Analyst

Simultaneous quantification of uranium(VI), samarium, nitric acid, and temperature with combined ensemble learning, laser fluorescence, and Raman scattering for real-time monitoring

Journal:	*Analyst*
Manuscript ID	AN-ART-06-2022-000998.R1
Article Type:	Paper
Date Submitted by the Author:	02-Aug-2022
Complete List of Authors:	Sadergaski, Luke; Oak Ridge National Laboratory, Isotope and Fuel Cycle Technology Andrews, Hunter; Oak Ridge National Laboratory, Radioisotope Science and Technology Division;

Simultaneous quantification of uranium(VI), samarium, nitric acid, and temperature with combined ensemble learning, laser fluorescence, and Raman scattering for real-time monitoring

Luke R. Sadergaski, Hunter B. Andrews*

Radioisotope Science and Technology Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37830, USA

*sadergaskilr@ornl.gov

ABSTRACT: Laser-induced fluorescence spectroscopy (LIFS), Raman spectroscopy, and a stacked regression ensemble was developed for near real-time quantification of uranium (VI) (1–100 μg·mL^{-1}), samarium (0–200 μg·mL^{-1}) and nitric acid (0.1–4 M) with varying temperature (20°C– 45°C). LIFS applications range from fundamental lab-scale studies to real-time process monitoring at industrial levels, such as nuclear reprocessing applications, provided the phenomena

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).

affecting the fluorescence spectrum are accounted for (e.g., absorption, quenching, complexation). Multiple chemometric models were examined and compared to a more traditional multivariate regression approach called partial least squares (PLS). Results obtained on synthetic samples selected using D-optimal experimental design indicated that a stacked regression method, which included ridge regression, random forest, PLS, and an eXtreme gradient boost algorithm, successfully measured uranium (VI) concentrations directly in nitric acid without measuring luminescence lifetimes or standard addition. The top model resulted in percent root-mean-square error of prediction values of 5.2, 1.9, 3.0, and 2.3% for U(VI), Sm^{3+}, HNO_3, and temperature, respectively. The approach may be useful for quantifying fluorescent fission products (e.g., Sm^{3+}) to provide information on burnup of irradiated nuclear fuel. This novel framework reinforces the applicability of LIFS for real-time applications in nuclear fuel cycle applications.

KEYWORDS: uranium, samarium, machine learning, real-time monitoring, optimal designs, ensemble learning, stacked regression, multivariate analysis

1. INTRODUCTION

Optical spectroscopy is a powerful option for in-line process control in many industrial applications and nuclear fuel cycle reprocessing separation schemes like PUREX (**P**lutonium **U**ranium **RE**duction E**X**traction).[1–5] PUREX is a liquid–liquid extraction process that purifies uranium and plutonium by first extracting uranium (U) and plutonium (Pu) from aqueous solutions (3–4 M nitric acid) with a hydrocarbon phase containing tributyl phosphate. This extraction step is followed by a second U and Pu partitioning stage and a third U stripping stage in dilute nitric acid.[6] All PUREX-related tasks must be performed remotely in heavily shielded hot cell

enclosures. Additional complications arise from radiation-induced decomposition and criticality risks, which makes this operation one of the most complicated chemical processes ever attempted. Processing must be supported by numerous analytical measurements. Traditional techniques, including radiochemistry (e.g., alpha and gamma spectroscopy) or inductively coupled plasma mass spectrometry, generally require retrieving samples which require significant dilutions (1,000–10,000-fold) and transfer out of the hot cell for analysis.[2] Hot cell measurements using optical spectroscopy are much faster and can be acquired in situ, in real-time using fiber-optic cables.

One form of optical spectroscopy, time-resolved laser-induced fluorescence spectroscopy (TRLIFS), was evaluated extensively for monitoring PUREX streams (e.g., raffinate) several decades ago,[7–11] but little work has been done in the last decade. More recently, Raman spectroscopy and UV-Vis spectrophotometry have been evaluated for monitoring applications, although these options are orders of magnitude less sensitive to uranium (VI) concentration than TRLIFS.[1,2,12–14] TRLIFS is highly sensitive and selective to many actinide (An) species. It is often used for ultratrace analysis and complexation studies with both spectral and temporal features.[15,16] TRLIFS can detect several actinides including UO_2^{2+}, Am^{3+}, Cm^{3+}, Cf^{3+}, Bk^{3+}, and Es^{3+}. Only UO_2^{2+} and Cm^{3+} have a suitably large energy gap ($>10^4$ cm^{-1}) for the luminescence level, while the others have a smaller gap, resulting in luminescence quantum yields highly sensitive to specific compounds and solution types.[15] It is challenging to measure the aqueous luminescence properties of other actinides like Np and Pu due to strong nonradiative quenching by the solvent (e.g., O–H band). This effect can be lessened using other solvent types such as heavy water (D_2O).[17] Despite the incredible sensitivity of the time-resolved technique, the analysis time is often 10–15 minutes, which is too slow for situations where real-time feedback is required.[10]

Laser-induced fluorescence spectroscopy (LIFS) is useful for the qualitative and quantitative analysis of molecular species such as the free uranyl (UO_2^{2+}) and uranyl nitrate complexes ($UO_2(NO_3)_n^{(2-n)+}$, n = 1 or 2) in aqueous solutions or U(VI) solid-state compounds.[18] Fluorescence spectra can be measured using a charge-coupled device to increase the timeliness of data collection, as opposed to single-channel photomultiplier tubes (PMTs) often used for time-resolved data. However, applying laser fluorometry is difficult in complex systems with overlapping bands, noise, matrix effects, chemical interactions, and baseline offsets.[5,11,15,19] Although it is possible to reroute the flow during processing to maintain a constant temperature and generate U(VI) complexes to improve quantum yield (i.e., sensitivity),[19] it is preferable to perform the analysis in-line (i.e., directly in nitric acid) and without complicating hot cell operations. This necessitates directly accounting for the complicating effects of a nitric acid medium and other real-world process conditions such as temperature fluctuations.

Dynamic temperatures drastically change the lifetime and intensity of U(VI) fluorescence, resulting in two highly correlated independent variables (i.e., multicollinearity), possibly the most challenging quenching phenomena.[5,10,11,15] It is critical to account for temperature because high-level waste solutions vary from 30°C–40°C and PUREX operations often take place at temperatures ranging from 20–40°C.[5] Additional complicating factors arise because many fuel cycle processing streams contain a variety of fission and corrosion products with absorption or fluorescence properties that can complicate LIFS spectra.[9] Lanthanide (Ln) fission products include Ce^{3+}, Pr^{3+}, Nd^{3+}, Sm^{3+}, Eu^{3+}, and Gd^{3+}. Ln elements such as Eu^{3+}, Tb^{3+}, Gd^{3+}, Dy^{3+}, Sm^{3+}, Ce^{3+}, and Tm^{3+} can be characterized by TRLIFS.[14,20–23] Two lanthanides, Eu^{3+} and Sm^{3+}, have visible luminescence properties and range from ~10 to 200 $\mu g \cdot mL^{-1}$ in PUREX raffinate streams.[9] The fission yield of Sm^{3+} is generally greater than Eu^{3+} and is less characterized in the literature.

LIFS could provide real-time feedback regarding nuclear fuel burnup and provide operational benefits compared to other optical techniques.[24] Multivariate regression analysis has not been used to describe U(VI) fluorescence spectra under relevant conditions. It is important to reexamine this system with modern data analytics to achieve greater success with the LIFS technique.

This work explores the use of multivariate chemometrics, or machine learning, to account for these multifaceted effects. One of the most traditional supervised techniques is called partial least squares regression (PLSR).[25–28] This factor analysis method iteratively relates two data matrices, the independent X (i.e., spectra) and dependent Y (i.e., concentrations), using combinations of latent variables (LV). PLSR models are built using a training set that covers the expected conditions and a validation set that tests the model's ability to predict samples not included in the training set. These samples can be selected using optimal experimental designs that are the most flexible and effective option when a small number of experimental runs is desired.[29–33]

Although PLSR has been used with great success to model systems with overlapping spectral features in numerous systems,[27] it does not always account for systems with a high degree of multicollinearity.[34,35] Multicollinearity must be accounted for in a LIFS system for monitoring U(VI) concentration because two independent variables (i.e., U(VI) concentration and temperature) are highly correlated in the regression model. To address this, ensemble learning combines multiple models to make a prediction in classification or regression problems to improve robustness and accuracy compared to single models. This process can be divided into three phases: (1) develop a set of candidate models (i.e., generation phase); (2) select a subset of the models (i.e., pruning phase); and (3) combine models to generate predictions (i.e., integration phase). In addition to PLSR, several advanced models are considered in this article, including ridge regression (RR), random forest (RF), and an eXtreme gradient boost (XGB) algorithm.[36–42] These

are evaluated individually and compared to a stacked regression approach, a form of ensemble learning. The primary goal of this work is to demonstrate how ensemble learning methods can be used to handle the dynamic, overlapping, covarying, and nonlinear spectral response to provide a new U(VI) analysis method independent of time-resolved fluorescence spectra.

