Submitted Paper

Applied Spectroscopy

Applied Spectroscopy
2022, Vol. 76(2) 173–183

DOI: 10.1177/00037028211053852
journals.sagepub.com/home/asp
SAGE

Measuring Nd(III) Solution Concentration in the Presence of Interfering Er(III) and Cu(II) Ions: A Partial Least Squares Analysis of Ultraviolet–Visible Spectra

Poki Tse[1,2], Jenifer Shafer[1], Samuel A. Bryan[1], Gilbert L. Nelson[3], and Amanda M. Lines[1]

Abstract

Optical spectroscopy is a powerful characterization tool with applications ranging from fundamental studies to real-time process monitoring. However, it can be difficult to apply to complex samples that contain interfering analytes which are common in processing streams. Multivariate (chemometric) analysis has been examined for providing selectivity and accuracy to the analysis of optical spectra and expanding its potential applications. Here we will discuss chemometric modeling with an in-depth comparison to more simplistic analysis approaches and outline how chemometric modeling works while exploring the limits on modeling accuracy. Understanding the limitations of the chemometric model can provide better analytical assessment regarding the accuracy and precision of the analytical result. This will be explored in the context of UV–Vis absorbance of neodymium (Nd^{3+}) in the presence of interferents, erbium (Er^{3+}) and copper (Cu^{2+}) under conditions simulating the liquid–liquid extraction approach used to recycle plutonium (Pu) and uranium (U) in used nuclear fuel worldwide. The selected chemometric model, partial least squares regression, accurately quantifies Nd^{3+} with a low percentage error in the presence of interfering analytes and even under conditions that the training set does not describe.

Keywords

On-line monitoring, chemometric analysis, interference effect, ultraviolet–visible spectroscopy, UV–Vis spectroscopy, used nuclear fuel reprocessing, UNF reprocessing

Date received: 28 May 2021; accepted: 16 September 2021

Introduction

Optical spectroscopy is a widely applied technique that has been utilized to support research and process monitoring for decades.[1–9] A variety of optical techniques are available and can provide a level of chemical characterization that cannot be supplied via many other techniques. Examples include enabling identification and quantification of chemical speciation,[10] oxidation state,[11] and coordination environment.[12,13] In addition, spectroscopy generally has a simple experimental setup[14] and can be easily converted to process-friendly configurations to allow remote measurements.[15,16] However, the efficacy of all optical approaches is limited by signal interferences which can hinder accurate data analysis.[10,17–22]

Numerous data analysis approaches have been developed to overcome these challenges and expand the effective utilization of optical spectroscopy for complex sample analysis. Of greatest interest is chemometric analysis, which was first introduced in 1972, that applies mathematical and statistical methods to obtain the maximum relevant system information by analyzing multiple wavelengths simultaneously.[23,24] Utilizing multiple wavelengths establishes a more comprehensive relationship between the spectroscopic data and analyte concentration by extracting spectral changes in the bands of other species due to the changes of

[1]Pacific Northwest National Laboratory, Richland, WA 99352, USA
[2]Department of Chemistry, Colorado School of Mines, Golden, CO 80401, USA
[3]Department of Chemistry, The College of Idaho, Caldwell, ID 83605, USA

Corresponding authors:
Amanda M. Lines, Pacific Northwest National Laboratory, PO Box 999, MS P7-25, Richland, WA 99352, USA.
Email: amanda.lines@pnnl.gov
Gilbert L. Nelson, College of Idaho, 2112 Cleveland Blvd, Caldwell, ID 83605, USA.
Email: GNelson@collegeofidaho.edu

analyte or matrix concentration.[25] Several different methods for performing chemometric analysis have been advanced.[23,26–28] These methods may not all perform equally well given the conditions of an experiment or the type of chemical system or spectroscopy used and some methods are much more sensitive to outlier data. Partial least squares (PLS) analysis, which is used for this study, is a commonly used linear chemometric method that can accurately quantify analyte concentration without requiring pure component spectra when calculating the regression factor between analyte concentration and spectral changes.[23,27,28] However, the model may have a better correlation between analyte concentration and spectral changes by including the pure component spectra for complex systems.

While chemometric modeling has been used to complete the spectral analysis of many systems,[10,11,13–20] open questions remain regarding what their limits are in relation to accurate analysis of complex samples. Herein these limits are explored in the context of impacts that could be expected in off-normal processing conditions or unanticipated research system changes. Examples include ingrowth of optically interfering corrosion products, iron (Fe^{3+}), or analytical targets appearing at higher concentrations than accounted for in model training sets. Understanding the limitations of the chemometric model can better assess the accuracy and precision of the analytical result.

The chemical system for this demonstration is a mixture of neodymium (Nd^{3+}), erbium (Er^{3+}), and copper (Cu^{2+}) in nitric acid. This system is a simplified, but sufficient optical facsimile, to the liquid–liquid extraction aqueous phase that selectively extracts plutonium (Pu) and uranium (U) from used nuclear fuel (UNF) solutions.[29] A separation process of this nature would have transition metals, with large and broad optical signatures, arising from facility corrosion as well as significant quantities of lanthanides, with narrow and weakly absorbing optical signatures, from the fission that had occurred in the reactor. The [Nd^{3+}] in the actual used nuclear fuel is around 8.4 mM[30] with a molar absorptivity around 10 $M^{-1}\,cm^{-1}$ at 798 nm. The [Cu^{2+}] in the actual used nuclear fuel is around 0.42 mM[31] with molar absorptivity around 8.5 $M^{-1}\,cm^{-1}$ at 778 nm. The [Er^{3+}] is small compared to other lanthanides produced during fission but it has a stronger molar absorptivity at about 2.6 $M^{-1}\,cm^{-1}$ at 523 nm.

The process extraction efficiencies are dependent on solution conditions and these solution conditions are historically monitored using grab sample collection. The batch and off-line nature of grab sample collection makes immediate detection of changes in the solution conditions impossible. Implementing spectroscopic on-line monitoring can enable real-time or near real-time assessment of system chemistry, significantly improving system control while lessening safety concerns about the collection of grab samples. This is possible because most species present in these processes (e.g., lanthanides, nitric acid, extractant molecules, and actinides) have unique spectral signatures that can be utilized to quantify their concentration. Previous work has already demonstrated that ultraviolet–visible (UV–Vis) and Raman spectroscopy, in combination with chemometric analysis, can been applied to accurately quantify lanthanides and actinides in simulated and actual UNF samples.[4,30,32–35] However, analyte quantification under strong baseline effects, which could be due to the presence of solid particles, the presence of interferents, or the high radiation background (which can cause the darkening of optical windows and lead to a decrease in the amount of light transmitted to the detector), has not yet been fully examined.[18]

Of the lanthanides, neodymium was closely examined here because Nd^{3+} exhibits multiple bands in the visible region and is observed in UNF reprocessing solutions at significant concentrations as a fission product.[4] Neodymium has also been used as a non-radioactive surrogate for Pu^{3+} and Pu^{4+} ion, both spectroscopically and chemically.[36] These properties make Nd^{3+} an ideal candidate for chemometric analysis demonstrations. The high absorption band of Cu^{2+} at >560 nm as a perturbation of the baseline for the Nd^{3+} bands in that region and is also a common contaminant in fuel reprocessing systems.[31] These properties allow Cu^{2+} to test the model's ability to quantify analytes in the presence of strong perturbation of the baseline. Erbium is used as a non-radioactive surrogate for lanthanide metal fission products and was selected based on the band at $\sim$520 nm which overlaps with the Nd^{3+} spectrum and can be used to test the model capability of quantifying Nd^{3+} in the presence of interfering bands. By adding both Cu^{2+} and Er^{3+}, significant interferences are present that could impact quantification of Nd^{3+}.

