Comparing sensor fusion and multimodal chemometric models for monitoring U(VI) in complex environments representative of irradiated nuclear fuel

Luke R. Sadergaski,[1] Hunter B. Andrews, [1*] Brandon A. Wilson[2]*

[1]Radioisotope Science and Technology Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA

[2]Nuclear Energy and Fuel Cycle Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA

*Corresponding authors: sadergaskilr@ornl.gov, andrewshb@ornl.gov

ABSTRACT: Optical sensors and chemometric models were leveraged for the quantification of uranium(VI) (0–100 $\mu g \cdot mL^{-1}$), europium (0–150 $\mu g \cdot mL^{-1}$), samarium (0–250 $\mu g \cdot mL^{-1}$), praseodymium (0–350 $\mu g \cdot mL^{-1}$), neodymium (0–1,000 $\mu g \cdot mL^{-1}$), and HNO_3 (2–4 M) with varying corrosion product (iron, nickel, and chromium) levels using laser fluorescence, Raman scattering, and ultraviolet–visible–near-infrared absorption spectra. In this article, an efficient approach to develop and evaluate tens of thousands of partial least squares regression (PLSR) models, built from fused optical spectra or multimodal acquisitions, is discussed. Each PLSR model was optimized with unique preprocessing combinations and features selected using genetic algorithm filters. The 7-factor D-optimal design training set contained just 55 samples to minimize the number of samples. The performance of PLSR models was evaluated using an automated latent variable selection script. PLS1 regression models tailored to each species outperformed a global PLS2 model. PLS1 models built using fused spectra data and a multimodal (i.e., analyzed separately) approach yielded similar information, resulting in a percent root mean square error of prediction values of 0.9%–5.7% for the seven factors. The optical techniques and the data processing strategies established in this study allow for the direct analysis of numerous species without measuring luminescence lifetimes or reliance on a standard addition approach, making it optimal for near real-time, in situ measurements. Nuclear reactor modeling helped bound training set conditions and identified elemental ratios of lanthanide fission products to characterize burnup of irradiated nuclear fuel. Leveraging fluorescence, spectrophotometry, experimental design, and chemometrics can enable the remote quantification and characterization of complex systems with numerous species, monitor system performance, help identify the source of materials, and enable rapid high-throughput experiments in a variety of industrial processes and fundamental studies.

KEYWORDS: actinide, lanthanide, sensor fusion, optical spectroscopy, machine learning, online monitoring

INTRODUCTION

Optical sensors and chemometrics have gained popularity for in-line process monitoring and can be deployed remotely using fiber-optic cables for rapid, in situ analysis in harsh radiological environments.[1–3] Although multivariate analysis is a powerful option for quantification, one drawback is the dependence on large datasets, which are representative of the process. This is reasonable in many nonradiological environments, where such techniques were originally implemented.[4] Developing efficient methods to train chemometric models with as few samples as possible is imperative for processing in harsh/restrictive environments and where radioactive materials can be in short supply.[5–7] Design of experiments can be used to reduce the training set size while maintaining or improving prediction performance but are often limited to only two or three variables.[8,9] The designed approach must be extended to include numerous factors and test the limits for describing real-world systems.

In complex nuclear fuel cycle separation schemes, numerous corrosion product (CP) and fission product (FP) species are present amid the primary analytes, often including uranium, plutonium, nitric acid (HNO_3), and nitrate (NO_3^-).[10,11] CPs typically include common transition metals (e.g., iron, nickel, chromium), and FPs include lanthanides (e.g., cerium, praseodymium, neodymium, samarium, europium, and gadolinium), metals (e.g., zirconium, ruthenium, cesium, strontium) and nonmetals (e.g., bromine, iodine).[12,13] Various forms of optical spectroscopy can be used for the quantification of process solutions, each with unique benefits and drawbacks. For example, the lanthanide elements praseodymium(III) and neodymium(III) have relatively high molar extinction coefficients (~10 $cm^{-1} \cdot M^{-1}$), whereas europium(III) and samarium(III) have much lower extinction coefficients but efficient visible photoluminescence properties.[14] Therefore, the greatest utility could be achieved when multiple techniques are deployed simultaneously.[15]

Spectrophotometry, or ultraviolet–visible–near-infrared (UV-Vis-NIR) absorption spectroscopy, and photoluminescence spectroscopy are highly sensitive to many transition metals, lanthanides, and actinides.[16,17] Combining steady-state laser-induced fluorescence spectroscopy (LIFS) and chemometrics can yield quantitative results using a charge-coupled device to increase the timeliness of data collection instead of single-channel photomultiplier tubes, which are often used for time-resolved data.[18] Recent work demonstrated that it is possible to quantify fluorescent species like uranium(VI) directly in HNO_3 process media using ensemble learning techniques.[19]

The system must be trained such that the chemometric models can account for complicating effects of CPs, including fluorescence quenching and self-absorption. One common supervised technique is called partial least squares regression (PLSR).[20,21] Individual PLSR models can be built for multimodal (MM) datasets that span different types of sensors or ones with values of various scales and distributions fused into a global features space.[15] Data fusion methods combine information from multiple sources, which can enhance the spectral analysis.[22] *Low-level* and *mid-level* data fusion methods refer to the concatenation of all the raw data or extracted features of the raw data, respectively.[23] *High-level* fusion occurs when results are fused from individual models. Data fusion strategies, as opposed to separate MM strategies, have been used to improve prediction performance.

This study demonstrates how optical techniques, including LIFS, Stokes Raman scattering, and UV-Vis-NIR spectrophotometry techniques; design of experiments; and chemometrics can be harnessed together for the accurate quantification of multiple chemical species. The developed

methodology is widely applicable in many fields.[11,15,22,23,25] This work highlights the following: (1) sensor fusion and MM data were compared to quantify uranium(VI) and multiple lanthanides [samarium(III), europium(III), praseodymium(III) and neodymium(III)] in the milligram-per-liter range with varying HNO_3 and CP levels; (2) the D-optimal sample selection method was extended to a seven analyte system; and (3) nuclear reactor modeling was applied to identify ratios of lanthanide FPs for use as burnup indicators. This work addresses the challenges of modeling numerous species and presents a novel approach for advanced monitoring applications.