Here, we demonstrate how to combine ensemble learning, LIFS, and Stokes Raman scattering to account for varying temperature, conditions that historically limit the timeliness of U(VI) monitoring applications. This work also includes the simultaneous quantification of a lanthanide fission product (i.e., Sm^{3+}). Calibration and validation fluorescence spectral data sets were selected by determinant(D)-optimal designs to minimize the samples required in the training set, which spanned U(VI) (1–100 $\mu g \cdot mL^{-1}$), Sm^{3+} (0–200 $\mu g \cdot mL^{-1}$), HNO_3 (0.1–4 M) concentrations and temperatures (20°C–45°C), conditions highly applicable to the nuclear fuel cycle. Specific points of scientific advancement covered in this work include: (1) the ensemble method allows for quantitative U(VI) predictions and varying temperature without recording lifetimes, (2) simultaneous quantification of a fluorescent lanthanide (i.e., Sm^{3+}) in the milligram-per-liter range, (3) stacked regression that combines LIFS and Raman spectroscopy, and (4) the augmented D-optimal design allows training/validation set samples to be effectively selected without user bias while minimizing the number of samples. Many spectroscopy-enabled online monitoring studies tend to neglect the temperature variable and include too many user-selected samples to be easily implemented in restrictive hot cell environments.[26] This work clearly articulates the challenges that arise due to fluctuating temperature and best practices to correct for such phenomena; this is applicable to many optical spectroscopy online monitoring applications extending within and beyond nuclear fuel cycle applications.

2. METHODS

All chemicals were commercially obtained (ACS grade) and used as received unless otherwise stated. Concentrated HNO_3 (70%) was purchased from Sigma-Aldrich. Certified 10,000 μg·mL^{-1} uranium (^{238}U, depleted) and samarium inductively coupled plasma optical emission spectroscopy standard solutions in 5% nitric acid were purchased from SPEX CertiPrep and Inorganic Ventures, respectively. Samples were prepared using deionized water with Milli-Q purity (18.2 MΩ·cm at 25°C).

2.1 Sample preparation

Calibration and validation samples contained uranium (1–100 μg·mL^{-1}), samarium (0–200 μg·mL^{-1}), and HNO_3 (0.1–4 M) and chosen to cover the anticipated solution conditions. Samples were prepared gravimetrically using volumetric pipettes. Each sample was prepared in individual 2 mL plastic microcentrifuge tubes (VWR Scientific, 525-1160) prior to spectroscopic analysis. A fluorescence 3-in-1 flow cuvette, purchased from Hellma (584.4-Q-1) with 5 × 2.5 mm optical path lengths, was used for each measurement to ensure consistent optical quality. The cuvette was periodically rinsed with 2% nitric acid and stored with ultrapure water on lint-free Kimwipes. The cuvette Z-height of 8.5 mm was necessary to accommodate Quantum Northwest's qpod 2e temperature-controlled sample compartment holder purchased from Avantes (CUV-UV/Vis-TC). Two collimating lenses (CUV-TC-QCL-UV) were placed at 90° in the sample compartment. Fluorescence measurements were performed at varying temperatures (i.e., 20°C–50°C) with an accuracy of ±0.05°C. Sample solutions were thermally equilibrated for at least 2 min prior to recording each spectrum to eliminate spectral variations due to temperature fluctuation. A syringe was used to inject the rinse and sample solutions.

2.2 Fluorescence and Raman spectroscopy

Laser fluorescence and Stokes Raman spectra were collected with a fully automated imaging iHR 320 spectrometer (Horiba Scientific) and a CW (continuous wave) LBX 405 nm laser (Oxxius) operating at 100 mW. Two multimode fibers—a 105 µm core diameter (M105L02S-A) and a 600 µm core diameter fiber (M134L01)—were used on the excitation and emission side, respectively. Static measurements were recorded in triplicate from 410 to 790 nm using a 600 grooves mm^{-1} grating and a 100 µm slit size. Each spectrum comprised 5,585 data points.

Lifetimes were collected using a Fluorolog-QM spectrometer (Horiba) and a DeltaTime kit for a DeltaDiode 405 nm laser (DD-405L) source operating with an average power of 2 mW and average pulse width of 50 ps. Single-photon-counting statistics with time-correlated single-photon counting were used to calculate lifetimes using a single-channel R928P PMT. The lifetime analysis depends on a model or fitting function that describes the decay of luminescence intensity. The lifetime was calculated using a fitting algorithm *D(t)* with the PowerFit-10 application in Horiba software by Equation 1:

$$D(t) = \sum a_i \exp\left\{\frac{-t}{\tau_i}\right\}, \quad (1)$$

where a_i is the preexponential factor, t is time, and τ_i is the fluorescence lifetime. The room temperature was stable at 22°C during lifetime measurements. Excitation spectra were acquired in steady state using a 75 W Xenon arc lamp (see Supporting Information [SI]).

2.3 Design of experiments

Experimental designs were built using Design-Expert (v.11.0.5.0) by Stat-Ease Inc., within the Unscrambler software package by Camo Analytics. Optimal designs incorporate mixture and process variables, contain different high and low components, and feature constraints with factor

limits. D-optimal samples were chosen by iteratively minimizing the determinant of the variance–covariance matrix X^TX using a quadratic process order.[25] A combination of point and coordinate exchange search options was used to select points. Fraction of design space (FDS) was used to evaluate the model and calculated by mean error type: $\delta = 2$, $\sigma = 1$, and $\alpha = 0.05$.[33]

A minimum of 10 model points was required for three numeric factors (i.e., U(VI), Sm, and HNO_3 concentrations). The design was augmented with 15 additional lack-of-fit (LOF) points, which were included either as calibration or validation samples. LOF points are chosen to maximize the distance to other runs while maintaining the optimality criterion.[32] Temperature was treated in a separate design. At least three model points were required for a single numeric factor (i.e., temperature), and the design was augmented using 2 LOF points. Combining the two designs resulted in a total of 125 samples.

2.4 Multivariate analysis and preprocessing

Several chemometric methods were evaluated including PLSR, RR, RF, an XGB algorithm, and combining multiple methods through stacked regressions. PLSR iteratively relates two data matrices, the independent X (i.e., spectra) and dependent Y (i.e., concentrations), using combinations of latent variables. The optimal number of latent variables is selected through cross validation (CV); the optimal number typically includes up to the last latent variable that marks a significant increase in the explained variance (or a significant reduction in the error of prediction). PLSR can be implemented in two forms, the first (PLS-1) models the response of only a single factor and the second (PLS-2) models the response of multiple factors simultaneously. PLS-2 is typically used to model systems with multicollinearity; however, in some cases multiple PLS-1 models can have better predictive capabilities by leveraging different preprocessing strategies (i.e., trimming, smoothing, derivatives, etc.). RR is similar to ordinary least squares regression;

however, instead of simply minimizing the sum of squared error, there is an additional factor, α, that penalizes the slope of the regression.[39,41] This factor is selected through CV and must be greater than zero. RR is typically useful when analyzing data with high multicollinearity (i.e., correlated independent variables). RF is a supervised form of machine learning that fits a user-defined number of independent decision trees to the data and then combines the predictions of the many trees to provide a single prediction.[39,40] The XGB algorithm is an open-source implementation of the gradient boosting algorithm that uses a loss function to prevent overfitting. XGB also uses a multitude of decision trees like RF, but actively uses residual information to optimize the model.[39]

Stacked regression is an ensemble machine learning method in which a final metamodel is trained on the outputs of several submodels to provide a single combined estimate. The final model regression is built using CV of the submodel predictions, learning the best way to combine the submodel predictions.[36,40,42] This allows for multiple, typically different, regression methods to be employed simultaneously on a single data set, allowing the strengths of each model to be retained while mitigating the error of the individual regressions.[36,42] Ensemble methods, like RF, help to mitigate issues of overfitting and typically provide more accurate predictions than the single models.

