

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

5 *Luke R. Sadergaski,\* Hunter B. Andrews*

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

8 [\\*sadergaskilr@ornl.gov](mailto:*sadergaskilr@ornl.gov)

9 ABSTRACT: Laser-induced fluorescence spectroscopy (LIFS), Raman spectroscopy, and a  
10 stacked regression ensemble was developed for near real-time quantification of uranium (VI) (1–  
11 100  $\mu\text{g}\cdot\text{mL}^{-1}$ ), samarium (0–200  $\mu\text{g}\cdot\text{mL}^{-1}$ ) and nitric acid (0.1–4 M) with varying temperature  
12 (20°C– 45°C). LIFS applications range from fundamental lab-scale studies to real-time process  
13 monitoring at industrial levels, such as nuclear reprocessing applications, provided the phenomena  
14 affecting the fluorescence spectrum are accounted for (e.g., absorption, quenching, complexation).  
15 Multiple chemometric models were examined and compared to a more traditional multivariate  
16 regression approach called partial least squares (PLS). Results obtained on synthetic samples  
17 selected using D-optimal experimental design indicated that a stacked regression method, which  
18 included ridge regression, random forest, PLS, and an eXtreme gradient boost algorithm,

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>).

19 successfully measured uranium (VI) concentrations directly in nitric acid without measuring  
20 luminescence lifetimes or standard addition. The top model resulted in percent root-mean-square  
21 error of prediction values of 5.2, 1.9, 3.0, and 2.3% for U(VI),  $\text{Sm}^{3+}$ ,  $\text{HNO}_3$ , and temperature,  
22 respectively. The approach may be useful for quantifying fluorescent fission products (e.g.,  $\text{Sm}^{3+}$ )  
23 to provide information on burnup of irradiated nuclear fuel. This novel framework reinforces the  
24 applicability of LIFS for real-time applications in nuclear fuel cycle applications.

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

## 27 **1. INTRODUCTION**

28 Optical spectroscopy is a powerful option for in-line process control in many industrial  
29 applications and nuclear fuel cycle reprocessing separation schemes like PUREX (**P**lутonium  
30 **U**ranium **R**eduction **E**Xtraction).<sup>1-5</sup> PUREX is a liquid-liquid extraction process that purifies  
31 uranium and plutonium by first extracting uranium (U) and plutonium (Pu) from aqueous solutions  
32 (3-4 M nitric acid) with a hydrocarbon phase containing tributyl phosphate. This extraction step  
33 is followed by a second U and Pu partitioning stage and a third U stripping stage in dilute nitric  
34 acid.<sup>6</sup> All PUREX-related tasks must be performed remotely in heavily shielded hot cell  
35 enclosures. Additional complications arise from radiation-induced decomposition and criticality  
36 risks, which makes this operation one of the most complicated chemical processes ever attempted.  
37 Processing must be supported by numerous analytical measurements. Traditional techniques,  
38 including radiochemistry (e.g., alpha and gamma spectroscopy) or inductively coupled plasma  
39 mass spectrometry, generally require retrieving samples which require significant dilutions  
40 (1,000-10,000-fold) and transfer out of the hot cell for analysis.<sup>2</sup> Hot cell measurements using

41 optical spectroscopy are much faster and can be acquired in situ, in real-time using fiber-optic  
42 cables.

43 One form of optical spectroscopy, time-resolved laser-induced fluorescence spectroscopy  
44 (TRLIFS), was evaluated extensively for monitoring PUREX streams (e.g., raffinate) several  
45 decades ago,<sup>7-11</sup> but little work has been done in the last decade. More recently, Raman  
46 spectroscopy and UV-Vis spectrophotometry have been evaluated for monitoring applications,  
47 although these options are orders of magnitude less sensitive to uranium (VI) concentration than  
48 TRLIFS.<sup>1,2,12-14</sup> TRLIFS is highly sensitive and selective to many actinide (An) species. It is often  
49 used for ultratrace analysis and complexation studies with both spectral and temporal features.<sup>15,16</sup>  
50 TRLIFS can detect several actinides including  $\text{UO}_2^{2+}$ ,  $\text{Am}^{3+}$ ,  $\text{Cm}^{3+}$ ,  $\text{Cf}^{3+}$ ,  $\text{Bk}^{3+}$ , and  $\text{Es}^{3+}$ . Only  
51  $\text{UO}_2^{2+}$  and  $\text{Cm}^{3+}$  have a suitably large energy gap ( $>10^4 \text{ cm}^{-1}$ ) for the luminescence level, while  
52 the others have a smaller gap, resulting in luminescence quantum yields highly sensitive to specific  
53 compounds and solution types.<sup>15</sup> It is challenging to measure the aqueous luminescence properties  
54 of other actinides like Np and Pu due to strong nonradiative quenching by the solvent (e.g., O-H  
55 band). This effect can be lessened using other solvent types such as heavy water ( $\text{D}_2\text{O}$ ).<sup>17</sup> Despite  
56 the incredible sensitivity of the time-resolved technique, the analysis time is often 10–15 minutes,  
57 which is too slow for situations where real-time feedback is required.<sup>10</sup>