This paper will be used to test the feasibility of utilizing PLS for accurate quantification of Nd^{3+} in the presence of overlapping bands (created by Er^{3+}) and a perturbation of the baseline (created by Cu^{2+}). The results of this paper set the groundwork for quantification of actinides in the presence of interferences in the actual UNF solution. This paper will also discuss model robustness with solution conditions not captured in the training set.

Experimental

Chemical

All metal, neodymium (Nd^{3+}), copper (Cu^{2+}), erbium (Er^{3+}), and iron (Fe^{3+}), hydrated nitrate salts ($Nd(NO_3)_3 \cdot 6H_2O$, $Cu(NO_3)_2 \cdot 3H_2O$, $Er(NO_3)_3 \cdot 5H_2O$, $Fe(NO_3)_3 \cdot 9H_2O$) were procured (Sigma-Aldrich) as reagent-grade salts and used as received. Concentrated nitric acid (HNO_3, ACS reagent grade, 70%) was procured (Sigma-Aldrich) and diluted to 1 M HNO_3. The 1 M HNO_3 solution was standardized by titration with NaOH. Stock metal solutions (0.33 M for Nd^{3+}, Er^{3+}, and Fe^{3+} and 0.5 M

for Cu^{2+}) were prepared by dissolving an appropriate amount of the metal salt in $\geq$18 MΩ · cm deionized water. The training and validation set sample solutions were later prepared by pipetting stock metal solutions into 1 M HNO_3.

Equipment

Spectra were collected using a UV–Vis spectrometer (CCD detector, Spectra Solutions Inc.) and associated Spectra Soft software (version 1.3). All UV–Vis spectra were collected in a 1 cm pathlength cuvette and referenced to deionized water. Acquisition times of 0.5 s were utilized in all measurements. Every five spectra were collected and averaged into one spectrum for modeling. All chemometric models were generated using the PLS toolbox (Eigenvector Research) in Matlab software (v.R2019a).

Chemometric Modeling

A PLS chemometric model was used to explore limits of applicability to complex spectral data. To explore these limits, a well-developed training set is required for the robust model development needed to accurately quantify Nd^{3+} in the presence of strong interferent effects. Besides the single component spectra (spectra that only contain either Nd^{3+}or Er^{3+} or Cu^{2+}), Nd^{3+} samples that contained either Cu^{2+} or Er^{3+} were also used to train the model to account for interferent effects. The training set was developed from the collection of 250 UV–Vis spectra, representing 25 samples where 10 replicate spectra of each sample were included in the matrix, see Table S1 (Supplemental Material) for more details. The validation set used to test the model was tested with solution conditions that the model had not seen before, i.e., both Cu^{2+} and Er^{3+} present. To make this testing more difficult for the model, samples with both of these metal interferents were not included in the training set. If both interferents were present in the training set it is expected that the model result would improve. It should be noted that all samples were prepared at the same [HNO_3] to maintain a constant [NO_3^-] for all samples in the training set. Any changes of nitric acid concentration will significantly affect the UV–Vis band of Nd^{3+} and affect the accurate quantification of Nd^{3+}. However, previous study has shown that a PLS model can accurately quantify the [Nd^{3+}] even when the [HNO_3] varies so long as the variation is represented in the training set.[36,37] This observation can also be applied to other metal ions. Sample spectra were referenced to a deionized water spectrum using the UV–Vis instrumental software before exporting to Matlab for study.

Preprocessing of UV–Vis spectra was done to minimize the instrumental noise and allow the model to focus on data that contained the chemical information.[11,37] Raw and post-processed training set spectra are presented in Fig. S1 (Supplemental Material). Spectral preprocessing included limiting the spectral range modeled to 468–860 nm as well as applying a first derivative followed by mean center (Fig. S1b). The first derivative can account for the perturbation of the baseline caused by the high absorbance band of Cu^{2+},[40] while mean centering can avoid the bias created by samples with higher analyte concentrations and allow the training set to better focus on the difference between samples.[27] Mean centering of the data is also applied in the concentration matrix to equalize the weighting of samples used in constructing the model.[27]

After preprocessing the training set samples, PLS models were built and applied to validation sets which include two data sets that the model was not trained to evaluate. Samples in validation set A contain Nd^{3+}, both interferences (Cu^{2+} and Er^{3+}), have different analyte concentrations than the training set but within the same concentration range, and were used to evaluate the model performance under the normal conditions when the concentration of the analytes is within the ranges covered by the training set. This set was is a collection of 70 UV–Vis spectra, seven representative samples in which 10 replicate spectra of each sample were included in the matrix. Validation Set B contains solution conditions that are not captured in the training set, including the presence of a new analyte, Fe^{3+}, and higher analyte concentrations than training sets samples. This set was developed from the collection of 14 samples where 10 replicate spectra of each sample were included for a total of 140 UV–Vis spectra. The analyte concentration in Validation Sets A and B is presented in Table I.

Results and Discussion

Single Variate Analysis

While the relationship between absorbance and [Nd^{3+}] in the pure component sample for most selected wavelengths is linear as expected (Fig. S2, Supplemental Material) consideration of how single variate analysis can be impacted by the presence of interfering analytes is relevant. Examples of analyte pure component spectra and multiple component spectra are shown in Fig. 1.

Within the system studied here, the high absorption band of Cu^{2+} generates a large perturbation of the baseline after 560 nm. The absorbance of Nd^{3+} bands at >700 nm has plateaued after the addition of 0.12 M Cu^{2+} (Fig. 2). Although the [Nd^{3+}] is held constant at 0.07 M, the absorbance of Nd^{3+} bands at >700 nm increases and reaches plateau with the addition of Cu^{2+} (Fig. 2a). Although the effect of Cu^{2+} on the absorbance of the Nd^{3+} band at 576 nm is less significant compared to the bands at >700 nm, Cu^{2+} has a large effect on the Nd^{3+} spectral signature across a range of wavelengths. A similar effect was observed when holding Nd^{3+} constant with variable

Table I. Analyte concentration in both validation set and model performance on validation samples.

Sample no.	Val set	Nd^{3+} (M)	Cu^{2+} (M)	Er^{3+} (M)	Fe^{3+} (M)	Nd^{3+} (M) (PLS)	%Error[b]	Pass/fail[c]
26	A	0	0	0	0	−0.002	N/A	Pass
27	A	0.03	0.12	0.04	0	0.031	5.0	Pass
28	A	0.05	0.17	0.06	0	0.053	6.0	Pass
29	A	0.07	0.22	0.08	0	0.067	3.7	Fail
30	A	0.03	0.22	0.09	0	0.03	1.0	Fail
31	A	0.07	0.17	0.07	0	0.07	3.2	Pass
32	A	0.11	0.12	0.05	0	0.11	4	Fail
33	B	0	0	0	0	−0.002	N/A	Pass
34	B	0.19	0	0	0	0.18	5.7	Fail
35	B	0.01	0	0	0	0.009	9.0	Pass
36	B	0.19	0.12	0.03	0	0.18	6.9	Fail
37_0[a]	B	0.07	0.12	0.03	0.1	0.075	7.0	Pass
37_1[a]	B	0.07	0.12	0.03	0.1	0.075	7.0	Pass
37_2[a]	B	0.07	0.12	0.03	0.1	0.075	7.1	Fail
37_3[a]	B	0.07	0.12	0.03	0.1	0.074	6.0	Fail
37_4[a]	B	0.07	0.12	0.03	0.1	0.061	12.4	Fail
38	B	0.03	0.44	0	0	0.025	17.0	Fail
39	B	0.07	0	0.1	0	0.071	1.6	Pass
40	B	0.01	0.03	0.1	0	0.0095	5.4	Pass
41	B	0.3	0	0	0	0.28	6.4	Fail
42	B	0.07	0	0.18	0	0.08	18.2	Fail

[a]Sample 37 was selected to test the model accuracy under the addition of some artificial baseline effect, the detailed discussion of which is included in Figure S5. _0 (without adding artificial baseline); _1 (adding constant), _2 (adding linear slope), _3 (adding curve) and _4 (adding curve).