Spectral data was preprocessed prior to modeling including smoothing, derivatives, and trimming to optimize performance. Smoothing removes instrument noise, and the use of derivatives makes calibration models more resistant to baseline shifts expected in monitoring applications. The smoothing and derivatives were performed using several Savitzky–Golay (SG) filters. SG derivatives fit a polynomial to the data based on a user-defined number of left/right smoothing points and polynomial order. Higher than first and second derivatives were not included

here because they suppressed the relatively broad spectral features needed for regression. Numerous derivative orders, polynomial orders, and smoothing points were tested.[30] Trimming the spectra reduces the dimensionality of large spectra files and can help reduce the error of prediction in multivariate modeling. The trimmed region included in the chemometric model was varied based on the behavior being modeled. All regression models and data preprocessing were completed in Python 3 using modules from the Scikit Learn and XGBoost packages.[38,40]

2.5 Statistical comparison

Model performance was evaluated using calibration, CV, and validation (i.e., prediction) metrics. The most important calibration/validation statistics typically include R^2 correlation values, root mean square error (RMSE) of the calibration (RMSEC), and RMSE of the CV (RMSECV). Although these statistics may suggest that a PLSR model is satisfactory, testing the prediction performance of PLSR models on samples not included in the training set is important because RMSECV is only an estimate.[5] Prediction statistics typically include RMSE of the prediction (RMSEP), RMSEP%, bias, and standard error of prediction. RMSEs for the calibration, CV, and validation were calculated using Equation 2:

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

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 model values using Equation 3:

$$\mathrm{RMSEP\%} = \frac{RMSEP}{y_{med}} \times 100\%, \tag{3}$$

where y_{med} represents the median of each analyte concentration range. Each RMSE value is in units of analyte concentration. In general, lower RMSEC, RMSECV, RMSEP, and RMSEP% values

indicate better model performance. A Tukey-Kramer significance test was used to statistically compare the RMSEP values for multiple regression models following a method outlined previously.[29,30] Additional details can be found in the SI.

3. RESULTS AND DISCUSSION

3.1 Fluorescence and Raman spectra

The absorption and photoluminescence spectrum of the uranyl ion (UO_2^{2+}) is extensively characterized in the literature.[13] Each electron in this nearly linear moiety is paired, thus the ground-state electronic level is a singlet.[18] Higher-energy levels occur when one of the bonding electrons is transferred to the 5f nonbonding atomic orbitals of the uranium ion ~20,000 cm^{-1} (500 nm) above the ground singlet. Photoexcitation occurs when photons with a higher energy than the first excited electronic level are absorbed. The excited uranyl ion relaxes rapidly by a nonradiative process to the first excited electronic level, followed by fluorescence to the symmetric and asymmetric vibrational levels associated with the ground-state singlet. This process is described by a notional energy-level diagram in Figure 1a.

The Sm^{3+} 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, respectively (Figure 1b). The electronic structures of lanthanides elements are characterized primarily by the 4f orbital.[21] The 4f orbitals are shielded by the filled 5s and 5p shells, indicating that the surrounding matrix has a minimal effect on $4f^n$ energy levels and corresponding optical transitions.[21] As a result, 4f electrons do not play a significant role in chemical bonding, and wavelength-dependent shifts due to the host medium are minimal. Most research on lanthanide luminescence in aqueous environments has been concerned with Eu^{3+} and Tb^{3+}. The other two visibly luminescent ions, Sm^{3+} and Dy^{3+}, have

received less attention because they have inferior luminescence quantum yields (i.e., more efficient nonradiative relaxation). Three out of the four Sm^{3+} emission bands significantly overlap with the uranyl fluorescence spectrum (see Figure 1).

With increasing temperature, the fluorescence intensity of U(VI) decreases dramatically (i.e., via dynamic quenching) because the molecular collisional rate increases (Figure 1a). The shape of the fluorescence spectrum changes minimally with increasing temperature (Figure S1). This implies that the change in temperature has a minimal effect on the nature of the UO_2^{2+} species. The temperature-induced radiative deexcitation rate of UO_2^{2+} is likely due to a chemical process involving hydrogen abstraction from the solvent molecules coordinating UO_2^{2+}.[5] On the other hand, the fluorescence spectrum of Sm^{3+} changes slightly in wavelength position and intensity as a function of temperature. These thermochromic shifts likely correspond to slight variations in the coordination environment surrounding Sm^{3+}.[23]

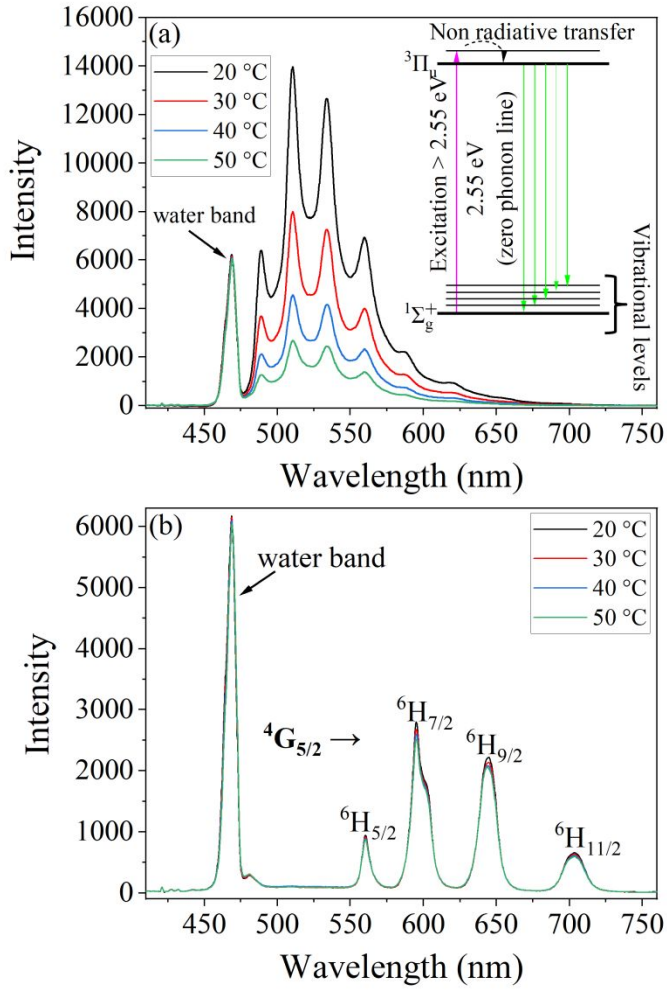

Figure 1. Laser fluorescence spectra (λ_{ex} = 405 nm) of an aqueous solution containing 100 μg·mL^{-1} uranium (a) or 100 μg·mL^{-1} Sm^{3+} (b) in 1 M HNO_3 from 20°C–50°C. A notional uranyl energy-level diagram is included in the figure, and each Sm^{3+} transition is labeled. The Raman NO_3^- symmetric N–O stretch and the O–H stretching band appeared at 421 nm and 455–475 nm, respectively.

The U(VI) spectrum changes significantly as a function of acid concentration (Figure 2a) due to the formation of free uranyl species (UO_2^{2+}) and uranyl nitrate complexes ($UO_2(NO_3)_n^{(2-n)+}$, n = 1 or 2) at acid concentrations >~0.1 M HNO_3, which coexist in various proportions. Each species—

UO_2^{2+}, $UO_2NO_3^+$ and $UO_2(NO_3)^{2+}$—has a unique spectrum contribution. The maximum intensity of U(VI) fluorescence from ~0.5 to 5 M is relatively consistent; however, the lifetime changes significantly from 3.1–1.1 μs over this range and continues to decrease at even higher HNO_3 concentrations (see SI). The Sm^{3+} emission spectrum is also highly sensitive to changes in nitric acid concentration and likely corresponds to the formation of nitrate complexes (Figure 2b). To use the entire spectrum for regression, multivariate chemometrics must be applied to account for the overlapping and covarying spectral features.