58 Laser-induced fluorescence spectroscopy (LIFS) is useful for the qualitative and quantitative  
59 analysis of molecular species such as the free uranyl ( $\text{UO}_2^{2+}$ ) and uranyl nitrate complexes  
60 ( $\text{UO}_2(\text{NO}_3)_n^{(2-n)+}$ ,  $n = 1$  or  $2$ ) in aqueous solutions or U(VI) solid-state compounds.<sup>18</sup> Fluorescence  
61 spectra can be measured using a charge-coupled device to increase the timeliness of data  
62 collection, as opposed to single-channel photomultiplier tubes (PMTs) often used for time-  
63 resolved data. However, applying laser fluorometry is difficult in complex systems with

64 overlapping bands, noise, matrix effects, chemical interactions, and baseline offsets.<sup>5,11,15,19</sup>  
65 Although it is possible to reroute the flow during processing to maintain a constant temperature  
66 and generate U(VI) complexes to improve quantum yield (i.e., sensitivity),<sup>19</sup> it is preferable to  
67 perform the analysis in-line (i.e., directly in nitric acid) and without complicating hot cell  
68 operations. This necessitates directly accounting for the complicating effects of a nitric acid  
69 medium and other real-world process conditions such as temperature fluctuations.

70 Dynamic temperatures drastically change the lifetime and intensity of U(VI) fluorescence,  
71 resulting in two highly correlated independent variables (i.e., multicollinearity), possibly the most  
72 challenging quenching phenomena.<sup>5,10,11,15</sup> It is critical to account for temperature because high-  
73 level waste solutions vary from 30°C–40°C and PUREX operations often take place at  
74 temperatures ranging from 20–40°C.<sup>5</sup> Additional complicating factors arise because many fuel  
75 cycle processing streams contain a variety of fission and corrosion products with absorption or  
76 fluorescence properties that can complicate LIFS spectra.<sup>9</sup> Lanthanide (Ln) fission products  
77 include Ce<sup>3+</sup>, Pr<sup>3+</sup>, Nd<sup>3+</sup>, Sm<sup>3+</sup>, Eu<sup>3+</sup>, and Gd<sup>3+</sup>. Ln elements such as Eu<sup>3+</sup>, Tb<sup>3+</sup>, Gd<sup>3+</sup>, Dy<sup>3+</sup>, Sm<sup>3+</sup>,  
78 Ce<sup>3+</sup>, and Tm<sup>3+</sup> can be characterized by TRLIFS.<sup>14,20–23</sup> Two lanthanides, Eu<sup>3+</sup> and Sm<sup>3+</sup>, have  
79 visible luminescence properties and range from ~10 to 200 µg·mL<sup>-1</sup> in PUREX raffinate streams.<sup>9</sup>  
80 The fission yield of Sm<sup>3+</sup> is generally greater than Eu<sup>3+</sup> and is less characterized in the literature.  
81 LIFS could provide real-time feedback regarding nuclear fuel burnup and provide operational  
82 benefits compared to other optical techniques.<sup>24</sup> Multivariate regression analysis has not been used  
83 to describe U(VI) fluorescence spectra under relevant conditions. It is important to reexamine this  
84 system with modern data analytics to achieve greater success with the LIFS technique.

85 This work explores the use of multivariate chemometrics, or machine learning, to account for  
86 these multifaceted effects. One of the most traditional supervised techniques is called partial least

87 squares regression (PLSR).<sup>25–28</sup> This factor analysis method iteratively relates two data matrices,  
88 the independent  $X$  (i.e., spectra) and dependent  $Y$  (i.e., concentrations), using combinations of  
89 latent variables (LV). PLSR models are built using a training set that covers the expected  
90 conditions and a validation set that tests the model's ability to predict samples not included in the  
91 training set. These samples can be selected using optimal experimental designs that are the most  
92 flexible and effective option when a small number of experimental runs is desired.<sup>29–33</sup>

93 Although PLSR has been used with great success to model systems with overlapping spectral  
94 features in numerous systems,<sup>27</sup> it does not always account for systems with a high degree of  
95 multicollinearity.<sup>34,35</sup> Multicollinearity must be accounted for in a LIFS system for monitoring  
96 U(VI) concentration because two independent variables (i.e., U(VI) concentration and  
97 temperature) are highly correlated in the regression model. To address this, ensemble learning  
98 combines multiple models to make a prediction in classification or regression problems to improve  
99 robustness and accuracy compared to single models. This process can be divided into three phases:  
100 (1) develop a set of candidate models (i.e., generation phase); (2) select a subset of the models  
101 (i.e., pruning phase); and (3) combine models to generate predictions (i.e., integration phase). In  
102 addition to PLSR, several advanced models are considered in this article, including ridge  
103 regression (RR), random forest (RF), and an eXtreme gradient boost (XGB) algorithm.<sup>36–42</sup> These  
104 are evaluated individually and compared to a stacked regression approach, a form of ensemble  
105 learning. The primary goal of this work is to demonstrate how ensemble learning methods can be  
106 used to handle the dynamic, overlapping, covarying, and nonlinear spectral response to provide a  
107 new U(VI) analysis method independent of time-resolved fluorescence spectra.