METHODS

All chemicals were commercially obtained and used as received unless otherwise stated. Concentrated HNO_3 (70%) was purchased from Sigma-Aldrich. A certified 10,000 $\mu g \cdot mL^{-1}$ uranium (^{238}U, depleted) standard was purchased from SPEX CertiPrep. Certified 10,000 $\mu g \cdot mL^{-1}$ inductively coupled plasma–optical emission spectroscopy standard solutions in 5% HNO_3 for each analyte were purchased from Inorganic Ventures. Samples were prepared using deionized (DI) water with MilliporeSigma Milli-Q purity (18.2 MΩ·cm at 25°C).

Sample preparation and design

Calibration and validation samples contained uranium(VI) (0–100 $\mu g \cdot mL^{-1}$), europium (0–150 $\mu g \cdot mL^{-1}$), samarium (0–250 $\mu g \cdot mL^{-1}$), praseodymium (0–350 $\mu g \cdot mL^{-1}$), neodymium ($\mu g \cdot mL^{-1}$) and HNO_3 (2–4 M) with varying CP (iron, nickel, and chromium) levels and were chosen to cover the anticipated solution conditions. Lanthanide concentrations were based on multiple UO_2 burnup and decay simulations (Figure S1). The discrete corrosion concentration levels were 0, 2,000, and 4,000 ppm, which were divided as 25% Cr, 25% Ni, and 50% Fe. This relative ratio of CP concentrations approximates the expected abundance of each species in real samples, where the source is generally stainless steel.[12] Training set samples were selected by a D-optimal using Design-Expert (v.11.0.5.0) by Stat-Ease, Inc., which included 55 samples as 42 required model and 13 lack-of-fit (LOF) points (Table S1). The validation set comprised 16 samples with acid/corrosion levels selected by additional D-optimal designs (Tables S2/S3). Additional details on the designs are provided in the Supporting Information.

Samples were prepared in individual 2 mL plastic microcentrifuge tubes (VWR Scientific, 525-1160) with volumetric pipettes and gravimetrically measured using a Mettler Toledo model XS204 balance with an accuracy of ±0.0001 g. A fluorometer flow cuvette, purchased from Starna (83.2.2F-Q-10/Z15) with a 10 × 2 mm optical pathlength, was used for each measurement to ensure consistent optical quality. Sample solutions were injected using a Fluid Metering, Inc. FMI pump with 1/16 in. tubing at a rate of approximately 1 $mL \cdot min^{-1}$.

Optical spectroscopy

Spectra were collected using iHR320 (Horiba Scientific), QEPro, and NIRQuest (Ocean Insight) spectrometers. Light sources consisted of a continuous-wave LBX 405 nm laser (Oxxius) and a stabilized broadband (360–2600 nm) light source (Thorlabs). Optical measurements were performed in triplicate at a constant temperature of 22.5 °C ± 0.1 °C. The spectrophotometers were referenced to DI water. Raman/fluorescence spectra were processed using LabSpec 6 software (Horiba Scientific), and absorption spectra were processed using OceanView software (Ocean Insight). Full details of instrument configurations are provided in the Supporting Information.

A measurement fusion (low-level) strategy was evaluated in this work. LIFS/Raman spectra were scaled by dividing each point by the maximum Raman water band intensity. Absorption spectra were normalized by dividing by the maximum absorption value to scale the intensities to one. UV-Vis and NIR absorption spectra were stitched (i.e., appended) near 910 nm, where the data was referenced to zero prior to analysis (Figure S2).

Multivariate analysis and preprocessing

PLSR was used to correlate spectral features to analyte concentrations to develop predictive models. Cross validation (CV) was used to determine the optimal number of latent variables (LVs) to include in the model. The best number of LVs was chosen by the last LV to show a significant difference in the model-explained variance. PLSR models can be built with one or multiple **Y** variables (PLS1 and PLS2, respectively). PLS2 models tend to account for covariance or multicollinearity between species represented in the spectral dataset. However, PLS1 models can be tailored to the spectral features of each species, which may require unique combinations of preprocessing and feature selection strategies.[18] This study evaluated PLS1 and PLS2 models with LV selection performed two ways. First, each model was constructed using an automated LV selection script, which determined the optimal number of LVs based on prediction error.[20] The second approached used the traditional *rule of thumb* assumption that the number of LVs should be equal to the number of factors in the design space (i.e., 7 LVs).

Spectral data were preprocessed prior to modeling. The preprocessing steps included scatter correction (standard normal variate [SNV]), noise/baseline correction (Savitsky–Golay filter [SG]), and/or mean centering (MC).[18] Details for the various 576 preprocessing combinations evaluated are provided in the Supporting Information. Trimming spectral data reduces the dimensionality by removing features that impede PLSR model correlations. A genetic algorithm (GA) was used for feature selection based on previous work.[18,26] The selected regions included in the chemometric model varied based on each analyte. All regression models and data preprocessing were completed in Python 3 using modules from the Scikit Learn.[27]

Statistical comparison

Model performance was evaluated using calibration, CV, and validation metrics. The primary statistics used to evaluate prediction performance was the root mean square error of the calibration, CV, and prediction (RMSEC, RMSECV, and RMSEP, respectively). PLSR model prediction performance testing on samples not included in the training set is important because the RMSECV is only an estimate, especially when using a designed sample matrix.[5,8] RMSE values were calculated using Equation 1:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}, \#(1)$$

where $\hat{y}_i$ is the predicted concentration, y_i is the measured concentration, and n is the number of samples. *RMSEP%* was calculated by dividing the RMSEP by the median of each respective analyte concentration range. PLSR models were optimized by minimizing the RMSEP. Prediction performance is strong when *RMSEP%* < 5%, satisfactory when *RMSEP%* ≤ 5%–10%, and indicative when *RMSEP%* ≤ 10%–15%.[9]