The electronic transitions of Sm^{3+} vary in spectral shape and intensity with changing nitric acid concentration. Despite these solvatochromic shifts, the lifetime remains relatively constant from 1 to 5 M HNO_3 at 3.2–3.6 μs. (Figure S3). Different energy levels in lanthanide fluorescence spectra occur due to several interactions within the ion, including the Coulombic interaction (10^4 cm^{-1}), spin-orbit coupling (10^3 cm^{-1}), and crystal field splitting (10^2 cm^{-1}).[21] These free ion levels are described by the term symbols $^{(2S+1)}L_J$ where 2S + 1 denotes the total spin multiplicity, L the total orbital angular momentum, and J denotes the total angular momentum of the f electrons. The electric field in a coordinating environment further splits individual J-levels. This splitting appears as fine structure on each band. This information is only visible with higher-resolution

spectrometers and is often ignored in the literature. However, fine structure is useful for determining symmetry and coordination environment.[15] The fine structure in the Sm^{3+} emission bands is particularly evident in the $^4G_{5/2} \rightarrow {}^6H_{7/2}$ emission line near 596 nm. The shoulder near 604 nm becomes less pronounced with increasing acidity. This suggests that the transitions acquire strength at least in part due to the coordinating environment and have electric dipole (ED) character. ED-allowed transitions are more sensitive to the coordinating environment, and stronger ligand fields lead to more intense transitions.

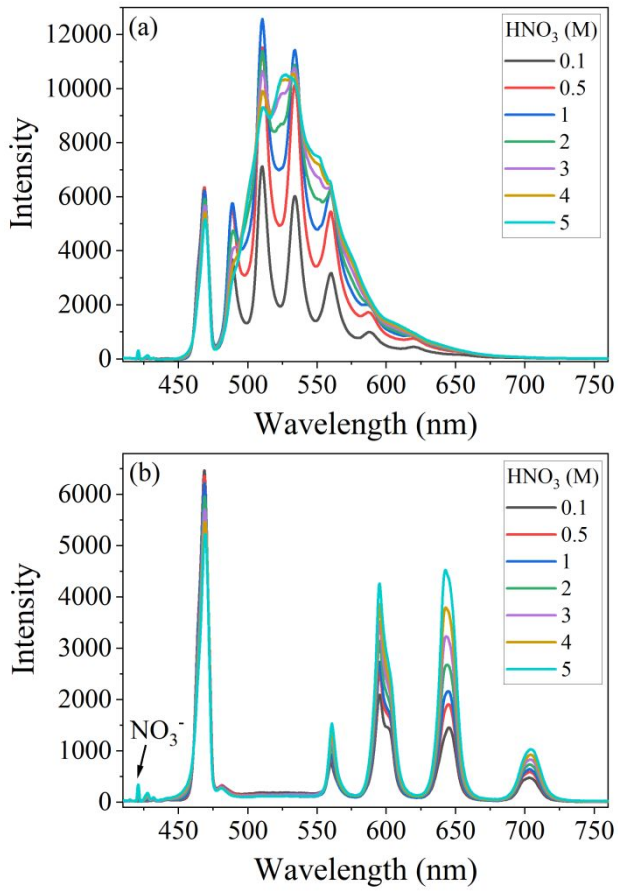

Figure 2. Example laser fluorescence spectra (λ_{ex} = 405 nm) of solutions containing 100 μg·mL^{-1} U(VI) (a) and 100 μg·mL^{-1} Sm^{3+} (b) with varying nitric acid concentration (0.1 M–5 M HNO_3). The Raman nitrate (NO_3^-) symmetric N–O stretch peak is labeled (b).

In addition to U(VI) and Sm^{3+} emission peaks, Stokes Raman scattering features corresponding to free acid (H^+), nitrate anions (NO_3^-), and the O–H stretching region were identified simultaneously.[26,30] Unperturbed nitrate ions have three Raman active bands including ν_1 (~1048 cm^{-1}), ν_3 (~1415 cm^{-1}), and ν_4 (~717 cm^{-1}). The O–H vibrational stretching region consists of several overlapping bands attributed to various H_2O and O–H (free and bound) vibrations (455–475 nm). An isosbestic point at 461.9 nm was observed with increasing nitric acid concentration

(Figure 2). This Raman band is sensitive to the presence of cation, anions, ion strength, and temperature and is generally best described using multivariate data analysis.[25,26]

One advantage of using an excitation wavelength of 405 nm is that the water band at ~460 nm does not overlap significantly with the absorption spectrum of U(VI) (see SI). Either 337 nm or 355 nm laser excitation sources are normally used for U(VI) fluorescence studies.[8] These wavelengths coincide with absorption bands of nitric acid, uranium, or several fission products, complicating measurement due to self-absorption effects.[5,9] The fluorescence signal is dependent on the optical pathlength for laser excitation and the pathlength for emission collection.[10] The complications that arise due to self-absorption are minimized in this work, which utilizes a 405 nm laser and a smaller pathlength cuvette (i.e., 5 × 2.5 mm) instead of the more common 10 × 10 mm.

3.2 Univariate Analysis to LOD and LOQ

A key component of a quality analytical chemistry paper includes the transparent calculation of limit of detection (LOD) and limit of quantification (LOQ). For this experiment, spectra were collected at room temperature (22°C) using a 4 s integration time and 100 mW laser power at 405 nm. Calibration regression curves were generated for the primary 510 nm U(VI) peak, the most intense Sm^{3+} emission line 595 nm peak, and the most resolved line relative to uranium at 702 nm (see SI). The calibration curves for each emission line were used to calculate LOD and LOQ by Equations 4 and 5:

$$LOD = \frac{3 \cdot s}{m} \quad (4)$$

$$LOQ = \frac{10 \cdot s}{m} \quad (5)$$

The variable *s* refers to standard deviation or noise of the blank, and *m* is the slope obtained by plotting the intensity of the peak versus concentration. LOD and LOQ values are shown in Table 1. Sample U(VI) and Sm^{3+} concentrations ranged from 1 to 8 μg·mL^{-1} in 1.0 M nitric acid (Figure S2). LODs could improve using longer integration times, higher laser power, or more efficient optics. LODs for Sm^{3+} would improve at higher acidity (Figure 2), and U(VI) would decrease at higher temperatures. These LOD and LOQ values do not reflect the exact detection limits for multivariate regression models but serve as a benchmark for the system used in this study. Due to the complicated nature of the emission bands, multivariate approaches were applied to the measurements.

Table 1. LOD and LOQ of U(VI) and Sm^{3+} in 1 M HNO_3 using a 405 nm laser at 22°C and 4 s integration time.

Species	LOD (μg·mL^{-1})	LOQ (μg·mL^{-1})	slope (m)
U(VI) 510.3 nm	0.019	0.064	132
Sm^{3+} 595.2 nm	0.062	0.21	27.4
Sm^{3+} 702.0 nm	0.75	2.5	7.32

3.3 D-optimal design sample selection

Models were built using calibration and validation sets selected by D-optimal experimental designs. Recent work has shown that this approach can minimize resource consumption while maintaining or improving PLSR prediction performance. Selecting training sets using a one-factor-at-a-time approach generally results in many samples, particularly when more than two factors are considered. The expected application of this work is monitoring in a restrictive glove box or hot cell environment. Therefore, D-optimal designs were evaluated to minimize the number of samples

in the training set as well as minimize time and resource consumption.[5] A D-optimal design (25 points) was used to select the concentrations studied in this work. This design comprised 10 required model points and was augmented with 15 LOF points. Inclusion of LOF points generally indicates that a higher-order process model is needed. To achieve an FDS of 0.99, five LOF points were included in the calibration set. The 10 remaining LOF points were used as the validation set.

U(VI) concentration spanned the conditions expected in raffinate waste streams (1–100 $\mu g \cdot mL^{-1}$).[5,9] Detecting less than 1 $\mu g \cdot mL^{-1}$ uranium generally is not cause for concern in most applications. The model covered the acid concentration range from 0.1 to 4 HNO_3, which covers both a raffinate stream (i.e., 3–4 M HNO_3) and low-acid-strip solutions.[5] Sm^{3+} concentrations were chosen based on potential burnups.[9] Generally, Sm^{3+} is found in raffinate solutions at a higher acid concentration. Accounting for acid dependence in this data set adds complexity, which tests the limits of the modeling approach. It was included in some model points and not included in others to determine if it strengthened U(VI) model predictions, particularly with varying temperature.

Table 2. D-optimal selected concentrations with space and build type.