[b]Calculated as $\frac{|\text{PLS measured } [Nd^{3+}] \text{ (column 7)} - \text{volumetrically measured } [Nd^{3+}] \text{ (column 3)}|}{\text{volumetrically measured } [Nd^{3+}] \text{ (column 3)}} \times 100\%$.

[c]Pass refers to sample Q and T^2 values that are both within 95% confidence limit of Q and T^2 value of training set.

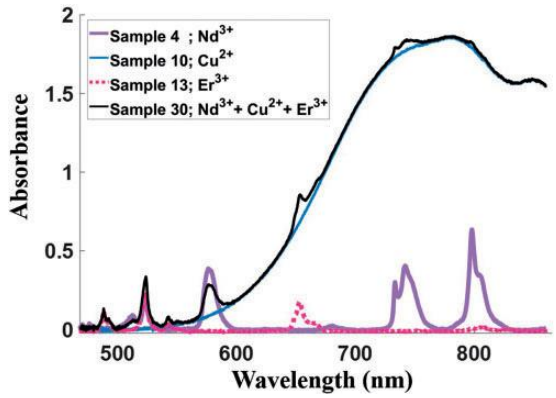

Figure 1. Representative UV–Vis spectra of pure Nd^{3+} (sample 4), Cu^{2+} (sample 10), Er^{3+} (sample 13) and sample consisting of Nd^{3+}, Cu^{2+}, and Er^{3+} (sample 30). The analyte sample concentrations can be found in Table I and Table S1.

Er^{3+} (Fig. 2b). This provides an example of how single variate analysis can be impacted by the presence of interfering components. Here, a Beer's law analysis relying on bands impacted by perturbation of the baseline will provide inaccurate quantification of Nd^{3+} (Fig. 2).

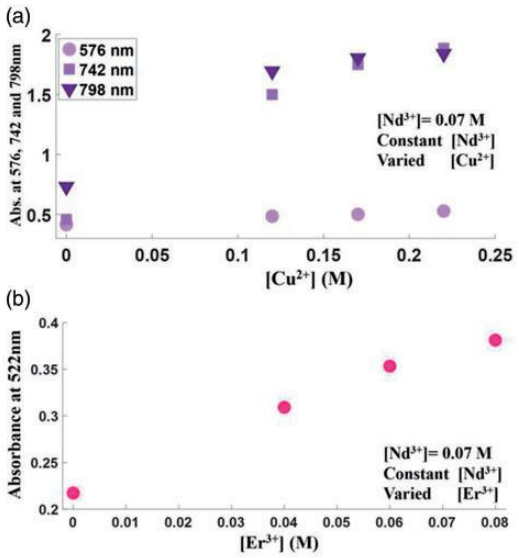

Figure 2. (a) The influence of Cu^{2+} on the absorbance of Nd^{3+} bands at 576 nm, 742 nm, and 798 nm. (b) The influence of Er^{3+} on the absorbance of Nd^{3+} bands at 522 nm. The absorbance of 0 M Cu^{2+} and Er^{3+} (at 0.07 M Nd^{3+}) was extrapolated from Figure S2.

To highlight an important note, many modeling approaches will attempt to produce optical training sets that capture all possible variants to avoid the negative impacts of unaccounted for interferents or conditions. However, cost and time limitations can impair collection of an expansive training set. Additional, truly unexpected and unknown interferents can contaminate the process and, ultimately, the analysis.[40] The goal here is to explore the limits of modeling accuracy particularly under these conditions.

Loadings

The data was treated using PLS analysis, which relates and captures the major variances in spectroscopic and concentration data into new abstract latent variable (LV) factors. Having a sufficient number of LVs is important for analyte quantification since PLS requires enough LVs to capture the spectral variance of interest in the system. An excess number of LVs can result in capturing noise in the training set which will affect the model measurement accuracy on the unknown (to the model) samples.[27] For the Nd^{3+} focused training set, the first LV captures the major variance in the sample while the second LV captures the next remaining major variance and so on. A loading matrix is composed of these abstract factors which constructs a new coordinate space to represent the range of variance of the samples. The corresponding score matrix (vide infra) then consists of the coordinates of those samples in the new coordinate space.

Loading plots show the spectroscopic variances captured in each individual LV and the loading plots for the four LVs are shown in Fig. 3a. The loading of each LV was normalized to have each LV plotted on a similar scale for easier comparison to the analyte UV–Vis spectra (Fig. 3b). Additional LVs are shown in Fig. S3. The first LV mainly captured Nd^{3+} spectral signatures which is expected. The next three LVs captured the remaining Nd^{3+} spectral signatures in addition to some Cu^{2+} and Er^{3+} spectral signatures. The loading plot of the fifth and sixth LVs still captures some Nd^{3+} spectral signatures, but they capture less than 1% of the total spectral variances. The loading plot of the seventh LV primarily captured spectral noise. Models were limited to four LVs as these captured the vast majority of variance (99.4%), accounted for the chemical species present, and avoided overfitting of the model.

Scores

The distribution of the training set samples in the new coordinate space can be presented by a score plot as a score matrix contains the sample coordinates in a new coordinate space. To identify what species is linked to a particular LV, the relationship between the analyte concentration and the score of the LV can be observed

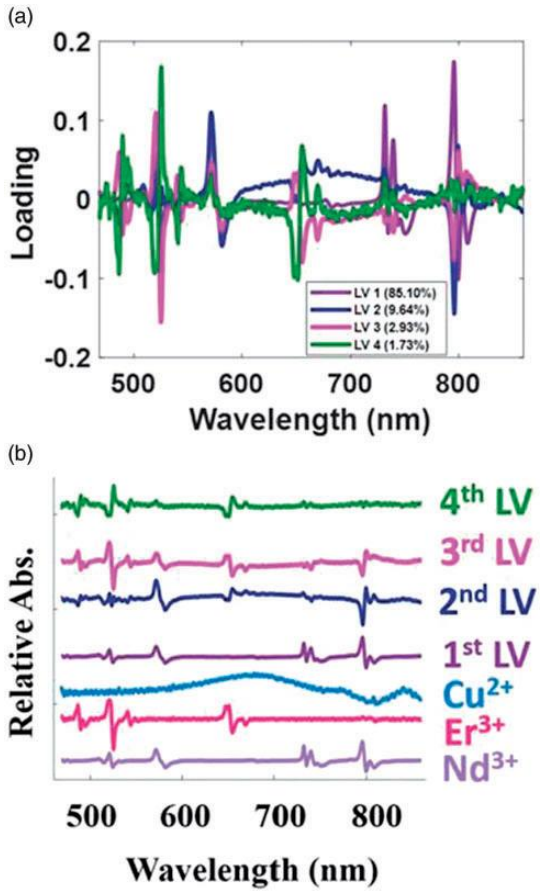

Figure 3. (a) Loading plot of the first four LVs. (b) Nd^{3+}, Cu^{2+}, and Er^{3+} spectra after first-derivative and normalization (for Nd^{3+}: normalized to 796 nm; for Cu^{2+}: normalized to 678 nm; for Er^{3+}: normalized to 521 nm). A normalized loading plot of the first LV (normalized to 796 nm, captures 85.1% of the spectral variance), second LV (normalized to 572 nm, captures 9.64% spectral variance), third LV (normalized to 521 nm, captures 2.93% spectral variance), and fourth LV (normalized to 525 nm, captures 1.73% spectral variance). This preprocessing method is only used for display purposes.