Fission product ingrowth modeling in nuclear fuel

The models were applied to evaluate a persistent challenge: that of FP ingrowth in nuclear fuel simulated for different reactors and fuel enrichments. Ratios of FPs change depending on the level of ^{235}U and ^{239}Pu fission during the fuel operation. The ratios reveal information on the fuel burnup, operating length, and attractiveness of the spent fuel for nuclear nonproliferation. The simulation of the FP ingrowth in nuclear fuel was modeled for a natural, 5%, and 20% enriched UO_2 fuel sample irradiated in the High Flux Isotope Reactor (HFIR) at Oak Ridge National Laboratory (Figure S2). The HFIR reactor was chosen because fuel samples are currently being irradiated as part of an active research effort into intentional forensics concepts for nuclear fuels.[25] Details pertaining to Monte Carlo N-Particle and SCALE software simulations of FP ingrowth during reactor burnup are provided in the Supporting Information.[28]

RESULTS AND DISCUSSION

Optical spectra

Combining absorption, florescence, and Raman spectral features and simultaneous chemometric analysis can be highly useful for characterizing and monitoring various chemical species in multiple fields.[15,19,22,23,29] The absorption, fluorescence, and Raman spectra of the chemical species characterized in this study are shown in Figure 1 with labeled dominant fluorescence emissions and absorption peaks. Molecular species such as the free uranyl ion (UO_2^{2+}) and uranyl nitrate complexes ($UO_2(NO_3)_n^{(2-n)+}$, n = 1 or 2) can be quantified by LIFS.[19] The samarium(III) emission spectrum consists of four bands near 563, 596, 643, and 720 nm, corresponding to the $^4G_{5/2} \rightarrow {}^6H_{5/2}$, $^6H_{7/2}$, $^6H_{9/2}$, and $^6H_{11/2}$ transitions (Figure 1a).[18] The europium(III) spectrum consists of five peaks from the first excited $^5D_0$ state to the $^7F_{0-4}$ Stark levels.

Although the 4f orbitals of the lanthanides are shielded by the filled 5s and 5p shells, optical transitions with electric dipole character are sensitive to the coordination environment.[14] The absorption spectra of praseodymium(III) and neodymium(III) are shown in Figure 1b. Each species is described by several absorption bands, with maximum molar extinction coefficients near 10 $M^{-1} \cdot cm^{-1}$. Neodymium(III) absorption bands originate from the ground state $^4I_{9/2}$. The peak near 580 nm is one of the most intense neodymium(III) peaks and corresponds to the $^4I_{9/2} \rightarrow {}^4G_{5/2}$ hypersensitive electronic transition, which is sensitive to the local symmetry and coordination environment.[30] The broad absorption peaks near 600 and 400 nm correspond to the CPs. These peaks are mainly prescribed to chromium(III) ions in solution.[31] NIR absorption water bands with negative absorbance values were observed at 960 and 1,200 nm. These regions are sensitive to various ionic species (e.g., H^+ and NO_3^-) and temperature.[8,9]

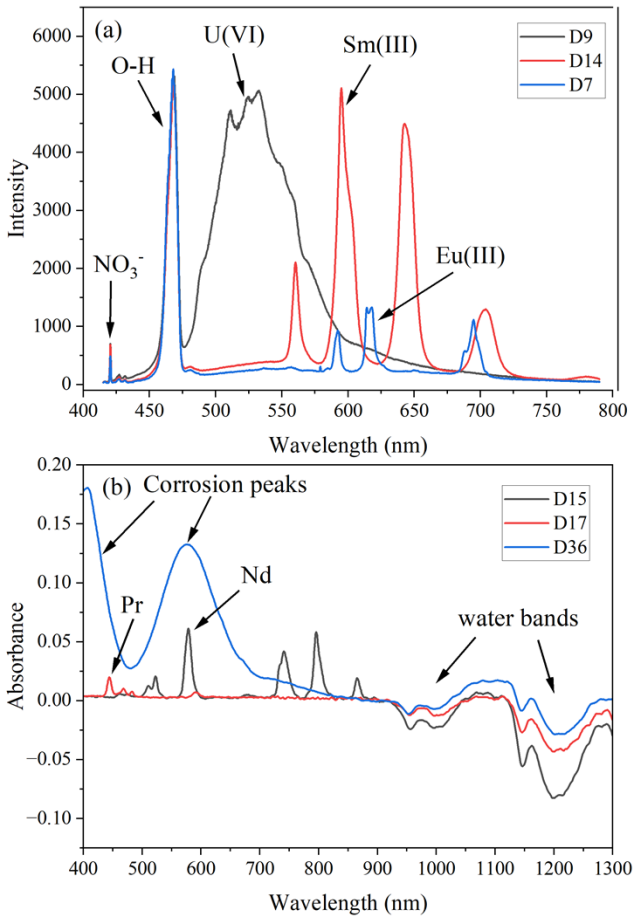

Figure 1. (a) Laser fluorescence spectra (λ_{ex} = 405 nm) of aqueous solutions and (b) absorption spectra aqueous solutions. The samples correspond to designed spectra (see Table S1).

In addition to emission peaks, the spectra contained Stokes Raman scattering features. The Raman NO_3^- symmetric N–O stretch corresponding to the strongest Raman active nitrate peak ν_1 (~1,048 cm^{-1}) was identified near 421 nm. The Raman O–H vibrational stretching region (broad peak 455–475 nm) is sensitive to the presence of cations, anions, ionic strength, and temperature. This band consists of several overlapping bands attributed to various H_2O and O–H (free and bound).[20]

The fluorescence signal intensity is dependent on many phenomena and physical parameters.[10,11] Many species found in reprocessing solutions and HNO_3 absorb light in the UV region, which is below the 405 nm excitation source used in this study.[10,12] *Self-absorption* refers to absorption of the excitation source and/or the emission signals from the sample itself. Thus, using a 405 nm excitation source minimizes self-absorption effects compared with lasers having wavelengths <405 nm.