108 Here, we demonstrate how to combine ensemble learning, LIFS, and Stokes Raman scattering  
109 to account for varying temperature, conditions that historically limit the timeliness of U(VI)

110 monitoring applications. This work also includes the simultaneous quantification of a lanthanide  
111 fission product (i.e.,  $\text{Sm}^{3+}$ ). Calibration and validation fluorescence spectral data sets were selected  
112 by determinant(D)-optimal designs to minimize the samples required in the training set, which  
113 spanned U(VI) ( $1\text{--}100\ \mu\text{g}\cdot\text{mL}^{-1}$ ),  $\text{Sm}^{3+}$  ( $0\text{--}200\ \mu\text{g}\cdot\text{mL}^{-1}$ ),  $\text{HNO}_3$  ( $0.1\text{--}4\ \text{M}$ ) concentrations and  
114 temperatures ( $20^\circ\text{C}\text{--}45^\circ\text{C}$ ), conditions highly applicable to the nuclear fuel cycle. Specific points  
115 of scientific advancement covered in this work include: (1) the ensemble method allows for  
116 quantitative U(VI) predictions and varying temperature without recording lifetimes,  
117 (2) simultaneous quantification of a fluorescent lanthanide (i.e.,  $\text{Sm}^{3+}$ ) in the milligram-per-liter  
118 range, (3) stacked regression that combines LIFS and Raman spectroscopy, and (4) the augmented  
119 D-optimal design allows training/validation set samples to be effectively selected without user bias  
120 while minimizing the number of samples. Many spectroscopy-enabled online monitoring studies  
121 tend to neglect the temperature variable and include too many user-selected samples to be easily  
122 implemented in restrictive hot cell environments.<sup>26</sup> This work clearly articulates the challenges  
123 that arise due to fluctuating temperature and best practices to correct for such phenomena; this is  
124 applicable to many optical spectroscopy online monitoring applications extending within and  
125 beyond nuclear fuel cycle applications.

## 126 **2. METHODS**

127 All chemicals were commercially obtained (ACS grade) and used as received unless otherwise  
128 stated. Concentrated  $\text{HNO}_3$  (70%) was purchased from Sigma-Aldrich. Certified  $10,000\ \mu\text{g}\cdot\text{mL}^{-1}$   
129 uranium ( $^{238}\text{U}$ , depleted) and samarium inductively coupled plasma optical emission spectroscopy  
130 standard solutions in 5% nitric acid were purchased from SPEX CertiPrep and Inorganic Ventures,  
131 respectively. Samples were prepared using deionized water with Milli-Q purity ( $18.2\ \text{M}\Omega\cdot\text{cm}$  at  
132  $25^\circ\text{C}$ ).

133 **2.1 Sample preparation**

134 Calibration and validation samples contained uranium (1–100  $\mu\text{g}\cdot\text{mL}^{-1}$ ), samarium (0–200  
135  $\mu\text{g}\cdot\text{mL}^{-1}$ ), and  $\text{HNO}_3$  (0.1–4 M) and chosen to cover the anticipated solution conditions. Samples  
136 were prepared gravimetrically using volumetric pipettes. Each sample was prepared in individual  
137 2 mL plastic microcentrifuge tubes (VWR Scientific, 525-1160) prior to spectroscopic analysis. A  
138 fluorescence 3-in-1 flow cuvette, purchased from Hellma (584.4-Q-1) with  $5 \times 2.5$  mm optical  
139 path lengths, was used for each measurement to ensure consistent optical quality. The cuvette was  
140 periodically rinsed with 2% nitric acid and stored with ultrapure water on lint-free Kimwipes. The  
141 cuvette Z-height of 8.5 mm was necessary to accommodate Quantum Northwest's qpod 2e  
142 temperature-controlled sample compartment holder purchased from Avantes (CUV-UV/Vis-TC).  
143 Two collimating lenses (CUV-TC-QCL-UV) were placed at  $90^\circ$  in the sample compartment.  
144 Fluorescence measurements were performed at varying temperatures (i.e.,  $20^\circ\text{C}$ – $50^\circ\text{C}$ ) with an  
145 accuracy of  $\pm 0.05^\circ\text{C}$ . Sample solutions were thermally equilibrated for at least 2 min prior to  
146 recording each spectrum to eliminate spectral variations due to temperature fluctuation. A syringe  
147 was used to inject the rinse and sample solutions.

148 **2.2 Fluorescence and Raman spectroscopy**

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

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

$$162 \quad D(t) = \sum a_i \exp\left\{\frac{-t}{\tau_i}\right\}, \quad (1)$$

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

### 166 **2.3 Design of experiments**

167 Experimental designs were built using Design-Expert (v.11.0.5.0) by Stat-Ease Inc., within the  
168 Unscrambler software package by Camo Analytics. Optimal designs incorporate mixture and  
169 process variables, contain different high and low components, and feature constraints with factor  
170 limits. D-optimal samples were chosen by iteratively minimizing the determinant of the variance–  
171 covariance matrix  $X^T X$  using a quadratic process order.<sup>25</sup> A combination of point and coordinate  
172 exchange search options was used to select points. Fraction of design space (FDS) was used to  
173 evaluate the model and calculated by mean error type:  $\delta = 2$ ,  $\sigma = 1$ , and  $\alpha = 0.05$ .<sup>33</sup>

174 A minimum of 10 model points was required for three numeric factors (i.e., U(VI), Sm, and  
175 HNO<sub>3</sub> concentrations). The design was augmented with 15 additional lack-of-fit (LOF) points,  
176 which were included either as calibration or validation samples. LOF points are chosen to  
177 maximize the distance to other runs while maintaining the optimality criterion.<sup>32</sup> Temperature was

178 treated in a separate design. At least three model points were required for a single numeric factor  
179 (i.e., temperature), and the design was augmented using 2 LOF points. Combining the two designs  
180 resulted in a total of 125 samples.