(Figs. 4 and S4). If a significant change in the score values of a particular LV occurs with a change in a particular analyte concentration, this indicates that the LV captures the spectral variance of the analyte. As shown in Fig. 4a, the first LV mainly captures the Nd^{3+} spectral changes because samples that contain various [Nd^{3+}] have different score values along the first LV. Samples that do not contain Nd^{3+} (pure Er^{3+} and Cu^{2+}) have similar score values (Figs. 4a and S4g).

The second LV captures the remaining Nd^{3+} spectral signature with some Cu^{2+} spectral signature because samples that contains various [Nd^{3+}] and [Cu^{2+}] have different score values on the second LV (Figs. 4b and S4a). This may further be seen in the T1 versus T2 plot (Fig. 4c) where a positive correlation is seen between the scores of LV 1 and scores of LV 2 for samples having a changing, increasing

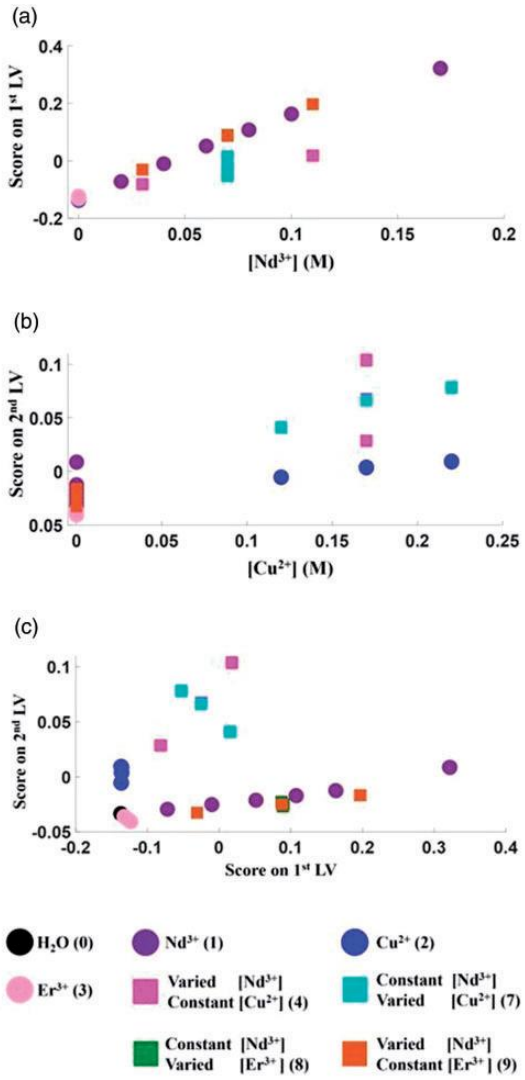

Figure 4. (a–b) Score plot of the first two LVs plotted against analyte concentration. (c) Score plot of LV1 versus LV2. The number in the legend represents the batch number of the samples and it can refer to Table S1. The circles indicate pure component samples while the squares indicate two component samples.

$[Nd^{3+}]$ (purple circles, green, and orange squares). Solutions in this figure containing both Cu^{2+} and Nd^{3+} show an even stronger correlation between these score values. The pink squares representing solutions with increasing $[Nd^{3+}]$ at constant $[Cu^{2+}]$ have a strong correlation, which is a much more positive change with the score of LV 2. The green squares representing solutions that have a constant $[Nd^{3+}]$ but a varying $[Cu^{2+}]$ show a strong negative correlation between these score values. Solutions involving only H_2O, Cu^{2+}, or Er^{3+} remain fairly localized. Samples containing only Er^{3+} exhibit similar score values on both the first and second LVs but have different score values on the third and fourth LVs (Figs. S4i and S4h). The third and fourth LVs also capture the remaining Nd^{3+} and Cu^{2+} behavior (Figs. S4b, S4c, S4e, and S4f). The fifth and sixth LVs capture the remaining Nd^{3+}, Cu^{2+}, and Er^{3+} spectral variance, but they capture less than 1% of the total spectral variances, which is of less use for Nd^{3+} quantification. Although the fifth and sixth latent variables capture only a small part of the total variation in the overall spectra, it is apparent from plots of the score values for each LV versus component concentration that these latent variables are picking up some subtleties of either solute–solute interactions or solute–solvent interactions (see Figs. S4j and S4k). LV 5 does show some variability with changes in $[Nd^{3+}]$ but also shows considerably more variability in the Nd^{3+}–Cu^{2+} solutions. The variability in the Nd^{3+}–Cu^{2+} solutions is much more than in the Cu^{2+} solutions alone indicating some accounting for solute–solute interactions. Latent variable 6 shows very little variability for changes in $[Nd^{3+}]$ or $[Er^{3+}]$ but indicates considerable variation in Cu^{2+} alone and even more for the interactions present in the Nd^{3+}– Cu^{2+} solutions. Although LVs 5 and 6 are not significant for the predictive model here they may be important to track changes in the UNF process. Interestingly, it appears the seventh LV captures the variance in the replicate measurement which is not a variance of interest in this system, so it was not included in this model (Fig. S4d). But for long-term process monitoring applications it would be very useful to include the seventh LV in modeling to keep track of the variations. In general, the scores plots provide further confirmation that the model should be limited to four LVs.

Model Performance Evaluation: Normal Condition

After choosing a suitable preprocessing method, and number of LVs to quantify Nd^{3+} in the presence of interferents (Cu^{2+} and Er^{3+}), the model performance was evaluated by comparing the model-measured $[Nd^{3+}]$ to the volumetrically prepared Nd^{3+} samples. The statistical approach utilized in this manuscript is the root mean square error of cross validation (RMSECV) and prediction (RMSEP) which can be calculated using Eq. 1:

$$\mathbf{RMSE(CV\ or\ P)} = \sqrt{\frac{\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}{m}} \quad (1)$$

where $\hat{y}_i$ is the model-measured analyte concentration in a given sample, y_i is the prepared analyte concentration of a given sample, and m is the total number of the samples.[41] Cross validation selectively excluded a set of samples from the training set and used the remaining samples in the training set to build a model to quantify analytes in the excluded samples as an estimate of the model performance of the unknown samples, the RMSECV.[41] The cross-validation method applied in this manuscript is venetian blinds with seven splits and each group of 10 replicate spectra of a sample removed as a contiguous block so all the replicate

spectra were excluded during cross validation. The RMSEP evaluates model performance on a validation set (values unknown to the model).[41] The PLS model accurately quantified [Nd^{3+}] even in the presence of strong interference effects (Fig. 5). The average measured error of [Nd^{3+}] on cross-validation data (RMSECV) (again the training set contains only Nd^{3+} as pure samples or mixed with either Cu^{2+} or Er^{3+}) is 0.003 M. This error is one to two orders of magnitude lower than the [Nd^{3+}] range in the training set.