There was evidence for self-absorption and/or uranium(VI) quenching with increased corrosion levels (Figure S3). The overall O–H band intensity and peak area decreased slightly with increasing levels of CPs because of absorption of the incident laser (405 nm) and absorption of the emission signal intensity. The uranyl emission intensity changed more significantly than that of

samarium(III) and europium(III). Uranium(VI)'s emission intensity was reduced by 15% in the sample highlighted in Figure S3. This reduction is likely a combination of self-absorption effects and quenching caused by iron(III) ions.[11] Lanthanide emission intensity appeared to be less dependent on the presence of CPs in the sample. The absorption bands and emission peaks overlapped extensively, rendering univariate analysis ineffective. Thus, the spectra were modeled with multivariate chemometrics to account for the overlapping and covarying spectral features.

Fused PLSR models with preprocessing combinations

Numerous preprocessing combinations and spectral features were evaluated for PLSR models built from fused spectra (Figure S2) and multimodal data using an automated Python script (see Supporting Information for preprocessing techniques).[18] This approach first tested all preprocessing combinations then applied a genetic algorithm (GA) for feature selection. PLS1 and PLS2 models were evaluated. An overview of the model development workflow is shown in Figure 2.

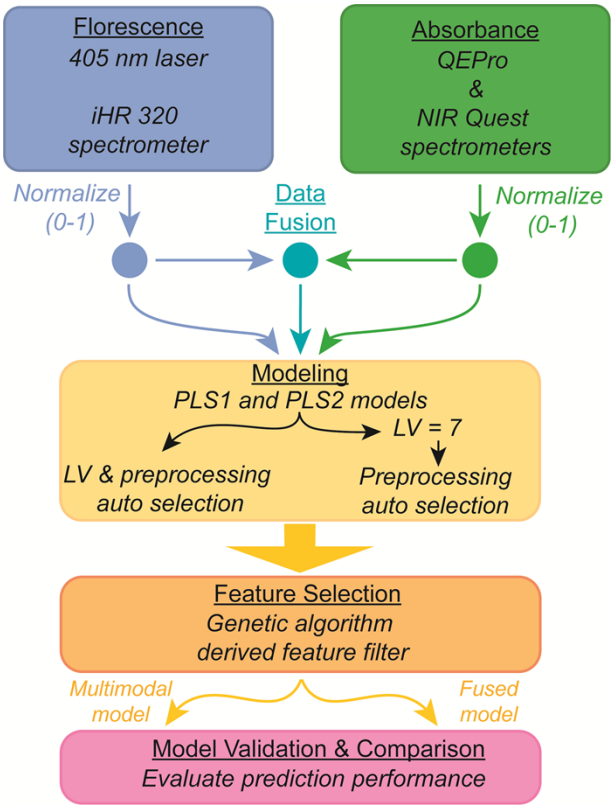

Figure 2. Schematic of data processing workflow for comparing fused and multimodal datasets.

The standardization between comparisons was the best prediction performance (i.e., the lowest achievable RMSEP). Other methods for standardization could include picking one preprocessing combination at random and forcing it upon each model. However, this option would effectively bootstrap the models because one preprocessing combination may be inherently more suitable for

one set of data over the other. Fused and nonfused spectra are unique; therefore, optimizing the regression of each set of spectra with different preprocessing methods was expected.

The PLS1 models leveraged unique preprocessing combinations for each species (Table 1). The PLS2 model selected one preprocessing combination representing the lowest average of all predicted RMSEP values (Table S4). PLSR models with the lowest *RMSEP%* values are shown in Table 1. PLS1 models were generated using the first 20 samples (20T) in the D-optimal design, which did not include any CPs (i.e., iron, nickel, and chromium) (Table S1). Using these PLSR models to predict samples in the validation set tested the model's predictive capability on samples outside of the trained conditions. The performance of this model was compared with that of the PLS1 models built using the entire D-optimal set with 55 samples (55T) to determine how important it is to include corrosion species in the model and whether models can perform well without including these species.[32] Finally, a PLS2 model was built using one preprocessing combination with the entire D-optimal set (55T) to compare with PLS1 models.

The average RMSEP%, excluding corrosion levels for the PLS1 models from the 20T and 55T sets, was 11.4% and 5.2%, respectively; the 55T set model reached a satisfactory level of prediction performance. Including corrosion levels, the PLS1 and PLS2 models for the 55T set had an average *RMSEP%* value of 4.7% and 8.1%, respectively. Applying unique preprocessing schemes and PLS1 models tailored for each analyte reduced the average *RMSEP%* by 42%. The top preprocessing methods for each PLS1 model were different because each spectrum consisted of unique peak widths and scattering features. Uranium(VI) was the only analyte to consistently use SNV for the 20T and 55T sets. The SNV method standardizes the spectra by subtracting the mean of each spectrum and dividing by the standard deviation, which likely minimized the effects of CP–related self-absorption and quenching effects. Models built without preprocessing did not perform as well (data not shown here).

Multimodal PLSR models with preprocessing

PLS1 models were built using the absorption or the laser fluorescence data separately to compare an MM approach with the fused approach discussed in the previous section. Unique data preprocessing strategies were selected for each variable using the same method applied to the fused dataset (Figure 2). These combinations were unique relative to the fused approach (Table 1) in the 20T set. For example, MC was applied to the fused samarium(III) spectral dataset but not to the MM dataset. Additionally, the SNV for uranium(VI) was applied to the fused spectra dataset but not to the MM 55T dataset.

Table 1. RMSEP and preprocessing combinations for fused and MM datasets.