## 181 **2.4 Multivariate analysis and preprocessing**

182 Several chemometric methods were evaluated including PLSR, RR, RF, an XGB algorithm, and  
183 combining multiple methods through stacked regressions. PLSR iteratively relates two data  
184 matrices, the independent  $X$  (i.e., spectra) and dependent  $Y$  (i.e., concentrations), using  
185 combinations of latent variables. The optimal number of latent variables is selected through cross  
186 validation (CV); the optimal number typically includes up to the last latent variable that marks a  
187 significant increase in the explained variance (or a significant reduction in the error of prediction).  
188 PLSR can be implemented in two forms, the first (PLS-1) models the response of only a single  
189 factor and the second (PLS-2) models the response of multiple factors simultaneously. PLS-2 is  
190 typically used to model systems with multicollinearity; however, in some cases multiple PLS-1  
191 models can have better predictive capabilities by leveraging different preprocessing strategies (i.e.,  
192 trimming, smoothing, derivatives, etc.). RR is similar to ordinary least squares regression;  
193 however, instead of simply minimizing the sum of squared error, there is an additional factor,  $\alpha$ ,  
194 that penalizes the slope of the regression.<sup>39,41</sup> This factor is selected through CV and must be  
195 greater than zero. RR is typically useful when analyzing data with high multicollinearity (i.e.,  
196 correlated independent variables). RF is a supervised form of machine learning that fits a user-  
197 defined number of independent decision trees to the data and then combines the predictions of the  
198 many trees to provide a single prediction.<sup>39,40</sup> The XGB algorithm is an open-source  
199 implementation of the gradient boosting algorithm that uses a loss function to prevent overfitting.

200 XGB also uses a multitude of decision trees like RF, but actively uses residual information to  
201 optimize the model.<sup>39</sup>

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

210 Spectral data was preprocessed prior to modeling including smoothing, derivatives, and  
211 trimming to optimize performance. Smoothing removes instrument noise, and the use of  
212 derivatives makes calibration models more resistant to baseline shifts expected in monitoring  
213 applications. The smoothing and derivatives were performed using several Savitzky–Golay (SG)  
214 filters. SG derivatives fit a polynomial to the data based on a user-defined number of left/right  
215 smoothing points and polynomial order. Higher than first and second derivatives were not included  
216 here because they suppressed the relatively broad spectral features needed for regression.  
217 Numerous derivative orders, polynomial orders, and smoothing points were tested.<sup>30</sup> Trimming  
218 the spectra reduces the dimensionality of large spectra files and can help reduce the error of  
219 prediction in multivariate modeling. The trimmed region included in the chemometric model was  
220 varied based on the behavior being modeled. All regression models and data preprocessing were  
221 completed in Python 3 using modules from the Scikit Learn and XGBoost packages.<sup>38,40</sup>

## 222 2.5 Statistical comparison

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

$$231 \text{ RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}, \quad (2)$$

232 where  $\hat{y}_i$  is the predicted concentration,  $y_i$  is the measured concentration, and  $n$  is the number of  
233 samples. RMSEP% was calculated by dividing the RMSEP by the median model values using  
234 Equation 3:

$$235 \text{ RMSEP}\% = \frac{\text{RMSEP}}{y_{med}} \times 100\%, \quad (3)$$

236 where  $y_{med}$  represents the median of each analyte concentration range. Each RMSE value is in units  
237 of analyte concentration. In general, lower RMSEC, RMSECV, RMSEP, and RMSEP% values  
238 indicate better model performance. A Tukey-Kramer significance test was used to statistically  
239 compare the RMSEP values for multiple regression models following a method outlined  
240 previously.<sup>29,30</sup> Additional details can be found in the SI.

## 241 3. RESULTS AND DISCUSSION

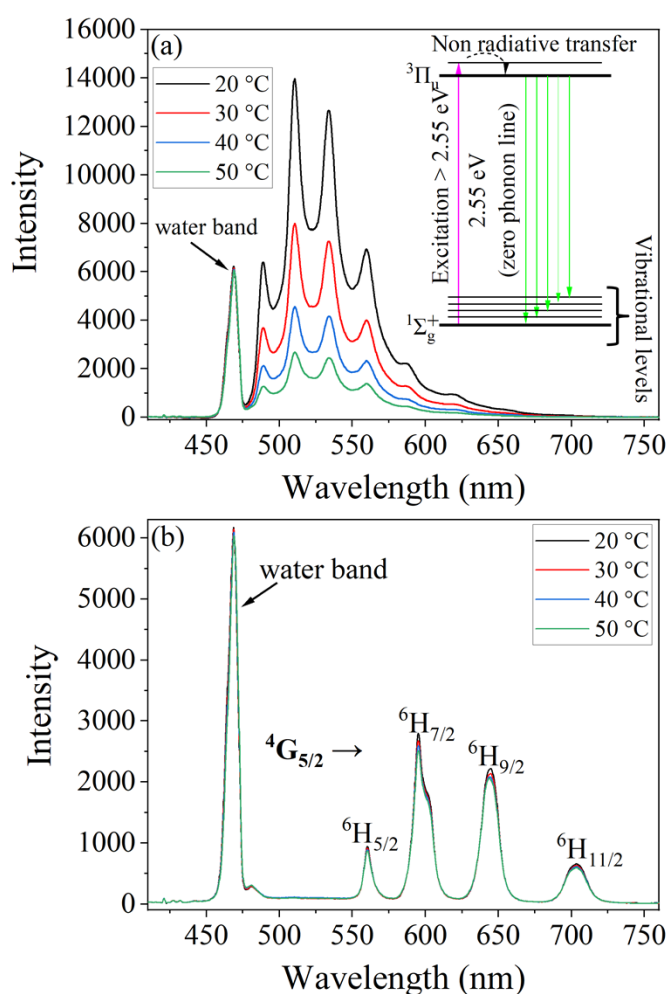
### 242 3.1 Fluorescence and Raman spectra

