Reynolds, J.G. and R. Carter, *Reconciliation of solute concentration data with water contents and densities of multi-component electrolyte solutions.* Journal of Solution Chemistry, 2008. 37(8): p. 1113-1125.

2. Miller, A.G., *Laser Raman Spectrometic Determination of Oxy Anions in Nuclear Waste Materials*. Analytical Chemistry, 1977. 49(13): p. 2044-2048.

3. Okemgbo, A.A., H.H. Hill, S.G. Metcalf, and M.A. Bachelor, *Determination of nitrate and nitrite in Hanford defense waste by reverse-polarity capillary zone electrophoresis*. Journal of Chromatography A, 1999. 844(1-2): p. 387-394.

4. Russell, R., L. Snow, and R. Peterson, *Methods to Avoid Post-Filtration Precipitation in Treatment of High-Level Waste*. Separation Science and Technology, 2010. 45(12-13): p. 1814-1821.

5. Reynolds, J.G., G.A. Cooke, D.L. Herting, and R.W. Warrant, *Salt Mineralogy of Hanford High-Level Nuclear Waste Staged for Treatment*. Industrial & Engineering Chemistry Research, 2013.

6. Bryan, S.A., T.G. Levitskaia, S.I. Sinkov, S.N. Schlahta, and J.M. Shaver, *Raman based process monitor for continuous real-time analysis of high level radioactive waste components*. Abstracts of Papers of the American Chemical Society, 2006. 232: p. 852-852.

7. Dunuwila, D.D., L.B. Carroll, and K.A. Berglund, *An Investigation of the Applicability of Attenuated Total-reflection Infrared Spectroscopy for Measurement of Solubility and Supersaturation of Aqueous Citric-acid Solutions*. Journal of Crystal Growth, 1994. 137(3-4): p. 561-568.

8. Dunuwila, D.D. and K.A. Berglund, *ATR-FTIR spectroscopy for in situ measurement of supersaturation*. Journal of Crystal Growth, 1997. 179(1-2).

9. Montgomery, D.C., Design and Analysis of Experiments. 6th ed. 2005: John Wiley & Sons, Inc.

10. Wu, J.C.F. and M.S. Hamada, *Experiments Planning, Analysis, and Optimization*. 2009, Wiley.

11. Herting, D.L., *Clean Salt Process—Final Report*, E. Management, Editor. 1996.

12. Geniesse, D.J., E.A. Nelson, J.H. Major, T.K. Nordahl, and D.W. Hamilton, *Fractional Crystallization of Hanford Single-Shell Tank Wastes—From Concept to Pilot Plant*, U.S.D.o. Energy, Editor. 2006.

13. Nassif, L., G. Dumont, H. Alysouri, and R.W. Rousseau, *Pretreatment of Hanford medium-curie wastes by fractional crystallization*. Environmental Science & Technology, 2008. 42(13): p. 4940-4945.

14. Goetz, M.J., G.L. Cote, R. Erckens, W. March, and M. Motamedi, *Quantitative application of in situ ATR-FTIR and Raman spectroscopy in crystallization processes*. Ieee Transactions on Biomedical Engineering, 1995. 42(7): p. 728-731.

15. Berger, A.J., T.W. Koo, I. Itzkan, G. Horowitz, and M.S. Feld, *Multicomponent blood analysis by near-infrared Raman spectroscopy*. Applied Optics, 1999. 38(13): p. 2916-2926.

16. Pelletier, M.J., *Quantitative analysis using Raman spectrometry*. Applied Spectroscopy, 2003. 57(1): p. 20A-42A.

17. Bryan, S.A., T.G. Levitskaia, A.M. Johnsen, C.R. Orton, and J.M. Peterson, *Spectroscopic monitoring of spent nuclear fuel reprocessing streams: an evaluation of spent fuel solutions via Raman, visible, and near-infrared spectroscopy*. Radiochimica Acta, 2011. 99(9): p. 563-571.

18. Levitskaia, T.G., et al., *Fourier Transform Infrared Spectroscopy and Multivariate Analysis for Online Monitoring of Dibutyl Phosphate Degradation Product in Tributyl Phosphate/nDodecane/Nitric Acid Solvent*. Industrial & Engineering Chemistry Research, 2013.

19. Lewiner, F., J.P. Klein, F. Puel, and G. Fevotte, *On-line ATR FTIR measurement of supersaturation during solution crystallization processes. Calibration and applications on three solute/solvent systems*. Chemical Engineering Science, 2001. 56(6): p. 2069-2084.

20. Togkalidou, T., M. Fujiwara, S. Patel, and R.D. Braatz, *Solute concentration prediction using chemometrics and ATR-FTIR spectroscopy*. Journal of Crystal Growth, 2001. 231(4): p. 534-543.

21. Togkalidou, T., H.H. Tung, Y.K. Sun, A. Andrews, and R.D. Braatz, *Solution concentration prediction for pharmaceutical crystallization processes using robust chemometrics and ATR FTIR spectroscopy*. Organic Process Research & Development, 2002. 6(3).

22. Cornel, J., C. Lindenberg, and M. Mazzotti, *Quantitative application of in situ ATR-FTIR and Raman spectroscopy in crystallization processes*. Industrial & Engineering Chemistry Research, 2008. 47(14): p. 4870-4882.

23. Boyd, S. and L. Vandenberghe, *Convex Optimization*. 2009: Cambridge University Press.

24. Naes, T. and H. Martens, *Principle Component Regression in NIR Analysis Viewpoint Background Details and Selection of Components*. Journal of Chemometrics, 1988. 2(2): p. 155-168.

25. Wold, S., K. Esbensen, and P. Geladi, *Principle Component Analysis*. Chemometrics and Intelligent Laboratory Systems, 1987. 2(1-3): p. 37-52.

26. Wold, S., A. Ruhe, H. Wold, and W.J. Dunn, *The Collinearity Problem in Linear-Regression: The Partial Least-squares (PLS) Approach to Generialized Inverses*. Siam Journal on Scientific and Statistical Computing, 1984. 5(3): p. 735-743.

27. Geladi, P. and B.R. Kowalski, *Partial Least-Squares Regression—A Tutorial*. Analytica Chimica Acta, 1986. 185: p. 1-17.

28. Kalivas, J.H., *Overview of two-norm (L2) and one-norm (L1) Tikhonov regularization variants for full wavelength or sparse spectral multivariate calibration models or maintenance*. Journal of Chemometrics, 2012. 26(6): p. 218-230.

29. Cortes, C. and V. Vapnik, Support-Vector Networks. Machine Learning, 1995. 20(3): p. 273-297.