		Pr	Nd	Sm	Eu	U	HNO_3	CPs
20T Fused	**LVs**	5	5	10	5	5	10	—
	RMSEP	15.3	11.4	13	3.57	7.84	0.243	—
	RMSEP%	8.7	2.3	10.4	7.1	15.7	24.3	—
	PP*	0,1,7,37,0	0,2,5,33,0	0,0,1,61,0	1,0,5,57,1	1,0,3,45,1	0,1,5,61,0	—
55T Fused	**LVs**	5	9	5	9	10	10	10
	RMSEP	8.01	8.41	3.23	3.06	4.08	0.082	0.018
	RMSEP%	4.6	1.7	2.6	6.1	8.2	8.2	1.8
	PP*	0,2,5,29,0	0,1,5,37,0	0,1,3,61,0	0,0,3,53,1	1,0,3,49,1	0,0,7,9,0	0,1,3,53,0

20T MM	**LVs**	5	7	10	5	5	8	—
	RMSEP	8.9	8.4	7.5	4	5.8	0.085	—
	RMSEP%	5.1	1.7	6	8.1	11.7	8.5	—
	PP*	0,2,3,13,0	0,1,5,29,0	0,0,1,21,1	1,0,1,61,1	1,0,3,13,1	0,0,1,5,0	—
	Data type	Abs.	Abs.	LIFS	LIFS	LIFS	Abs.	
55T MM	**LVs**	5	5	8	7	8	8	9
	RMSEP	7.2	8.6	2.9	3.5	3.1	0.072	0.012
	RMSEP%	4.1	1.7	2.3	7.1	6.1	7.2	1.2
	PP*	0,1,7,45,0	0,1,3,29,0	0,0,1,61,0	0,1,3,61,0	0,0,3,61,1	0,0,1,41,1	0,1,5,61,0
	Data type	Abs.	Abs.	LIFS	LIFS	LIFS	Abs.	Abs.

*Values in the PP row correspond to preprocessing: SNV (0=no, 1=yes); SG: derivative order (*X*), polynomial order (*Y*), and smoothing points (*Z*); and MC (0 = no, 1 = yes). Abs. stands for absorbance. **The PLS2 fused model was built using 7 LVs and the following preprocessing combination (1,2,7,41,0).

Praseodymium(III), neodymium(III), HNO_3, and corrosion levels were best measured using absorption spectra, and uranium(VI), samarium(III) and europium(III) were best modeled using the LIFS spectra. For example, the RMSEP% value using the 55T set for uranium(VI) based on absorption data was 34%. The LIFS model resulted in a RMSEP% value of 6.1%, which was very near the target of 5%. *RMSEP%* values for HNO_3 based on absorption NIR water bands were slightly lower than HNO_3 predictions based solely on the Raman water band. This difference could relate to the NIR band intensity not being altered by self-absorption effects. Information on the PLS regression coefficients for the top models for each analyte is provided in Figure S5.

PLSR models with feature selection

Spectral data can be trimmed to include only the regions that contribute to the regression model. This process can include manually trimming the spectra to include obvious peaks corresponding to relevant species and omitting regions with low signal intensity. This trimming can improve model performance by allowing the model to focus on regions of high correlations to the response matrix and ignore low-correlation regions. Feature selection is best applied to spectral data after preprocessing because features are altered by transformations or even lost (e.g., SG edge effects).

A GA was applied to the spectra after preprocessing to select the most important features in the spectra to improve the regression analysis. The *RMSEP%* values for most GA-filtered PLS1 models built from fused and MM datasets were substantially lower (Table 2). For the fused dataset, the *RMSEP%* was lowered by an average of 49.2% for the 20T PLS1 models and 26.9% for the 55T PLS1 models. The average fused dataset *RMSEP%* values were reduced to 5.1% and 3.9%, respectively. For the MM dataset, the *RMSEP%* values lowered on average by 13.1% (20T) and 19.8% (55T). This smaller reduction compared with the fused data is likely because irrelevant regions were already excluded from the spectra by analyzing one data source at a time.

Table 2. PLS1 model results with *RMSEP%* and percent change compared with preprocessing-only models after applying the GA filters*

20T	**Pr**	**Nd**	**Sm**	**Eu**	**U**	**HNO_3**	**CPs**
***RMSEP%* fused**	**1.9**	1.5	2.3	7.1	9.9	8.3	—
% change	−78.8	−33.9	−78.3	−1.23	−36.9	−66.1	—

***RMSEP%* MM**	4.3	1.6	3.3	7.5	11.6	8.2	—
% change	−14.7	−8.06	−45.5	−6.91	−0.07	−0.0342	—
55T	**Pr**	**Nd**	**Sm**	**Eu**	**U**	**HNO_3**	**CPs**
***RMSEP%* fused**	2.3	**1.3**	1.8	**4.8**	7.9	5.6	0.92
% change	−49.3	−23.4	−32.1	−21.4	−3.80	−31.5	−48.1
***RMSEP%* MM**	3.7	1.4	**1.6**	5.1	**5.7**	**5.5**	**0.86**
% change	−9.29	−20.1	−31.7	−27.4	−6.93	−23.2	−29.0

*Models were built using the top fused and MM models. The models with the lowest *RMSEP%* values are bolded/underlined.

Overall, the GA provided a spectral filter that improved the regression analysis and lowered *RMSEP%* values to either satisfactory or strong prediction levels for each species. For all species other than the praseodymium(III) in the fused dataset, the PLS1 models built from the training set of 55 samples resulted in lowered *RMSEP%* values. Figure 3 shows parity plots for predicted analyte concentrations compared with reference values for the top PLS1 models generated from the 55T set. In general, the predicted values for each species were close to the 1:1 line, which indicated highly accurate measurements. The uranium(VI) predictions had the largest spread around the 1:1 line; it had the largest *RMSEP%* value. These differences in uranium(VI) predictions were not correlated to corrosion levels. As an additional test, the slope and intercept of the parity plots were compared with their ideal values of 1 and 0 using an elliptical joint confidence region (EJCR).[33] All EJCR at 95% confidence contained the theoretical coordinates, ensuring the models were statistically accurate (more information in the Supporting Information).