Run	U(VI) ($\mu g \cdot mL^{-1}$)	HNO_3 (M)	Sm^{3+} ($\mu g \cdot mL^{-1}$)	Space type	Build type
1*†	100.00	3.22	160.00	Plane	Lack of fit
2	**100.00**	**0.10**	**0.00**	**Vertex**	**Model**
3	82.18	1.66	119.00	Interior	Lack of fit
4	55.45	0.12	0.10	Interior	Lack of fit
5	3.97	0.12	108.74	Interior	Lack of fit
6*	30.70	1.11	55.00	Interior	Lack of fit
7	13.38	4.00	87.00	Plane	Lack of fit
8*	63.41	4.00	130.00	Plane	Lack of fit
9	52.48	4.00	40.00	Plane	Lack of fit
10	61.39	2.48	199.00	Interior	Lack of fit
11	**1.00**	**4.00**	**0.00**	**Vertex**	**Model**
12	**100.00**	**1.50**	**200.00**	**Edge**	**Model**
13	15.85	2.36	0.00	Plane	Lack of fit

14†	30.70	1.27	163.00	Interior	Lack of fit
15	**59.91**	**2.42**	**0.00**	**Plane**	**Model**
16	**100.00**	**4.00**	**70.27**	**Edge**	**Model**
17	**1.00**	**2.42**	**119.00**	**Plane**	**Model**
18	**36.15**	**4.00**	**200.00**	**Edge**	**Model**
19*†	44.32	2.50	104.00	Interior	Lack of fit
20†	100.00	0.20	86.82	Plane	Lack of fit
21	**60.26**	**0.10**	**120.00**	**Plane**	**Model**
22	**1.00**	**0.10**	**0.00**	**Vertex**	**Model**
23*	1.00	3.03	200.00	Edge	Lack of fit
24	100.00	1.84	33.00	Plane	Lack of fit
25	**1.00**	**0.10**	**200.00**	**Vertex**	**Model**

Note: Abbreviations used in this table are derivative (Der.) and polynomial (Poly.). *LOF points included in the construction of the original calibration models. †Optimal LOF points for U(VI) calibration model. Required model points are bolded.

A second D-optimal design was generated for temperature from 20°C to 45°C (see SI). This design included three required model points at 20°C, 32.5°C, and 45°C and two LOF points at 26.25°C and 38.75°C. Although it is possible to combine temperature as a fourth factor in the concentration design, the number of temperature points in such a design may not account for the variation necessary to model it. Additionally, a four-factor design would create additional concentrations, resulting in more samples. Thus, the variable temperature, which is unique relative to concentration, was treated separately.

3.4 Stacked regression model development

In addition to the required model points, five LOF samples (1, 6, 8, 19, and 23) were arbitrarily chosen and included in the training set to begin building models. All five temperature levels were used to best model the effect on analyte spectral signatures. Each model was optimized by minimizing the RMSEP through numerous reiterations. PLSR (PLS-2) was attempted first, as it is one of the most widely applied techniques to correlate convoluted and covarying spectral features to analyte concentration.[2,3,27,28] The global PLSR model was built using most of the spectrum (410–750 nm) after applying a first derivative with a first-order polynomial and 31 smoothing points (1,

1, 31). Eight latent variables (i.e., factors) were included in the model based on the RMSECV versus latent variable plot (Figure S6). The number of latent variables was chosen based on when there was the last significant reduction in RMSECV (>10%) when an additional latent variable was added. The addition of the eighth latent variable corresponds to a 20, 69, 84, and 3.2% reduction in the RMSECV for U(VI), HNO_3, Sm^{3+}, and temperature, respectively. The global PLSR model predicted Sm^{3+} and HNO_3 concentration relatively well (see Table 3). The model was unable to account for the multicollinear effects of temperature and U(VI) concentration on the spectral intensity. Parity plots showing the wide spread of temperature and U(VI) predictions against the reference values are shown in Figure S7.

Individual PLSR models (PLS-1) were built using only the spectral regions most related to each species to reduce issues of multicollinearity. As denoted in Table 3, fewer factors were needed in the individual PLSR models due to trimming and reducing the signal dimensionality.[37] This lowered the RMSEP: 18, 63, and 60% difference for Sm^{3+}, HNO_3, and temperature (20%). However, this exercise only slightly lowered U(VI) predictions (11% difference), clearly stressing that U(VI) predictions are impacted the most by multicollinearity. A Tukey-Kramer test was performed at a 95% confidence level to verify there was a significant difference between the global and trimmed PLSR models. The two models showed similar levels of prediction bias for all factors and showed a significant improvement in prediction error for all factors except for Sm^{3+}. A description of the Tukey-Kramer test can be found in the Supporting Information along with plots of the bias and standard error of prediction confidence bands (Figure S8).

Next, other regression methods, including RR, RF, XGB, and stacked models, were investigated. Each regression model was optimized by tuning the preprocessing hyperparameters and trimming limits. The optimal models for Sm^{3+} and HNO_3 were developed using trimmed spectra (580–

750 nm) and RR. This removed most of the neighboring U(VI) signal from inhibiting the Sm^{3+} predictions. Similarly, the HNO_3 model only used the wavelength range from 400 to 480 nm, using only the Raman nitrite peak and O–H stretching band for prediction. The RR model only slightly outperformed the seven-factor PLSR model for HNO_3, but RR provided the benefit of not requiring the user to perform latent variable selection.

The temperature model was built using the second derivative of the wavelength range 415–500 nm. This range included the Raman O– H stretching band and the highest-energy U(VI) peak. This stacked regression model comprised an RR and PLSR serving as the submodels, and their predictions were analyzed by a 1,000-tree RF regression model. The resultant model had a strong predictive capability. The calibration and CV statistics for these three models are shown in Table 3 as part of the overall ensemble. One interesting observation during CV was that RMSECV was much larger than the RMSEC. This was due to Sample 16, an edge sample with U(VI) and HNO_3 levels at their maximum, being predicted poorly when left out of the model (Figure 5), indicating its importance to the model. Sample 16 was located at the edge of the experimental design space, which explains why it would not be well predicted when left out of the calibration set.

The U(VI) signal was the most influenced by changes in temperature and system acidity. The models developed to characterize temperature and acidity were used to correct the U(VI) signal. Using six D-optimal model point calibration samples (Table 2), the 510.3 nm peak intensity was plotted as a function of temperature. The peak intensity was then normalized to the maximum 510.3 nm peak of that sample (i.e., at 20°C), and the temperature was normalized to the model temperature range. The empirically derived plot (Figure 3) reveals a similar exponential relationship between the intensity and temperature. The coefficients of this relationship, shown as A and b in Figure 3, change only with acidity, not with U(VI) concentration. This relationship

allowed an additional preprocessing step for the U(VI) model where the trimmed spectra (450–550 nm) were transformed to their respective I_{max} values by scaling them using the relationship shown in Figure 3, taking into consideration their temperature and acidity.

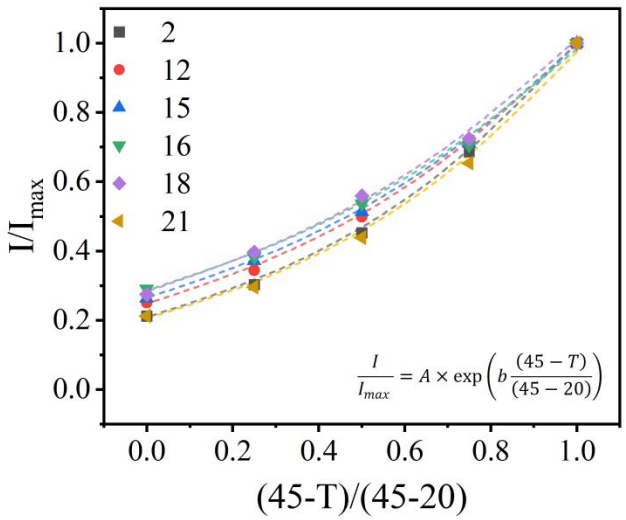

Figure 3. Temperature dependence of the 510.3 nm U(VI) fluorescence peak for six required model points. The coefficients of the exponential fit were determined to be functions of acidity: $A = -0.0033 \times C_{HNO3}^2 + 0.0334 \times C_{HNO3} + 0.2038$ and $b = 0.0152 \times C_{HNO3}^2 - 0.1416 \times C_{HNO3} + 1.5773$.