The low RMSECV shows the model should be able to accurately quantify Nd^{3+} in an unknown data set even in the presence of both interferents. This hypothesis was tested by applying the PLS model to quantify Nd^{3+} in validation set samples (validation set A) which contains Nd^{3+} with both interferents. The RMSEP value is 0.003 M and the similar values between RMSECV and RMSEP indicate the model is robust enough to handle some degree of variation which is expected for an unknown sample.[39] The RMSEP showed the model is robust enough to accurately quantify

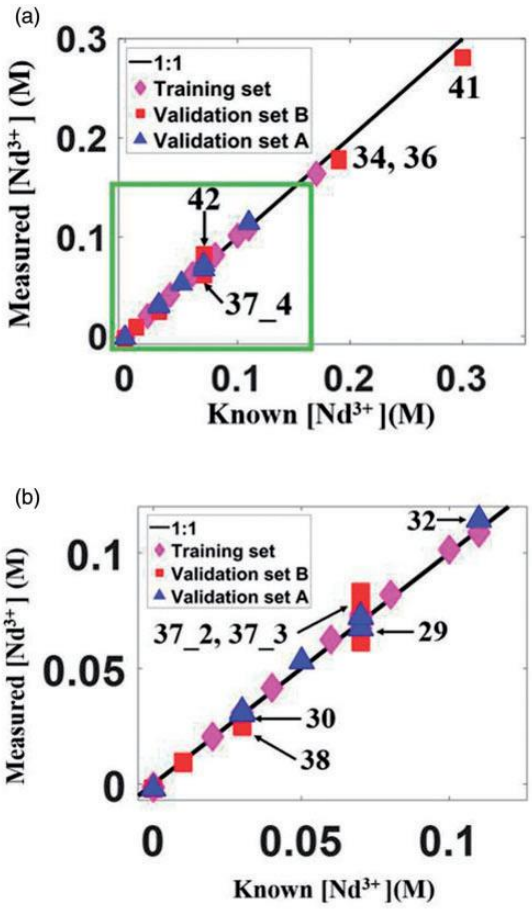

Figure 5. Measured (using PLS modeling) [Nd^{3+}] as a function of known concentration (volumetrically measured) of Nd^{3+} in the training set samples (purple diamonds), validation set A (blue triangles) and validation set B (red squares)). The individually labeled points are samples with relatively high Q and T^2 values (see next section). The composition of those samples is specified in Table I.

Nd^{3+} in the presence of both interferents even though no single training set sample included both interferents. With both interferents present, the percentage error of validation set A is below 7%. Figure S5 shows the RMSEC, RMSECV, and RMSEP out to seven latent variables. The values come close together at four LVs but the RMSEP diverges from the other statistics for LVs 5 and 6. If indeed LVs 5 and 6 account at least in part for solute–solute interactions this may indicate that these interactions are present in the validation set and are not fully accounted for in the training set. A well-constructed training set significantly improves the PLS model accuracy on quantifying Nd^{3+} in the unknown sample (validation A) which has a solution matrix that the model has not encountered before. The training set included both pure component spectra and multiple component spectra which helps the model to follow the Nd^{3+} spectral changes in the presence of the interferents. The RMSEP increased from 0.003 M to 0.02 M when the model excludes the multiple component spectra. Therefore, the model should be able to accurately quantify species even if it has fewer absorbance bands than Nd^{3+} when using a training set that includes samples that capture spectral changes of analytes in the presence of interferents or other physical effects (i.e., changes of temperature or pH). In summary, the PLS model is capable of accurately quantifying [Nd^{3+}] despite the presence of overlapping bands (Er^{3+}) and a perturbation of the baseline (Cu^{2+}) included in the validation set A.

Model Performance Evaluation Performance Evaluation: Unexpected Conditions

It is reasonable to anticipate that actual unknown samples will have unexpected conditions not covered in the training set. The common unexpected variances include the analyte concentration being higher or lower than the estimated concentration range covered in the training set or the appearance of new interferents not represented in the training set.

An additional, independent validation set (validation set B) was prepared in order to evaluate the model's ability to accurately quantify Nd^{3+} under off-normal or unexpected conditions. This new validation set includes samples with concentrations outside the analytes and interference concentration range covered in the training set. In addition, a new interferent not found in the training set, Fe^{3+}, was included as it is a common corrosion product in processing streams.[31] Although Fe^{3+} has no direct effect in the current wavelength region, it will have significant spectroscopic interference effect below 450 nm which can affect the quantification of actinides including U and may impact the solute–solute interactions modeled in the higher spectral region.[31] Since the influence of Fe^{3+} on the selected wavelengths was found to be insignificant, some artificial baseline effects (constant, liner slope, and curve) was

added to the collected spectra to provide a better evaluation of the ability of the model to tolerate unanticipated solution conditions (Fig. S5). The analyte concentrations of this validation set B is presented in Table I. The same static training set and the same resulting model employed to quantify the validation set A was used for this validation set B.

Two statistical methods are commonly used to evaluate model performance on unexpected conditions, which are Q statistics and Hotelling T^2 statistics (Eqs. 2 and 3), respectively.[42,43] The included references describe the equations of Q statistics and Hotelling T^2 statistics in principal component analysis (PCA) model, but it can be applied to other regression models.[44]

$$Q = x_i(I - PP^T)x_i^T \quad (2)$$

$$T^2 = \sum_{j=1}^{r} \frac{t_{sij}^2}{\lambda_j} \quad (3)$$

Here, x_i is a selected validation set sample, t_{sij}^2 is a square of the score element of the selected sample at a particular LV, P is the loading matrix, and λ_i is the total variance captured in the particular LV. Additional information on these approaches can be found elsewhere, but, most importantly, these are powerful indicators for normal or off-normal conditions.[42,43] Q statistics evaluate the difference between the actual sample spectra and the model constructed sample spectra. Therefore, Q statistics can be used to detect the unexpected spectral variances that are not covered in the training set, e.g., an unknown fingerprint. Hotelling T^2 statistics compares the variance in each sample compared to the total variance captured in the LV. Therefore, Hotelling T^2 statistics can be used to detect the unusual variance that is captured within the model, e.g., a known fingerprint but outside of calibration concentration range. The Q and Hotelling T^2 results of validation set A and B are shown in Fig. 6.

The model assessed four samples in validation set B with higher [Nd^{3+}] (samples 36 and 41) and [Er^{3+}] (sample 42) than the training set range ([Nd^{3+}]: 0–0.17 M and [Er^{3+}]: 0–0.09 M). Q and Hotelling T^2 analysis also successfully detected the untrained spectral signature accompanied with Fe^{3+} samples (especially sample 37_4) (Figs. 6a and S5). Those four samples were not considered the same as the population of the training set by both Q and Hotelling T^2 statistics at 95% confidence because those sample solution matrices were significantly different than the training set (Fig. 6; Table I). Most of the validation set A samples are within the 95% confidence interval of Q and T^2 which indicates validation set A and training set samples are similar (Fig. 6; Table I).