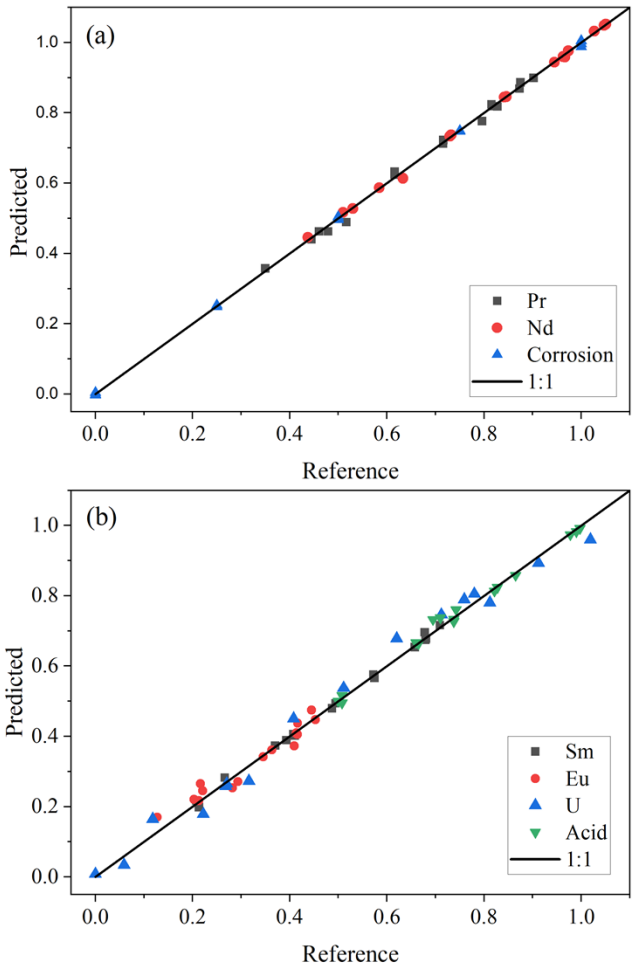

Figure 3. Parity plots for the top models' predictions of each analyte normalized to the highest concentration in the studied range compared with the 1:1 line.

It is up to the analyst to decide how many LVs are included in PLSR models, often limiting the number of models considered due to time constraints, which may inhibit prediction performance. This work used an automated LV selection script to test tens of thousands of PLSR models. However, in some cases, it chose LVs that appeared too unnecessarily high or low and LVs that were somewhat inconsistent between species. For example, the 20T PLSR model for samarium(III) selected ten LVs and the 55T set for samarium(III) choose five LVs. This could be a result of oscillations in RMSE that cannot be accounted by simple mathematics. Picking the optimal number of LVs is up to the user, and if too few are included, some of the explained variance in the system may not be accounted for; if too many are included, then the model can overfit noise. There are differing opinions on LV selection. Therefore, PLSR models were also built assuming a rule of thumb 7 LVs—one for each species. Prediction performance was similar to the automated LV selection script (details can be found in the Supporting Information).

The similarities between low-level fused and MM PLS1 model prediction performance indicated that a mid-level (e.g., feature fusion) method would not significantly improve the performance of the fusion approach.[23] The GA filters resulted in a mid-level fusion strategy option for the fused spectra because it chose features from multiple data sources for each analyte. However, including

features from each data source simultaneously did not significantly outperform PLS1 models built using independent sources, likely because most spectral features for the analytes were primarily observed with one technique or the other. Although Raman O–H water band and NIR features both contained information relevant to measuring acid concentration, the NIR features alone performed better than the fused model (see filter in Figure S6). Future work could compare the models described in this work, for the analysis of analytes measured by LIFS, to a high-level fusion approach, which fuses the results from individual models and additional regression techniques (e.g., ridge regression, random forest).

Predicting lanthanide ratios

The developed approach could be applied in the remote analysis of chemical process streams in real applications for the nuclear field. To articulate this application, the validation set was selected using concentrations derived from nuclear reactors modeling (Figure S1). Nuclear criticality safety, waste storage, treatment, fuel qualification, and nuclear code validation depend on precise elemental and isotopic measurements of irradiated nuclear fuel. The composition of the fuel post irradiation depends on the fuel properties, positioning within a reactor, power levels, irradiation duration, and many other conditions. Multicollector inductively coupled plasma–mass spectrometry or thermal ionization mass spectrometry are often used to monitor burnup (e.g., ^{148}Nd) compared with residual uranium and plutonium concentrations.[24] These methods typically require high-pressure ion chromatographic separations to measure the concentration and isotopic composition of neodymium, plutonium, and uranium.

The optical techniques used in this work are not sensitive to isotopic information and could provide elemental information useful for deconvoluting complicated inductively coupled plasma–mass spectrometry data owing to isobaric interferences without requiring separations. Additionally, the ratios between the lanthanide FPs are sample volume independent and could be used as a signature, in conjunction with some process knowledge, to help determine fuel type, reactor type, operating cycle length, plutonium attractiveness in the spent fuel, and burnup characteristics. Two example lanthanide ratios are shown in Figure 4. Additional examples are shown in Figure S4.

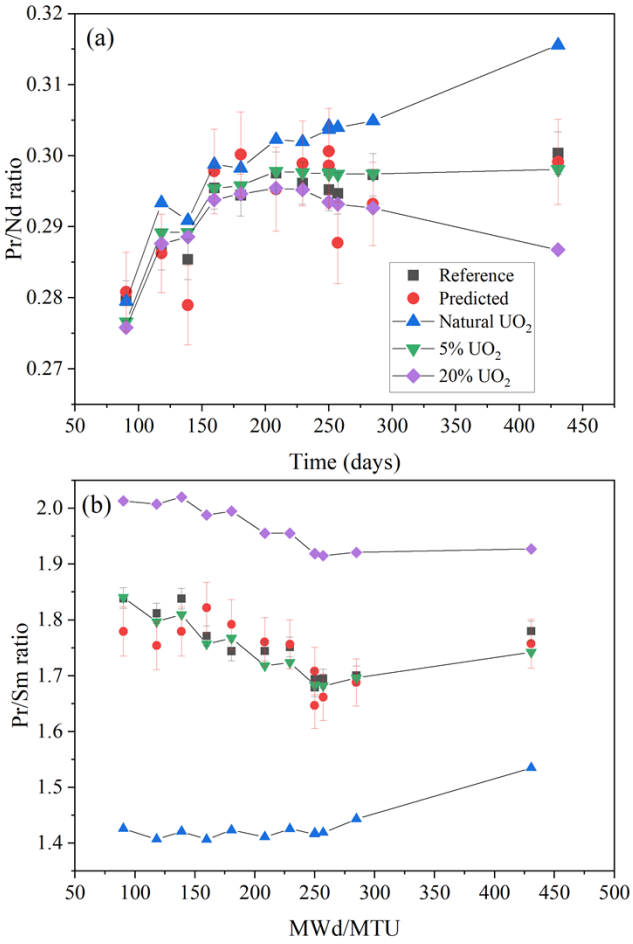

Figure 4. Reference, predicted, and modeled ratios for (a) praseodymium/neodymium and (b) praseodymium/samarium.