243 The absorption and photoluminescence spectrum of the uranyl ion ( $\text{UO}_2^{2+}$ ) is extensively  
244 characterized in the literature.<sup>13</sup> Each electron in this nearly linear moiety is paired, thus the  
245 ground-state electronic level is a singlet.<sup>18</sup> Higher-energy levels occur when one of the bonding  
246 electrons is transferred to the 5f nonbonding atomic orbitals of the uranium ion  $\sim 20,000 \text{ cm}^{-1}$  (500  
247 nm) above the ground singlet. Photoexcitation occurs when photons with a higher energy than the  
248 first excited electronic level are absorbed. The excited uranyl ion relaxes rapidly by a nonradiative  
249 process to the first excited electronic level, followed by fluorescence to the symmetric and  
250 asymmetric vibrational levels associated with the ground-state singlet. This process is described  
251 by a notional energy-level diagram in Figure 1a.

252 The  $\text{Sm}^{3+}$  emission spectrum consists of four bands near 563, 596, 643, and 720 nm,  
253 corresponding to the  $^4\text{G}_{5/2} \rightarrow ^6\text{H}_{5/2}$ ,  $^6\text{H}_{7/2}$ ,  $^6\text{H}_{9/2}$  and  $^6\text{H}_{11/2}$  transitions, respectively (Figure 1b). The  
254 electronic structures of lanthanides elements are characterized primarily by the 4f orbital.<sup>21</sup> The 4f  
255 orbitals are shielded by the filled 5s and 5p shells, indicating that the surrounding matrix has a  
256 minimal effect on 4f<sup>n</sup> energy levels and corresponding optical transitions.<sup>21</sup> As a result, 4f electrons  
257 do not play a significant role in chemical bonding, and wavelength-dependent shifts due to the host  
258 medium are minimal. Most research on lanthanide luminescence in aqueous environments has  
259 been concerned with  $\text{Eu}^{3+}$  and  $\text{Tb}^{3+}$ . The other two visibly luminescent ions,  $\text{Sm}^{3+}$  and  $\text{Dy}^{3+}$ , have  
260 received less attention because they have inferior luminescence quantum yields (i.e., more efficient  
261 nonradiative relaxation). Three out of the four  $\text{Sm}^{3+}$  emission bands significantly overlap with the  
262 uranyl fluorescence spectrum (see Figure 1).

263 With increasing temperature, the fluorescence intensity of U(VI) decreases dramatically (i.e., via  
264 dynamic quenching) because the molecular collisional rate increases (Figure 1a). The shape of the

265 fluorescence spectrum changes minimally with increasing temperature (Figure S1). This implies  
 266 that the change in temperature has a minimal effect on the nature of the  $\text{UO}_2^{2+}$  species. The  
 267 temperature-induced radiative deexcitation rate of  $\text{UO}_2^{2+}$  is likely due to a chemical process  
 268 involving hydrogen abstraction from the solvent molecules coordinating  $\text{UO}_2^{2+}$ .<sup>5</sup> On the other  
 269 hand, the fluorescence spectrum of  $\text{Sm}^{3+}$  changes slightly in wavelength position and intensity as  
 270 a function of temperature. These thermochromic shifts likely correspond to slight variations in the  
 271 coordination environment surrounding  $\text{Sm}^{3+}$ .<sup>23</sup>