Samples with higher [Cu^{2+}] (sample 38, 0.44 M) than the training set covered (0–0.22 M) were not considered as the same population as the training set in Q statistics. Because the associated [Nd^{3+}] in these samples is low (0.03 M), the analysis has smaller T^2 values (Fig. S6). Interestingly, these two statistical approaches were not able to consider the Fe^{3+} containing samples (sample 37_0) as abnormal within 95% confidence. This might be expected because the included [Fe^{3+}] is low and does not have any significant absorption band in the Nd^{3+} wavelength region of focus (Fig. S6). However, after adding artificial baseline effects (37_2 and 37_3), they were not considered as the same population as the training set due to its high Q and T^2 values. This is likely because those baseline effects cannot be fully accounted for or minimized by the first derivative preprocessing method.

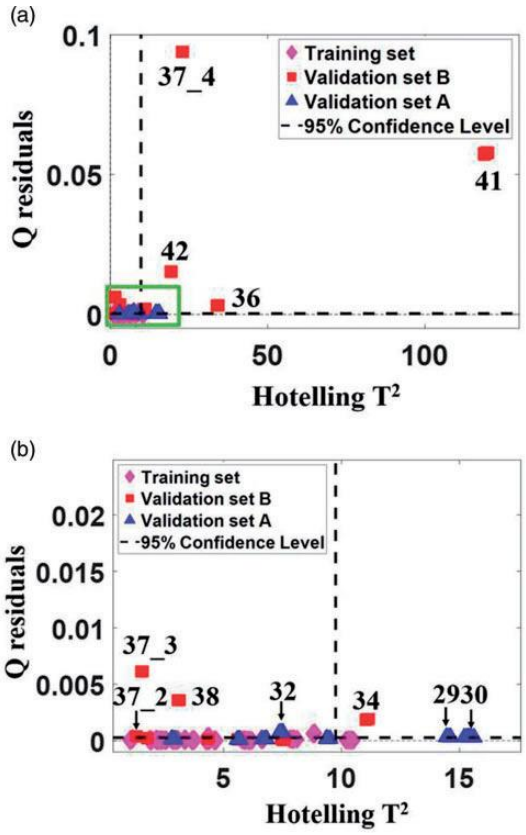

Figure 6. (a) Q and T^2 statistic result of the training set samples (purple diamonds), validation set A (blue triangles), and validation set B (red squares). (b) A closeup of Q and T^2 plot at the bottom left (green box in A). The named points are samples with relatively high Q and T^2 values. The composition of those samples is given in Table I.

Overall, Q and Hotelling T^2 statistics can be used to flag some unanticipated conditions in the new unknown samples which can be an indicator of the need to update the model to account the changes in the solution, but limitations exist. This suggests that the inclusion of additional analytical approaches is necessary to alert a system to off-normal conditions.[44] While some off-normal conditions

were not flagged by Q and T^2 statistics, the PLS model was still able to accurately quantify [Nd^{3+}] despite the presence of new interferents and abnormally high interferent concentration (Fig. 5). The RMSEP value is 0.009 M which is slightly higher than the RMSEP value of the normal validation set (validation set A, 0.003 M), but it is still one to two orders of magnitude lower than the concentration range in the training set. This shows the PLS model is sufficiently robust enough to handle some variation. The samples that were flagged by Q and Hotelling T^2 statistics (labelled as "Fail" in Table I) indicate that their spectral signatures were unaccounted for by the model. Although the model was still able to accurately quantify [Nd^{3+}] in those samples, the results should be approached with hesitation.

The model percent error of validation set B is higher than the percent error of samples in validation set A. This is expected as validation set B captured the unanticipated solution conditions that were not a part of the training set. The model will break down at a [Nd^{3+}] of 0.3 M given the loss of spectral response linearity (expected above absorbance values of one). Impressively the model could still function in this region, although this likely represents the end of the working range as above these concentration values, saturation of signal begins to have more notable spectral effects.

Similarly, the ability of the model to function in the face of high Cu^{2+} was impressive and likely only possible due to the broad shape of the band, allowing for straightforward application of a first derivative which decoupled the impact of the Cu^{2+} on the analysis. The introduction of species with sharper interfering bands has a different impact and is more easily discussed with the Er^{3+} interferent. The model would be anticipated to fail in the face of a broad band that also saturated signal, as a simple first derivative would not be able to account for saturated bands. In addition, model performance in the presence of increased interferents like Er^{3+} is not necessarily surprising, where additional literature suggests this challenge could be further exacerbated before causing model failure.[18,45] Finally, the addition of artificial baseline effects of Fe^{3+} containing samples creates different spectral features that are not covered in the training set and may cause the model to breakdown. Those baseline effects were minimized or accounted for after preprocessing by first derivative. Impressively, the model still accurately quantified Nd^{3+} even with the new interferent peak that was simulated in sample 37_4 around 650 nm. Overall, this work demonstrates a well-defined PLS model can exhibit surprising resilience to off-normal conditions, with failure points primarily appearing at the actual breakdown of spectral integrity, e.g., at saturation of response.

Conclusion

Spectroscopic on-line monitoring has been proposed to replace the periodic off-line analysis to characterize the chemical composition of UNF reprocessing in real-time. However, the traditional univariate analysis of these spectra is impaired by complicated solution matrices. Here, a multivariate analysis, PLS, is able to accurately quantify Nd^{3+} in the presence of the interferents Er^{3+} and Cu^{2+}, which, respectively, have absorption bands that interfere with Nd^{3+} and a perturbation of the baseline. Both types of spectral changes are common phenomena in processing streams and can prevent the accurate quantification of target analytes.

The ability of chemometric modeling, and, in this example specifically, PLS regression to handle analysis challenges such as interferents or perturbation of the baseline was tested. This was pushed further to explore accuracy of response in the face of unexpected conditions such as introducing species not accounted for in the original model or introducing species at much higher concentrations than included in modeling training sets. Two statistical approaches, Q and Hotelling T^2 statistics were able to detect most abnormal samples, but limitations exist and suggest additional statistics could be valuable in discovering anomalous conditions. Aside from helping build an understanding of model strengths and weaknesses, samples failing to pass metrics is highly valuable to this study. Indeed, several samples were specifically designed to push model boundaries and were expected to fail. Identifying the role of metrics in the process monitoring of complex streams is key to building trust in model outputs or providing operators avenues to understand potential off-normal conditions. For example, samples labeled as "Fail" in Table I indicate that those samples are not considered to be in the same population of training set samples; they are outside the 95% confidence interval. Among those failing samples in Table I, eight of the samples belonged to the validation set B which had solution conditions that are not included in the training set. This indicates that the model is sensitive enough to detect most of the abnormal samples and can be an indicator of the need to update the model to account for any changes in the UNF process solution. Having a more complete training set to describe the anticipated solution conditions (e.g., potential interferents and analytes) in the actual UNF processing stream can help to lower the number of samples that will be considered as an outlier from the training set. The PLS model was still able to accurately quantify [Nd^{3+}] despite the presence of new interferents and abnormally high analyte and interferent concentrations, although a discussion of limitations of model applicability provides context as to where accuracy could break down. Understanding the limitations of the chemometric model can provide a better analytical judgement on the accuracy of the analytical result. We suspect the limitations brought out in this study are largely a function of mismatches in the composition and complexity of our training set and that present in the validation set. To apply the lessons here to process monitoring UNF

requires examination of modeling statistics as the training set is refined to reflect UNF solution composition and conditions. This study shows that attention to modeling statistics can raise a warning flag if the UNF process diverges far outside the expected conditions and solution composition. In general, this demonstrates that the PLS model can accurately quantify analytes despite the presence of interference and the unexpected changes of the solution condition that will be anticipated in the actual processing stream.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This research was supported by the U.S. Department of Energy, Office of Nuclear Energy, through the Nuclear Technologies R&D Program and by U.S. Department of Energy, Office of Science, through a Small Business Innovation Research (SBIR) grant, and was performed at the Pacific Northwest National Laboratory (PNNL) a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy. PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC05-76RL01830. The Authors P. T. and J.C.S would like to acknowledge the National Science Foundation, grant #1925708, for the support provided related data interpretation and manuscript preparation.