The error bars in the predicted values are approximated as the ±*RMSEP* values for each analyte, and the error in reference values were derived from standard error propagation calculation (see Supporting Information). The praseodymium/neodymium ratio is relatively constant until many days after irradiation. However, the praseodymium/samarium ratios are unique, and this technique has sufficient accuracy to measure the difference between natural, 5%, and 20% enriched UO_2 fuel types. The praseodymium/samarium ratio is unique for different fuel types because uranium enrichment affects the ratio of ^{235}U and ^{239}Pu in the fuel, and these isotopes have unique FP yields, making it possible to track the level of ^{235}U enrichment. The real-time feedback within radiological facilities such as hot cells could provide rapid comparisons with simulations and codes. These data would be particularly useful in research applications in which a large phase space exists to be explored, namely where many unknowns are held constant, and many experiments are desired.

CONCLUSIONS

This work combined design of experiments, chemometrics, nuclear reactor modeling, an automated latent variable selection script, and sensor fusion approaches to optimize multivariate

regression models for near-real-time monitoring of numerous chemical species in complex environments. The work identified lanthanide FP ratios with nuclear reactor modeling indicative of nuclear fuel burnup characteristics and demonstrated a predictive model for quantification. A D-optimal design training set minimized the training set size, time, and resource consumption for a 7-factor design space. Including species representing CPs directly in the calibration set improved prediction performance. However, depending on the application, satisfactory PLSR models were still developed without including these variables when appropriate feature selection and preprocessing combinations were applied. The fused dataset performed similarly to the MM dataset. Analyzing optical spectra using both methods should be considered, but no clear advantages to low-level sensor fusion were observed in this work. The methodology addresses how to efficiently optimize chemometric models with multiple sensors and numerous species which can be leveraged by numerous fields characterizing complex processes. This work represents a significant step toward deploying optical sensors for monitoring complex chemical processing in harsh chemical environments.

Supporting Information. These files are available free of charge: extended instrument settings, statistical methods, design of experiments, and multivariate analysis sections and FP ingrowth simulation, spectra, and additional experimental results (Microsoft Word).

Author Contributions

All authors contributed to this work and have given approval to the final version of the manuscript.

Funding Sources

Funding for this work was provided by the National Nuclear Security Administration's Office of Defense Nuclear Nonproliferation Research and Development mission under contract DE-AC05-00OR22725. This work used resources at the Radiochemical Engineering Development Center operated by the US Department of Energy's Oak Ridge National Laboratory.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by the National Nuclear Security Administration and performed at Oak Ridge National Laboratory. This manuscript has been authored by UT-Battelle LLC under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

REFERENCES

1. Sadergaski, L. R.; DePaoli, D. W.; Myhre, K. G. Monitoring the Caustic Dissolution of Aluminum Alloy in a Radiochemical Hot Cell Using Raman Spectroscopy. *Appl. Spectrosc.* **2020**, 74 (10), 1252-1262.
2. Lascola, R.; O'Rourke, P. E.; Kyser, E. A. A Piecewise Local Partial Least Squares (PLS) Method for the Quantitative Analysis of Plutonium Nitrate Solutions. *Appl. Spectrosc.* **2017**, 7: 2579-2594.
3. L.R. Sadergaski, K.G. Myhre, L.H. Delmau. Multivariate chemometric methods and Vis-NIR spectrophotometry for monitoring plutonium-238 anion exchange column effluent in a radiochemical hot cell. *Talanta Open* **2022**, 5, 100120.
4. Dang, H. V.; Marini, F. Editorial: Chemometrics-based Spectroscopy for Pharmaceutical and Biomedical Analysis. *Front. Chem.* **2019**, 7, 153.
5. Colle, J-Y. C.; Manara, D.; Geisler, T.; Konings, R. Advances in the application of Raman spectroscopy in the nuclear field. *Spectrosc. Eur.* **2020**, 32 (6), 9-13.
6. Kirsanov, D.; Rudnitskaya, A.; Legin, A.; Babain, V. UV-Vis spectroscopy with chemometric data treatment: an option for on-line control in nuclear industry. *J. Radioanal. Nucl. Chem.* **2017**, 312, 461-470.
7. Tse, P.; Shafer, J.; Bryan, S. A.; Lines, A. M. Quantification of Raman-Interfering Polyoxoanions for Process Analysis: Comparison of Different Chemometric Models and a Demonstration on Real Hanford Waste. *Environ. Sci. Technol.* **2021**, 55, 12943-12950.
8. Sadergaski, L. R.; Toney, G. K.; Delmau, L. H.; Myhre, K. G. Chemometrics and Experimental Design for the Quantification of Nitrate Salts in Nitric Acid: Near-Infrared Spectroscopy Absorption Analysis. *Appl. Spectrosc.* **2021**, 75 (9), 1155-1167.
9. Andrews, H. B.; Sadergaski, L. R. Leveraging visible and near-infrared spectroelectrochemistry to calibrate a robust model for Vanadium(IV/V) in varying nitric acid and temperature levels. *Talanta* **2023**, 259, 124554.
10. Moulin, C.; Decambox, P.; Mauchien, P. Direct Uranium(VI) and Nitrate Determination in Nuclear Reprocessing by Time-Resolved Laser-Induced Fluorescence. *Anal. Chem.* **1996**, 68, 3204-3209.
11. Matusi, T.; Fujimori, H.; Suzuki, K. *J Nucl. Sci. Technol.* Effects of Coexisting Ions upon UO_2^{2+} Fluorescence in Fuel Reprocessing Solutions. **1988**, 25, 868-874.
12. Sinkov, S. I.; Hall, G. B.; Heller, F. D.; Lumetta, G. J. Neptunium redox speciation and determination of its total concentration in dissolved fuel simulant solutions by spectrophotometry *J. Radioanal. Nucl. Chem.* **2022**, 331, 5579-5595.
13. Tse, P.; Shafer, J.; Bryan, S. A.; Nelson, G. L.; Lines, A. M. 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. *Appl. Spectrosc.* **2022**, 76, 173-183.
14. Carnall, W. T.; Fields, P. R.; Rajnak, K. Electronic Energy Levels of the Trivalent Lanthanide Aquo Ions. I. Pr^{3+}, Nd^{3+}, Pm^{3+}, Sm^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, and Tm^{3+} *J. Chem. Phys.* **1968**, 49, 4424-4442.
15. Lines, A. M.; Hall, G. G.; Asmussen, S.; Allred, J.; Sinkov, S.; Heller, F.; Gallagher, N.; Lumetta, G. J.; Bryan, S. A. Sensor Fusion: Comprehensive Real-Time, On-Line Monitoring for Process Control via Visible, Near-Infrared, and Raman Spectroscopy. *ACS Sens.*, **2020**, 5, 2467-2475.
16. Izosimov, I. N.; Firsin, N. G.; Gorshkov, N. G.; Nekhoroshkov. S. N. Detection of lanthanides and actinides in solutions based on laser-induced luminescence and chemiluminescence. *Hyperfine Interact.* **2014**, 227, 271-281.
17. Deniau, B. H.; Decambox, P.; Mauchien, P.; Moulin, C. Time-Resolved Laser-Induced Spectrofluorometry of UO_2^{2+} in Nitric Acid Solutions. Preliminary Results for On-Line Uranium Monitoring Applications. *Radiochim. Acta* **1993**, 61, 23-28.