This step was essential to build an optimal U(VI) model, which also used a stacked regression. The submodels in the stacked ensemble included XGB, PLSR, and RR. Each submodel regressed the temperature and acid-adjusted preprocessed spectra to provide an individual U(VI) prediction. The submodels' estimates were then regressed themselves by the final model, in this case RR, which provided the final U(VI) concentration estimates. The final model is typically referred to as a metamodel, as it regresses the submodel predictions, rather than the spectra, to provide a final prediction. The calibration, CV, and validation statistics shown in Table 3 were calculated using

the known temperatures and acidities for the adjustment. The final ensemble model included combining all four models into one flow sheet where the temperature, HNO_3, and Sm^{3+} concentrations would be predicted first, and then these predictions were used for scaling the spectra before regression by the U(VI) model (Figure 4). The predictions of the final ensemble model are visualized in Figure 5.

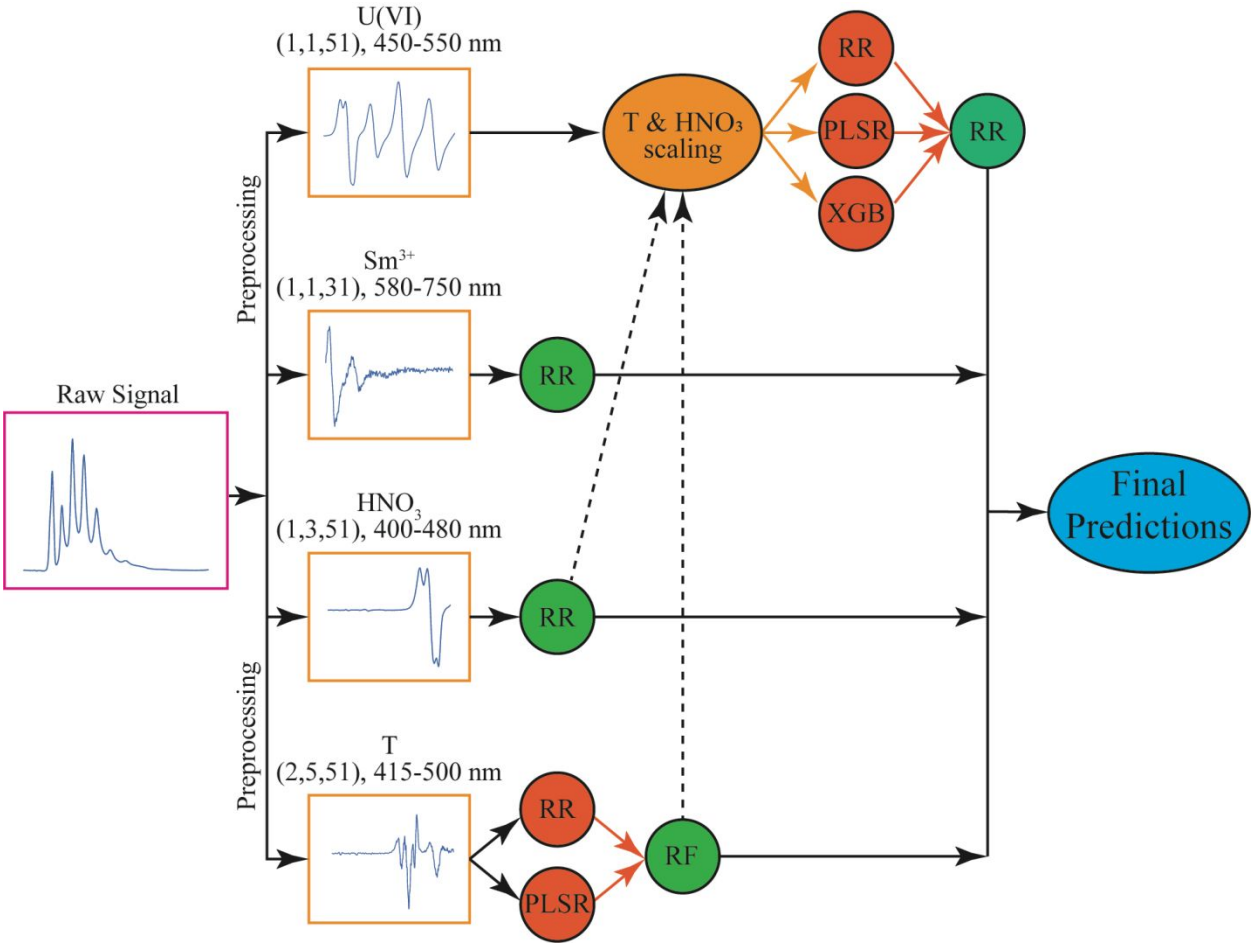

Figure 4. Overall ensemble scheme detailing how each model is connected.

The validation statistics in Table 3 show that the overall ensemble provides a strong predictive capability for all four factors, overcoming the issues of multicollinearity. The overall ensemble model's RMSEP values were significantly reduced compared to the trimmed PLSR models for all

species: 95, 22, 10, and 164% difference for U(VI), Sm^{3+}, HNO_3, and temperature, respectively. Again, a Tukey-Kramer test at a 95% confidence level was performed and for all four factors the overall ensemble model showed a statistically significant difference compared to the global and trimmed PLSR models (Figure S8).[29,30]

The RMSEP% for Sm^{3+}, HNO_3, and temperature all fall below the goal level of 5%. While the U(VI) RMSEP% misses this objective, the ensemble model vastly improves the prediction accuracy compared to the initial PLSR models. Although not shown in Table 3, the trimmed PLSR and ensemble models for U(VI) were reconstructed with and without scaling to investigate the impact of this step. The results showed that scaling improved the trimmed PLSR U(VI) RMSEP by 60%; however, the final ensemble model still offered a 10% relative reduction in RMSEP over the scaled and trimmed PLSR model.

Another note is that these models were developed with 10 model points and 5 randomly selected LOF points; given the complexity of the system more LOF points may improve the validation statistics. Similarly, an alternate set of LOF points may be better suited for the calibration set if they capture important features within the design space.

Table 3. PLSR and overall ensemble models' calibration and validation statistics for each analyte derived from multiple preprocessing strategies.

Model	Global PLSR	Trimmed PLSRs	Overall ensemble
Preprocessing			
U(VI)	(1,1,31), 410–750 nm, 8 LVs	(1,1,31), 450–550 nm, 7 LVs	(1,1,51), 450–550 nm, SR1
Sm^{3+}		(1,1,31), 580–750 nm, 6 LVs	(1,1,31), 580–750 nm, RR
HNO_3		(1,1,31), 410–480 nm, 7 LVs	(1,3,51), 410–480 nm, RR
°C		(1,1,31), 415–500 nm, 8 LVs	(2,5,51), 415–500 nm, SR2
Calibration/CV statistics			
R^2 (U(VI))	0.9672	0.9718	0.998
RMSEC	7.0	6.5	1.7

RMSECV	14	10	2.8
R^2 (Sm^{3+})	0.9988	0.9995	0.999
RMSEC	2.7	1.61	0.37
RMSECV	3.4	2.23	1.87
R^2 (HNO_3)	0.9976	0.9996	0.999
RMSEC	0.075	0.031	0.007
RMSECV	0.13	0.043	0.036
R^2 (°C)	0.9318	0.9939	0.999
RMSEC	2.3	0.69	0.003
RMSECV	3.2	1.1	0.74
Validation statistics			
RMSEP (U(VI))	9.4	8.4	3.0
RMSEP%	19%	17%	6.2%
RMSEP (Sm^{3+})	3.0	2.5	2.0
RMSEP%	3.0%	2.5%	2.0%
RMSEP (HNO_3)	0.14	0.073	0.066
RMSEP%	7.2%	3.7%	3.4%
RMSEP (°C)	2.6	1.4	0.14
RMSEP%	21%	11%	1.1%

Note: R^2 of the calibration, RMSE of the calibration (C), cross validation (CV), and prediction (P). Preprocessing information includes derivative information (order, polynomial, smoothing points); wavelength regions regressed by the model; and number of latent variables (LVs) or regression type. The first PLSR model employs the PLS-2 implementation and the trimmed PLS models utilize the PLS-1 implementation of PLSR. SR1 refers to scaling (Figure 3) and stacked regression (base models = PLSR, RR, XGB; final model = RR). SR2 refers to stacked regression (base models = PLSR, RR; final model = RF).