ORCID iD

Samuel A. Bryan https://orcid.org/0000-0002-8826-0880

Supplemental Material

All supplemental material mentioned in the text is available in the online version of the journal.

References

1. C. Madic, D.E. Hobart, G.M. Begun. "Raman Spectrometric Studies of Actinide(V) and -(VI) Complexes in Aqueous Sodium Carbonate Solution and of Solid Sodium Actinide(V) Carbonate Compounds". Inorg. Chem. 1983. 22(10): 1494–1503. doi: 10.1021/ic00152a015.
2. C. Madic, G.M. Begun, D.E. Hobart, R.L. Hahn. "Raman Spectroscopy of Neptunyl and Plutonyl Ions in Aqueous Solution: Hydrolysis of Np(VI) and Pu(VI) and Disproportionation of Pu(V)". Inorg. Chem. 1984. 23(13): 1914–1921. doi: 10.1021/ic00181a025.
3. B. Guillaume, G.M. Begun, R.L. Hahn. "Raman Spectrometric Studies of 'Cation–Cation' Complexes of Pentavalent Actinides in Aqueous Perchlorate Solutions". Inorg. Chem. 1982. 21(3): 1159–1166. doi: 10.1021/ic00133a055.
4. S.A. Bryan, T.G. Levitskaia, A.M. Johnsen, C.R. Orton, J.M. Peterson. "Spectroscopic Monitoring of Spent Nuclear Fuel Reprocessing Streams: An Evaluation of Spent Fuel Solutions via Raman, Visible, and Near-Infrared Spectroscopy". Radiochim. Acta. 2011. 99(9): 563–571. doi: 10.1524/ract.2011.1865.
5. J. Bürck. "Spectrophotometric Determination of Uranium and Nitric Acid by Applying Partial Least-Squares Regression to Uranium(VI) Absorption Spectra". Anal. Chim. Acta. 1991. 254(1–2): 159–165. doi: 10.1016/0003-2670(91)90022-W.
6. B.J. Colston, G.R. Choppin. "Evaluating the Performance of a Stopped-Flow Near-Infrared Spectrophotometer for Studying Fast Kinetics of Actinide Reactions". J. Radioanal. Nucl. Chem. 2001. 250(1): 21–26. doi: 10.1023/A:1013251827451.
7. P. Tse, S.A. Bryan, N.P. Bessen, A.M. Lines, J.C. Shafer. "Review of On-Line and Near Real-Time Spectroscopic Monitoring of Processes Relevant to Nuclear Material Management". Anal. Chim. Acta. 2020. 1107: 1–13. doi: 10.1016/j.aca.2020.02.008.
8. F. Baumgärtner, D. Ertel. "The Modern Purex Process and Its Analytical Requirements". J. Radioanal. Chem. 1980. 58(1–2): 11–28. doi: 10.1007/BF02533770.
9. G. Meinrath. "Uranium(VI) Speciation by Spectroscopy". J. Radioanal. Nucl. Chem. 1997. 224(1–2): 119–126. doi: 10.1007/BF02034623.
10. A.M. Lines, S.R. Adami, S.I. Sinkov, G.J. Lumetta, S.A. Bryan. "Multivariate Analysis for Quantification of Plutonium(IV) in Nitric Acid Based on Absorption Spectra". Anal. Chem. 2017. 89(17): 9354–9359. doi: 10.1021/acs.analchem.7b02161.
11. B. Debus, D. Kirsanov, C. Ruckebusch, M. Agafonova-Moroz, et al. "Restoring Important Process Information from Complex Optical Spectra with MCR-ALS: Case Study of Actinide Reduction in Spent Nuclear Fuel Reprocessing". Chemom. Intell. Lab. Syst. 2015. 146: 241–249. doi: 10.1016/j.chemolab.2015.05.023.
12. G. Meinrath. "Chemometric Analysis: Uranium(VI) Hydrolysis by UV–Vis Spectroscopy". J. Alloys Compd. 1998. 275–277: 777–781. doi: 10.1016/S0925-8388(98)00439-3.
13. G. Meinrath. "Chemometric and Statistical Analysis of Uranium(VI) Hydrolysis at Elevated U(VI) Concentrations". Radiochim. Acta. 1997. 77(4): 221–234. doi: 10.1524/ract.1997.77.4.221.
14. A.M. Lines, G.L. Nelson, A.J. Casella, J.M. Bello, et al. "Multivariate Analysis To Quantify Species in the Presence of Direct Interferents: Micro-Raman Analysis of HNO_3 in Microfluidic Devices". Anal. Chem. 2018. 90(4): 2548–2554. doi: 10.1021/acs.analchem.7b03833.
15. A.J. Casella, L.R.H. Ahlers, E.L. Campbell, T.G. Levitskaia, et al. "Development of Online Spectroscopic PH Monitoring for Nuclear Fuel Reprocessing Plants: Weak Acid Schemes". Anal. Chem. 2015. 87(10): 5139–5147. doi: 10.1021/ac504578t.
16. A.J. Casella, A.M. Lines, G.L. Nelson, J.M. Bello, S.A. Bryan. "MicroRaman Measurements for Nuclear Fuel Reprocessing Applications". Procedia Chem. 2016. 21: 466–472. doi: 10.1016/j.proche.2016.10.065.
17. A.M. Lines, S.R. Adami, A.J. Casella, S.I. Sinkov, et al. "Electrochemistry and Spectroelectrochemistry of the Pu(III/IV) and (IV/VI) Couples in Nitric Acid Systems". Electroanalysis. 2017. 29(12): 2744–2751. doi: 10.1002/elan.201700465.
18. A.M. Lines, P. Tse, H.M. Felmy, J.M. Wilson, et al. "Real-Time Analysis of Highly Complex Processing Streams: Quantification of Analytes in Hanford Tank Sample". Ind. Eng. Chem. Res. 2019. 58(47): 21194–21200. doi: 10.1021/acs.iecr.9b03636.
19. A.M. Lines, G.B. Hall, S. Asmussen, J. Alfred, et al. "Sensor Fusion: Comprehensive Real-Time, On-Line Monitoring for Process Control via Visible, Near-Infrared, and Raman Spectroscopy". ACS Sensors. 2020. 5(8): 2467–2475. doi: 10.1021/acssensors.0c00659.
20. A.M. Lines, G.B. Hall, S. Sinkov, T.G. Levitskaia, et al. "Overcoming Oxidation State-Dependent Spectral Interferences: Online Monitoring of U(VI) Reduction to U(IV) via Raman and UV–Vis Spectroscopy". Ind. Eng. Chem. Res. 2020. 59(19): 8894–8901. doi: 10.1021/acs.iecr.9b06706.
21. A.M. Lines, Z. Wang, S.B. Clark, S.A. Bryan. "Electrochemistry and Spectroelectrochemistry of Luminescent Europium Complexes". Electroanalysis. 2016. 28(9): 2109–2117. doi: 10.1002/elan.201600034.
22. A.M. Lines, J.D. Warner, W.R. Heineman, S.B. Clark, S.A. Bryan. "Spectroelectrochemical Sensor for Spectroscopically Hard-to-