18. Andrews, H. B.; Sadergaski, L. R.; Cary, S. K. Pursuit of the Ultimate Regression Model for Samarium(III), Europium(III), and LiCl Using Laser-Induced Fluorescence, Design of Experiments, and a Genetic Algorithm for Feature Selection. *ACS Omega*, **2023**, 8, 2281-2290.
19. Sadergaski, L. R.; Andrews, H. B. Simultaneous quantification of uranium(VI), samarium, nitric acid, and temperature with combined ensemble learning, laser fluorescence, and Raman scattering for real-time monitoring. *Analyst* **2022**, 147, 4014-4025.
20. Casella, A. J.; Levitskaia, T. G.; Peterson, J. M.; Bryan, S. A. Water O-H Stretching Raman Signature for Strong Acid Monitoring via Multivariate Analysis. *Anal. Chem.* 2013, 85 (8), 4120-4128.
21. Sadergaski, L. R.; Hagar, T. J.; Andrews, H. B. Desing of Experiments, Chemometrics, and Raman spectroscopy for the Quantification of Hydroxylammonium, Nitrate, and Nitric Acid. *ACS Omega*, **2022**, 7, 7287-7296.
22. Comino, F.; Ayora-Canada, M. J.; Aranda, V.; Diaz, A.; Dominguez-Vidal, A. Near-infrared spectroscopy and X-ray fluorescence data fusion for olive leaf analysis and crop nutritional status determination. *Talanta* **2018**, 188, 776-684.
23. Zhu, S.; Song, Z.; Shi, S.; Wang, M.; Jin, G. Fusion of Near-Infrared and Raman Spectroscopy for In-Line Measurement of Component Content of Molten Polymer Blends. *Sensors* **2019**, 19, 3463.
24. Roach, B. D.; Rogers, K. T.; Zirakparvar, N. A.; Delashmitt, J. S.; Metzger, S. C.; Manard, B. T.; Keever, T. J.; Giaquinto, J. M.; Hexel, C. The need for speed – Burnup determination of spent nuclear fuel. *Talanta Open*, **2022**, 6, 100152.
25. Wilson, B. A.; Conant, A.; Ulrich, T. L.; Kercher, A.; Sadergaski, L. R.; Gerczak, T.; Nelson, A. T.; Petrie, C. M.; Harp, J.; Shields, A. E. Nuclear fuel irradiation testbed for nuclear security applications. *Front. Nucl. Eng.* **2023**, 2, 1123134.
26. Andrews, H. B.; McFarlane, J.; Myhre, K. G. Monitoring Noble Gases (Xe and Kr) and Aerosols (Cs and Rb) in a Molten Salt Reactor Surrogate Off-Gas Stream Using Laser-Induced Breakdown Spectroscopy (LIBS). *Appl. Spectrosc.* **2022**, 76 (8), 877-886.
27. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.,* **2011**, 12, 2825-2830.
28. Werner, C. J.; et al. MCNP version 6.2 release notes. No. LA-UR-18-20808. Los Alamos, New Mexico: Los Alamos National Laboratory, 2018.
29. Goblirsch, T.; Mayer, T.; Penzel, S.; Rudolph, M.; Borsdorf, H. In Situ Water Quality Monitoring Using an Optical Multiparameter Sensor Probe. *Sensors*, **2023**, 23, 9545.
30. Picayo, G. A.; Etz, B. D.; Vyas, S.; Jensen, M. P. Characterization of the ALSEP Process at Equilibrium: Speciation and Stoichiometry of the Extracted Complex. *ACS Omega*, **2020**, 5, 8076-8089.
31. Nickson, I. D.; Boxall, C.; Jackson, A.; Whillock, G. O. H. A spectrophotometric study of cerium IV and chromium VI species in nuclear fuel reprocessing process streams. *IOP Conf. Ser. Mater. Sci. Eng.* **2010**, 9, 012011.
32. Czitrom, V. One-Factor-at-a-Time Versus Designed Experiments. *Am. Stat.* **1999**, 53 (2), 126-131.
33. González, A. G.; Herrador, M. A.; Asuero. A.G. Intra-laboratory testing of method accuracy from recovery assays. *Talanta*, 48, 729-736.

For Table of Contents Only