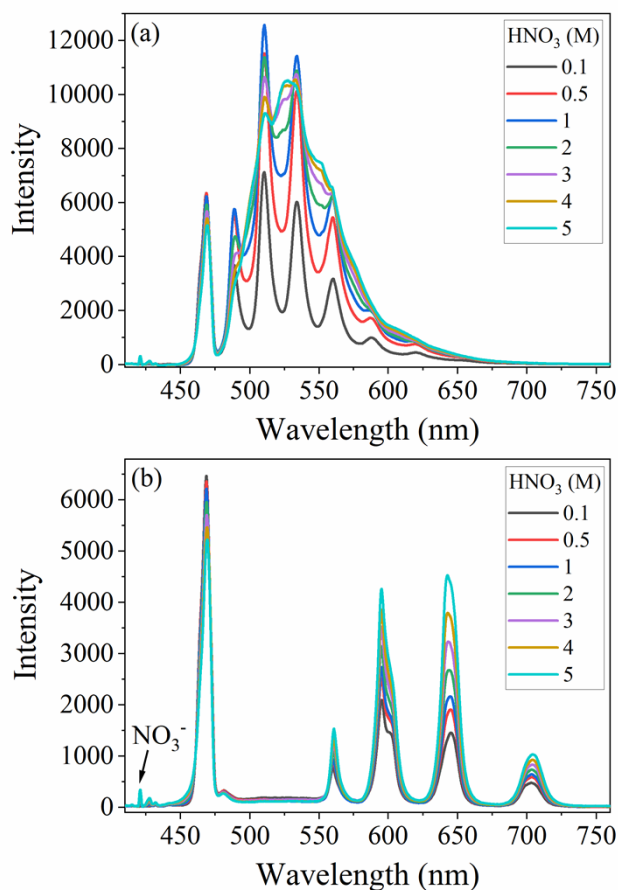
272  
 273 **Figure 1.** Laser fluorescence spectra ( $\lambda_{\text{ex}} = 405 \text{ nm}$ ) of an aqueous solution containing  $100 \mu\text{g}\cdot\text{mL}^{-1}$   
 274 uranium (a) or  $100 \mu\text{g}\cdot\text{mL}^{-1} \text{ Sm}^{3+}$  (b) in  $1 \text{ M HNO}_3$  from  $20^\circ\text{C}$ – $50^\circ\text{C}$ . A notional uranyl energy-

275 level diagram is included in the figure, and each  $\text{Sm}^{3+}$  transition is labeled. The Raman  $\text{NO}_3^-$   
276 symmetric N–O stretch and the O–H stretching band appeared at 421 nm and 455–475 nm,  
277 respectively.

278 The U(VI) spectrum changes significantly as a function of acid concentration (Figure 2a) due to  
279 the formation of free uranyl species ( $\text{UO}_2^{2+}$ ) and uranyl nitrate complexes ( $\text{UO}_2(\text{NO}_3)_n^{(2-n)+}$ ,  $n = 1$   
280 or 2) at acid concentrations  $> \sim 0.1$  M  $\text{HNO}_3$ , which coexist in various proportions. Each species—  
281  $\text{UO}_2^{2+}$ ,  $\text{UO}_2\text{NO}_3^+$  and  $\text{UO}_2(\text{NO}_3)_2^{2+}$ —has a unique spectrum contribution. The maximum intensity  
282 of U(VI) fluorescence from  $\sim 0.5$  to 5 M is relatively consistent; however, the lifetime changes  
283 significantly from 3.1–1.1  $\mu\text{s}$  over this range and continues to decrease at even higher  $\text{HNO}_3$   
284 concentrations (see SI). The  $\text{Sm}^{3+}$  emission spectrum is also highly sensitive to changes in nitric  
285 acid concentration and likely corresponds to the formation of nitrate complexes (Figure 2b). To  
286 use the entire spectrum for regression, multivariate chemometrics must be applied to account for  
287 the overlapping and covarying spectral features.

288 The electronic transitions of  $\text{Sm}^{3+}$  vary in spectral shape and intensity with changing nitric acid  
289 concentration. Despite these solvatochromic shifts, the lifetime remains relatively constant from 1  
290 to 5 M  $\text{HNO}_3$  at 3.2–3.6  $\mu\text{s}$ . (Figure S3). Different energy levels in lanthanide fluorescence spectra  
291 occur due to several interactions within the ion, including the Coulombic interaction ( $10^4$   $\text{cm}^{-1}$ ),  
292 spin-orbit coupling ( $10^3$   $\text{cm}^{-1}$ ), and crystal field splitting ( $10^2$   $\text{cm}^{-1}$ ).<sup>21</sup> These free ion levels are  
293 described by the term symbols  $(^{2S+1})L_J$  where  $2S + 1$  denotes the total spin multiplicity,  $L$  the total  
294 orbital angular momentum, and  $J$  denotes the total angular momentum of the  $f$  electrons. The  
295 electric field in a coordinating environment further splits individual  $J$ -levels. This splitting appears  
296 as fine structure on each band. This information is only visible with higher-resolution  
297 spectrometers and is often ignored in the literature. However, fine structure is useful for

298 determining symmetry and coordination environment.<sup>15</sup> The fine structure in the Sm<sup>3+</sup> emission  
299 bands is particularly evident in the <sup>4</sup>G<sub>5/2</sub> → <sup>6</sup>H<sub>7/2</sub> emission line near 596 nm. The shoulder near  
300 604 nm becomes less pronounced with increasing acidity. This suggests that the transitions acquire  
301 strength at least in part due to the coordinating environment and have electric dipole (ED)  
302 character. ED-allowed transitions are more sensitive to the coordinating environment, and stronger  
303 ligand fields lead to more intense transitions.



304  
305 **Figure 2.** Example laser fluorescence spectra ( $\lambda_{\text{ex}} = 405 \text{ nm}$ ) of solutions containing  $100 \mu\text{g}\cdot\text{mL}^{-1}$   
306 U(VI) (a) and  $100 \mu\text{g}\cdot\text{mL}^{-1}$  Sm<sup>3+</sup> (b) with varying nitric acid concentration (0.1 M–5 M HNO<sub>3</sub>).  
307 The Raman nitrate (NO<sub>3</sub><sup>-</sup>) symmetric N–O stretch peak is labeled (b).