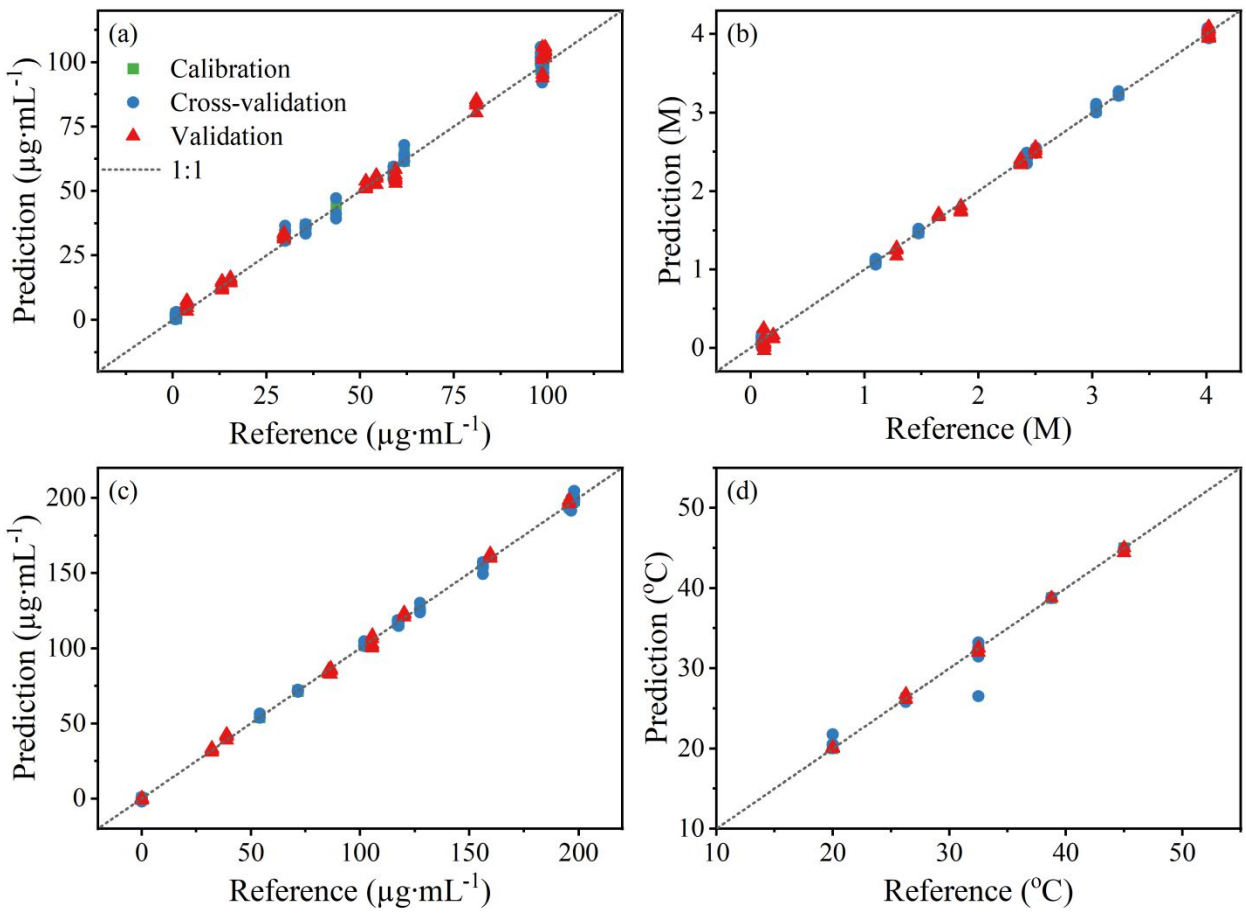

Figure 5. Predicted (a) U(VI), (b) HNO_3, (c) Sm^{3+} concentrations, and (d) temperature as determined by the overall ensemble model compared to normalized reference (known) values.

3.6 Optimizing LOF points

To find the minimal number of samples in the training set and determine which LOF points improved the calibration, each combination of 1–5 LOF points was tested in the training set, and the remaining LOF points were treated as the validation set. This power set resulted in 4,944 combinations that were evaluated in two ways including which LOF combinations best improved the overall RMSEP and which LOF combinations best improved the U(VI) RMSEP. The top overall combination and U(VI) combinations for each level of LOF inclusion are detailed in Table

S3 in the SI. The top U(VI) model used only four LOF points (1, 14, 19, 20) and resulted in RMSEP% values of 5.2%, 1.9%, 3.0%, and 2.3% for U(VI), Sm^{3+}, HNO_3, and temperature, respectively. This combination improved U(VI) prediction significantly with the trade-off of losing some temperature accuracy; however, given that U(VI) is the primary analyte of interest, this combination of model and LOF points would be best used to recalibrate a model in a new environment such as a hot cell.

Another benefit of considering all the top combinations was that the most important LOF points in the design space could be inferred. For overall prediction accuracy, Samples 1, 19, 20, 23, and 24 were found to be the most beneficial. For U(VI) prediction accuracy, Samples 1, 14, 19, 20, and 24 were found to be the most beneficial. Interestingly, samples 1, 19, 20, and 24 are represented in both lists; Samples 1, 20, and 24 are all plane points with the maximum U(VI) concentration, and sample 19 is an interior point nearly at the center of the overall design space.

4. CONCLUSIONS

The stacked regression approach built upon several multivariate analytical methods can account for nonlinear temperature fluctuations in uranyl fluorescence spectra without measuring luminescence lifetimes or using a separate temperature probe. Raman spectral fingerprints were combined with LIFS to improve acid and temperature predictions. The real-time feedback afforded by this novel approach makes it possible to study separation system dynamics prior to equilibrium. In addition, by simultaneously quantifying Sm^{3+}, a common fission product, it may be possible to characterize fuel burnup using the LIFS technique.

The work presented here was essential to demonstrate key concepts related to minimizing the number of samples in the training set and building stacked regression models to account for

temperature fluctuations, prior to characterizing a more complicated system. To demonstrate greater applicability in real-world processing solutions, future work will include additional factors in the training set to account for self-absorption, quenching, and additional peak overlap effects.[11,43] These species may include fission products (e.g., Zr, Mo, Ru); lanthanides (e.g., Eu^{3+}, Ce^{3+}, Nd^{3+}); corrosion products (e.g., Ni^{2+}, Cr^{3+}, Fe^{3+}) and phosphorous-containing breakdown products. This method will also be applied to a more challenging validation set that includes a flow loop demonstration.

ASSOCIATED CONTENT

Supporting Information. The files noted below are available free of charge.

Extended statistical methods section, lifetime decay curves, and additional experimental results (MS Word)

AUTHOR INFORMATION

Corresponding Author

*Telephone: +1 (865) 574-1167; email address: sadergaskilr@ornl.gov

Author Contributions

The manuscript was written using contributions of all authors. All authors 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 (NNSA) 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 (ORNL).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors wish to thank Alexander Braatz for helpful discussions on nuclear fuel processing. This work was supported by the NNSA and performed at ORNL.

REFERENCES

1. J-Y. C. Colle, D. Manara, T. Geisler and R. J. M. Konings, Advances in the application of Raman spectroscopy in the nuclear field. *Spectrosc. Eur.* 2020, **32** (6), 9–13.
2. D. Kirsanov, A. Rudnitskaya, A. Legin and V. Babain, UV-VIS spectroscopy with chemometric data treatment: An option for on-line control in nuclear industry. *J. Radioanal. Nucl. Chem.* 2017, **312**, 461–470.
3. L. R. Sadergaski, D. W. DePaoli and 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.
4. G. J. Lumetta, J. R. Allred, S. A. Bryan, G. B. Hall, T. G. Levitskaia, A. M. Lines and S. I. Sinkov, 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.

5. B. H. Deniau, P. Decambox, P. Mauchien and C. Moulin, 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.

6. R. A. Peterson, E. C. Buck, J. Chun, R. C. Daniel, D. L. Herting, E. S. Ilton, G. J. Lumetta and S. B. Clark, Review of the Scientific Understanding of Radioactive Waste at the U.S. DOE Hanford Site. *Environ. Sci. Technol.* 2018, **52**, 381–396.

7. H. Fujimori, T. Matsui, and K. Suzuki, K. Fluorometry of UO_2^{2+} Ion in Nitric Acid Solutions. *J Nucl. Sci. Technol.* 1986, **23**, 1069–1074.

8. C. Moulin, P. Decambox, L. Couston and D. Pouyat, Time-Resolved Laser-Induced Fluorescence of UO_2^{2+} in Nitric Acid Solutions. Comparison between Nitrogen and Tripled Nd-YAG laser. *J Nucl. Sci. Technol.* 1994, **31**, 691–699.