Detect Metals by In Situ Formation of a Luminescent Complex Using Ru(II) as a Model Compound". Electroanalysis. 2018. 30(11): 2644–2652. doi: 10.1002/elan.201800427.
23. M. Otto. "The Computer-Based Laboratoryrdquo;. Chemometrics: Statistics and Computer Application in Analytical Chemistry. Weinheim: Wiley-VCH, 2017. Chap. 1.1, Pp. 2–9. doi: 10.1002/9783527699377.
24. M. Otto. "Multiple Linear Regressionrdquo;. Chemometrics: Statistics and Computer Application in Analytical Chemistry. Weinheim: Wiley-VCH, 2017. Chap. 6.2, Pp. 231–239. doi: 10.1002/9783527699377.
25. D. Kirsanov, V. Babain, M. Agafonova-Moroz, A. Lumpov, A. Legin. "Approach to On-Line Monitoring of PUREX Process Using Chemometric Processing of the Optical Spectral Data". Radiochim. Acta. 2013. 101(3): 149–154. doi: 10.1524/ract.2013.2009.
26. K.R. Beebe, R.J. Pell, M.B. Seasholtz. Chemometrics: A Practical Guide. New York: Wiley, 1998. Pp. 56–337.
27. R. Kramer. Chemometric Techniques for Quantitative Analysis. New York: Marcel Dekker, 1998. Pp. 51–158.
28. S. de Jong, B.M. Wise, N.L. Ricker. "Canonical Partial Least Squares and Continuum Power Regression". J. Chemom. 2001. 15(2): 85–100. doi: 10.1002/1099-128x(200102)15:2<85::aid-cem601>3.0.co,2-9.
29. G.J. Lumetta, J.R. Alfred, S.A. Bryan, G.B. Hall, et al. "Simulant Testing of a Co-Decontamination (CoDCon) Flowsheet for a Product with a Controlled Uranium-to-Plutonium Ratio". Sep. Sci. Technol. 2019. 54(12): 1977–1984. doi: 10.1080/01496395.2019.1594899.
30. S.A. Bryan, T.G. Levitskaia, A.J. Casella, J.M. Peterson, et al. "Spectroscopic On-Line Monitoring for Process Control and Safeguarding of Radiochemical Streams in Nuclear Fuel Reprocessing Facilities". In: K.L. Nash, G.J. Lumetta, editors. Advanced Separation Techniques for Nuclear Fuel Reprocessing and Radioactive Waste Treatment. Oxford: Woodhead Publishing Limited, 2011. Chap. 4, Pp. 95–119. doi: 10.1533/9780857092274.1.95.
31. R.J. Lascola, R.R. Livingston, M.A. Sanders, J.E. McCarty, G.A. Cooper. On Line Spectrophotometric Measurement of Uranium and Nitrate in H Canyon. Aiken, SC: Westinghouse Savannah River Co., 2002. https://sti.srs.gov/fulltext/tr2002334/tr2002334.pdf [accessed Sep 29 2021].
32. S.A. Bryan, T.G. Levitskaia, A. Casella, J.M. Peterson. "Spectroscopic Online Monitoring for Process Control and Safeguarding of Radiochemical Fuel Reprocessing Streams". Paper presented at: WM Symposia 2013: International Collaboration and Continuous Improvement, Phoenix, Arizona: 24–28 February 2013, Paper no. 13553.
33. C.R. Orton, S.A. Bryan, J.M. Schwantes, T.G. Levitskaia, et al. "Advanced Process Monitoring Techniques for Safeguarding Reprocessing Facilities". In: Proceedings of the Symposium on International Safeguards: Preparing for Future Verification Challenges, Vienna, Austria: 1–5 November 2010.
34. O.Y. Rodionova, T.I. Tikhomirova, A.L. Pomerantsev. "Spectrophotometric Determination of Rare Earth Elements in Aqueous Nitric Acid Solutions for Process Control". Anal. Chim. Acta. 2015. 869: 59–67. doi: 10.1016/j.aca.2015.02.037.
35. O.Y. Rodionova, A.L. Pomerantsev. "Non-Linear Multivariate Curve Resolution Applied to the Spectrophotometric Determination of Cerium(III) in Aqueous Nitric Acid Solutions for Process Control". Anal. Methods. 2016. 8(2): 435–444. doi: 10.1039/c5ay02393a.
36. G.L. Nelson, A.M. Lines, J.M. Bello, S.A. Bryan. "Online Monitoring of Solutions Within Microfluidic Chips: Simultaneous Raman and UV–Vis Absorption Spectroscopies.". ACS Sensors. 2019. 4(9): 2288–2295. doi: 10.1021/acssensors.9b00736.
37. G.L. Nelson, H.E. Lackey, J.M. Bello, H.M. Felmy, et al. "Lines Enabling Microscale Processing: Combined Raman and Absorbance Spectroscopy for Microfluidic On-Line Monitoring". Anal. Chem. 2021. 93(3): 1643–1651. doi: 10.1021/acs.analchem.0c04225.
38. R.G. Brereton, J. Jansen, J. Lopes, F. Marini, et al. "Chemometrics in Analytical Chemistry—Part I: History, Experimental Design and Data Analysis Tools". Anal. Bioanal. Chem. 2017. 409(25): 5891–5899. doi: 10.1007/s00216-017-0517-1.
39. G.L. Nelson, A.M. Lines, J. Casella, J.M. Bello, S.A. Bryan. "Development and Testing of a Novel Micro-Raman Probe and Application of Calibration Method for the Quantitative Analysis of Microfluidic Nitric Acid Streams". Analyst. 2018. 143(5): 1188–1196. doi: 10.1039/c7an01761h.
40. L.R. Sadergaski, D.W. DePaoli, K.G. Myhre. "Monitoring the Caustic Dissolution of Aluminum Alloy in a Radiochemical Hot Cell Using Raman Spectroscopy". Appl. Spectrosc. 2020. 74(10): 1252–1262. doi: 10.1177/0003702820933616.
41. K. Nee, S.A. Bryan, T.G. Levitskaia, J.W.-J. Kuo, M. Nilsson. "Combinations of NIR, Raman Spectroscopy, and Physicochemical Measurements for Improved Monitoring of Solvent Extraction Processes Using Hierarchical Multivariate Analysis Models". Anal. Chim. Acta. 2018. 1006: 10–21. doi: 10.1016/j.aca.2017.12.019.
42. R.A. Potyrailo. "On-Line Measurement". In: K.H. Jurgen Buschow, R.W. Chan, P. Veyssiere, editors. Encyclopedia of Materials: Science and Technology. Amsterdam; London: Elsevier, 2001, pp.6401–6411doi: 10.1016/b0-08-043152-6/01133-5.
43. L.E. Mujica, J. Rodellar, A. Fernández, A. Güemes. "Q-Statistic and T2-Statistic PCA-Based Measures for Damage Assessment in Structures". Struct. Heal. Monit. 2011. 10(5): 539–553. doi: 10.1177/1475921710388972.
44. O.Y. Rodionova, A.L. Pomerantsev. "Detection of Outliers in Projection-Based Modeling". Anal. Chem. 2020. 92(3): 2656–2664. doi: 10.1021/acs.analchem.9b04611.
45. C.A. Schroll, A.M. Lines, W.R. Heineman, S.A. Bryan. "Absorption Spectroscopy for the Quantitative Prediction of Lanthanide Concentrations in the 3LiCl–2CsCl Eutectic at 723 K". Anal. Methods. 2016. 8(43): 7731–7738. doi: 10.1039/c6ay01520d.