308 In addition to U(VI) and Sm<sup>3+</sup> emission peaks, Stokes Raman scattering features corresponding  
309 to free acid (H<sup>+</sup>), nitrate anions (NO<sub>3</sub><sup>-</sup>), and the O–H stretching region were identified  
310 simultaneously.<sup>26,30</sup> Unperturbed nitrate ions have three Raman active bands including  $\nu_1$  (~1048  
311 cm<sup>-1</sup>),  $\nu_3$  (~1415 cm<sup>-1</sup>), and  $\nu_4$  (~717 cm<sup>-1</sup>). The O–H vibrational stretching region consists of  
312 several overlapping bands attributed to various H<sub>2</sub>O and O–H (free and bound) vibrations (455–  
313 475 nm). An isosbestic point at 461.9 nm was observed with increasing nitric acid concentration  
314 (Figure 2). This Raman band is sensitive to the presence of cation, anions, ion strength, and  
315 temperature and is generally best described using multivariate data analysis.<sup>25,26</sup>

316 One advantage of using an excitation wavelength of 405 nm is that the water band at ~460 nm  
317 does not overlap significantly with the absorption spectrum of U(VI) (see SI). Either 337 nm or  
318 355 nm laser excitation sources are normally used for U(VI) fluorescence studies.<sup>8</sup> These  
319 wavelengths coincide with absorption bands of nitric acid, uranium, or several fission products,  
320 complicating measurement due to self-absorption effects.<sup>5,9</sup> The fluorescence signal is dependent  
321 on the optical pathlength for laser excitation and the pathlength for emission collection.<sup>10</sup> The  
322 complications that arise due to self-absorption are minimized in this work, which utilizes a 405  
323 nm laser and a smaller pathlength cuvette (i.e., 5 × 2.5 mm) instead of the more common 10 × 10  
324 mm.

### 325 **3.2 Univariate Analysis to LOD and LOQ**

326 A key component of a quality analytical chemistry paper includes the transparent calculation of  
327 limit of detection (LOD) and limit of quantification (LOQ). For this experiment, spectra were  
328 collected at room temperature (22°C) using a 4 s integration time and 100 mW laser power at  
329 405 nm. Calibration regression curves were generated for the primary 510 nm U(VI) peak, the  
330 most intense Sm<sup>3+</sup> emission line 595 nm peak, and the most resolved line relative to uranium at

331 702 nm (see SI). The calibration curves for each emission line were used to calculate LOD and  
332 LOQ by Equations 4 and 5:

$$333 \quad LOD = \frac{3 \cdot s}{m} \quad (4)$$

$$334 \quad LOQ = \frac{10 \cdot s}{m} \quad (5)$$

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

344 **Table 1.** LOD and LOQ of U(VI) and Sm<sup>3+</sup> in 1 M HNO<sub>3</sub> using a 405 nm laser at 22°C and 4 s  
345 integration time.

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

346

### 347 3.3 D-optimal design sample selection

348 Models were built using calibration and validation sets selected by D-optimal experimental  
349 designs. Recent work has shown that this approach can minimize resource consumption while  
350 maintaining or improving PLSR prediction performance. Selecting training sets using a one-factor-  
351 at-a-time approach generally results in many samples, particularly when more than two factors are

352 considered. The expected application of this work is monitoring in a restrictive glove box or hot  
 353 cell environment. Therefore, D-optimal designs were evaluated to minimize the number of samples  
 354 in the training set as well as minimize time and resource consumption.<sup>5</sup> A D-optimal design (25  
 355 points) was used to select the concentrations studied in this work. This design comprised 10  
 356 required model points and was augmented with 15 LOF points. Inclusion of LOF points generally  
 357 indicates that a higher-order process model is needed. To achieve an FDS of 0.99, five LOF points  
 358 were included in the calibration set. The 10 remaining LOF points were used as the validation set.

359 U(VI) concentration spanned the conditions expected in raffinate waste streams (1–100  $\mu\text{g}\cdot\text{mL}^{-1}$ )  
 360 <sup>1</sup>).<sup>5,9</sup> Detecting less than 1  $\mu\text{g}\cdot\text{mL}^{-1}$  uranium generally is not cause for concern in most applications.  
 361 The model covered the acid concentration range from 0.1 to 4  $\text{HNO}_3$ , which covers both a raffinate  
 362 stream (i.e., 3–4 M  $\text{HNO}_3$ ) and low-acid-strip solutions.<sup>5</sup>  $\text{Sm}^{3+}$  concentrations were chosen based  
 363 on potential burnups.<sup>9</sup> Generally,  $\text{Sm}^{3+}$  is found in raffinate solutions at a higher acid  
 364 concentration. Accounting for acid dependence in this data set adds complexity, which tests the  
 365 limits of the modeling approach. It was included in some model points and not included in others  
 366 to determine if it strengthened U(VI) model predictions, particularly with varying temperature.