9. T. Matusi, H. Fujimori and K. Suzuki, Effects of Coexisting Ions upon UO_2^{2+} Fluorescence in Fuel Reprocessing Solutions. *J Nucl. Sci. Technol.* 1988, **25**, 868–874.

10. C. Moulin, P. Decambox and P. Mauchien, Direct Uranium(VI) and Nitrate Determinations in Nuclear Reprocessing by Time-Resolved laser-Induced Fluorescence. *Anal. Chem.* 1996, **68**, 3204–3209.

11. T. Matsui, T. Kitamori, H. Fujimori, K. Suzuki and M. Sakagami, In-Line Uranium Monitoring in Reprocessing Waste Solution by Time-Resolved Laser-Induced Fluorometry. *J Nucl. Sci. Technol.* 1992, **29**, 664–670.

12. N. S. Smith, G. S. Cerefice and K. R. Czerwinski, Fluorescence and absorbance spectroscopy of the uranyl ion in nitric acid for process monitoring applications. *J. Radioanal. Nucl. Chem.* 2013, **295**, 1553–1560.

13. L. M. Colletti, R. Copping, K. Garduno, E. J. W. Lujan, A. K. Mauser, A. Mechler-Hickson, I. May, S. D. Reilly, D. Rios, J. Rowley and A. B. Schroeder. The application of visible absorption spectroscopy to the analysis of uranium in aqueous solutions. *Talanta*, 2017, **175**, 390–405.

14. T. Saito, N. Aoyagi and T. Kimura. Time-resolved laser-induced fluorescence spectroscopy combined with parallel factor analysis: a robust technique for UO_2^{2+}. *J. Radioanal. Nucl. Chem.* 2015, **303**, 1129–1132.

15. I. N. Izosimov, N. G. Firsin, N. G. Gorshkov and S. N. Nekhoroshkov. Detection of lanthanides and actinides in solutions based on laser-induced luminescence and chemiluminescence. *Hyperfine Interact.* 2014, **227**, 271–281.

16. T. Kimura, R. Nagaishi, T. Ozaki, M. Arisaka and Z. Yoshida, Uranium(VI) Speciation at Elevated Temperatures and Pressures by Time-resolved Laser-induced Fluorescence Spectroscopy. *J Nucl. Sci. Technol.* 2002, **3**, 233–239.

17. C. Talbot-Eeckelaers, S. J. A. Pope, A. J. Hynes, R. Copping, C. J. Jones, R. J. Taylor, S. Faulkner, D. Sykes, F. R. Livens and I. May. Luminescence from Neptunyl(VI) Species in Solution. *J. Am. Chem. Soc.* 2007, **129**, 2442–2443.

18. R. G. Denning, Electronic Structure and Bonding in Actinyl Ions and their Analogs. *J. Phys. Chem. A* 2007, **111**, 4125–4143.

19. S. Ganesh, F. Khan, M. K. Ahmed, P. Velavendan, N. K. Pandey, K. Mudali and S. K. Pandey, Determination of ultra traces amount of uranium in raffinates of Purex process by laser fluorimetry. *J. Radioanal. Nucl. Chem.* 2012, **292**, 331–334.

20. D. Y. Chung, E. H. Lee, J. H. Yoo and T. Kimura, Laser-Induced Fluorescence Studies on Sm(III) and Cm(III) Complexes in the HNO_3/DHDECMP Extraction System. *J. Nucl. Sci.* 2002, **3**, 332–335.

21. W. T. Carnall, P. R. Fields and K. Rajnak, Electronic Energy Levels in 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.

22. K. B. Hong, K. W. Jung and K.-H. Jung, Application of Laser-Induced Fluorescence for Determination of Trace Uranium, Europium and Samarium. *Talanta* 1989, **36**, 1095–1099.

23. S. Goderski, M. Runowski, P. Wozny, V. Lavin and S. Lis, Lanthanide Upconverted Luminescence for Simultaneous Contactless Optical Thermometry and Manometry – Sensing under Extreme Conditions of Pressure and Temperature. *CS Appl. Mater. Interfaces* 2020, **12**, 40475–40485.

24. U. K. Maity, P. Manoravi, M. Joseph and N. Sivaraman, Laser-induced breakdown spectroscopy for simultaneous determination of lighter lanthanides in actinide matrix in aqueous medium. *Spectrochim. Acta B: At. Spectrosc.* 2022, **190**, 106393.

25. A. M. Lines, G. L. Nelson, J. M. Casella, J. M. Bello, S. E. Clark and S. A. Bryan, 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.

26. A. J. Casella, T. G. Levitskaia, J. M. Peterson and S. A. Bryan, Water O–H Stretching Raman Signature for Strong Acid Monitoring via Multivariate Analysis. *Anal. Chem.* 2013, **85** (8), 4120–4128.

27. P. Tse, J. Shafer, S. A. Bryan, G. L. Nelson and A. M. Lines. Measuring Nd(III) Solution Concentration in the Prescence of Interfering Er(III) and Cu(II) Ions: A Partial Least Squares Analysis of Ultraviolet-Visible Spectra. *Appl. Spectrosc.* 2022, **76**, 173–183.

28. P. Tse, J. Shafer, S. A. Bryan and A. M. Lines, Quantification of Raman-Interfering Polyoxoanions for Process Analysis" Comparison of Different Chemometric Models and a Demonstration on Real Hanford Waste. *Environ. Sci. Technol*. 2021, **55**, 12943–12950.

29. L. R. Sadergaski, G. K. Toney, L. H. Delmau and K. G. Myhre. 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.

30. L. R. Sadergaski, T. J. Hagar and H. B. Andrews. Design of Experiments, Chemometrics, and Raman Spectroscopy for the Quantification of Hydroxylammonium, Nitrate, and Nitric Acid. *ACS Omega*, 2022, **7**, 7287–7296.

31. V. Czitrom. One-Factor-at-a-Time Versus Designed Experiments. *Am. Stat.* 1999, **53** (2), 126–131.

32. B. Smucker, M. Krzywinski and N. Altman, Optimal experimental design. *Nat. Methods* 2018, **15**, 559–560.

33. A. Zahran, C. M. Anderson-Cook and R. H. Myers, Fraction of Design Space to Assess Prediction Capability of Response Surface Designs. *J. Qual. Tech.* 2003, **35** (4), 377–386.

34. P. Shan, Y. Zhao, Q. Wang, X. Sha, X. Lv, S. Peng and Y. Ying. Stacked ensemble extreme learning machine coupled with Partial Least Squares-based weighting strategy for nonlinear multivariate calibration. *Spectrochim. Acta A: Mol. and Biomol. Spectrosc.*, 2019, **215**, 97–111.

35. R. Lascola, P. E. O'Rourke and E. A. Kyser. A Piecewise Local Partial Least Squares (PLS) Method for the Quantitative Analysis of Plutonium Nitrate Solutions. *Appl. Spectrosc.* 2017, **7** (12), 2579–2594.

36. M. Mancini, V.-M. Taavitsainen and G. Toscano. Comparative study between Partial Least Squares and Rational function Ridge Regression models for the prediction of moisture content of woodchip samples using a handheld spectrophotometer. *J. Chemometr.* 2021, **35**, e3337. DOI: 10.1002/cem.3337.

37. H. B. Andrews and K. G. Myhre. Quantification of Lanthanides in a Molten Salt Reactor Surrogate Off-Gas Stream Using Laser-Induced Breakdown Spectroscopy. *Appl. Spectrosc.* 2022, doi:10.1177/00037028211070323.

38. L. Wei, Z. Yuan, Y. Zhong, L. Yang, X. Hu and Y. Zhang. An improved gradient boosting regression tree estimation model for soil heavy metal (arsenic) pollution monitoring using hyperspectral remote sensing. *Appl. Sci.* 2019, **9** (9), 1943.

39. L. Wei, Z. Yuan, M. Yu, C. Huang and L. Cao. Estimation of Arsenic Content in Soil Based on Laboratory and Field Reflectance Spectroscopy. *Sensors* 2019, **19** (18), 3904, doi:10.3390/s19183904.

40. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 2011, **12**, 2825–2830.

41. T. Chen and C. Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd SIGKDD Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, August 13–17, 2016; ACM, New York, 2016, 785–794.

42. D. Wang and X. Yue. The weighted multiple meta-models stacking method for regression problem. In *Proceedings of the 2019 Chinese Control Conference (CCC)*, 2019, 7511-7516.

43. 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.