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

Run	U(VI) ( $\mu\text{g}\cdot\text{mL}^{-1}$ )	$\text{HNO}_3$ (M)	$\text{Sm}^{3+}$ ( $\mu\text{g}\cdot\text{mL}^{-1}$ )	Space type	Build type
1*†	100.00	3.22	160.00	Plane	Lack of fit
<b>2</b>	<b>100.00</b>	<b>0.10</b>	<b>0.00</b>	<b>Vertex</b>	<b>Model</b>
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
<b>11</b>	<b>1.00</b>	<b>4.00</b>	<b>0.00</b>	<b>Vertex</b>	<b>Model</b>

12	<b>100.00</b>	<b>1.50</b>	<b>200.00</b>	<b>Edge</b>	<b>Model</b>
13	15.85	2.36	0.00	Plane	Lack of fit
14 <sup>†</sup>	30.70	1.27	163.00	Interior	Lack of fit
<b>15</b>	<b>59.91</b>	<b>2.42</b>	<b>0.00</b>	<b>Plane</b>	<b>Model</b>
<b>16</b>	<b>100.00</b>	<b>4.00</b>	<b>70.27</b>	<b>Edge</b>	<b>Model</b>
<b>17</b>	<b>1.00</b>	<b>2.42</b>	<b>119.00</b>	<b>Plane</b>	<b>Model</b>
<b>18</b>	<b>36.15</b>	<b>4.00</b>	<b>200.00</b>	<b>Edge</b>	<b>Model</b>
19 <sup>*†</sup>	44.32	2.50	104.00	Interior	Lack of fit
20 <sup>†</sup>	100.00	0.20	86.82	Plane	Lack of fit
<b>21</b>	<b>60.26</b>	<b>0.10</b>	<b>120.00</b>	<b>Plane</b>	<b>Model</b>
<b>22</b>	<b>1.00</b>	<b>0.10</b>	<b>0.00</b>	<b>Vertex</b>	<b>Model</b>
23 <sup>*</sup>	1.00	3.03	200.00	Edge	Lack of fit
24	100.00	1.84	33.00	Plane	Lack of fit
<b>25</b>	<b>1.00</b>	<b>0.10</b>	<b>200.00</b>	<b>Vertex</b>	<b>Model</b>

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

370 A second D-optimal design was generated for temperature from 20°C to 45°C (see SI). This  
371 design included three required model points at 20°C, 32.5°C, and 45°C and two LOF points at  
372 26.25°C and 38.75°C. Although it is possible to combine temperature as a fourth factor in the  
373 concentration design, the number of temperature points in such a design may not account for the  
374 variation necessary to model it. Additionally, a four-factor design would create additional  
375 concentrations, resulting in more samples. Thus, the variable temperature, which is unique relative  
376 to concentration, was treated separately.

### 377 3.4 Stacked regression model development

378 In addition to the required model points, five LOF samples (1, 6, 8, 19, and 23) were arbitrarily  
379 chosen and included in the training set to begin building models. All five temperature levels were  
380 used to best model the effect on analyte spectral signatures. Each model was optimized by  
381 minimizing the RMSEP through numerous reiterations. PLSR (PLS-2) was attempted first, as it is  
382 one of the most widely applied techniques to correlate convoluted and covarying spectral features  
383 to analyte concentration.<sup>2,3,27,28</sup> The global PLSR model was built using most of the spectrum (410–

384 750 nm) after applying a first derivative with a first-order polynomial and 31 smoothing points (1,  
385 1, 31). Eight latent variables (i.e., factors) were included in the model based on the RMSECV  
386 versus latent variable plot (Figure S6). The number of latent variables was chosen based on when  
387 there was the last significant reduction in RMSECV (>10%) when an additional latent variable  
388 was added. The addition of the eighth latent variable corresponds to a 20, 69, 84, and 3.2%  
389 reduction in the RMSECV for U(VI), HNO<sub>3</sub>, Sm<sup>3+</sup>, and temperature, respectively. The global  
390 PLSR model predicted Sm<sup>3+</sup> and HNO<sub>3</sub> concentration relatively well (see Table 3). The model was  
391 unable to account for the multicollinear effects of temperature and U(VI) concentration on the  
392 spectral intensity. Parity plots showing the wide spread of temperature and U(VI) predictions  
393 against the reference values are shown in Figure S7.

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

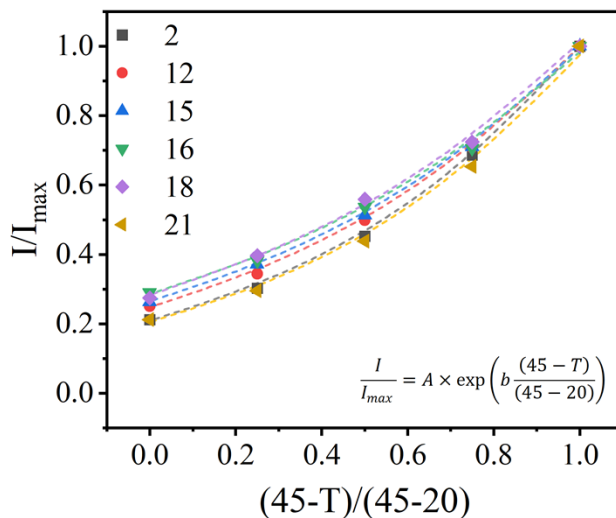
405 Next, other regression methods, including RR, RF, XGB, and stacked models, were investigated.  
406 Each regression model was optimized by tuning the preprocessing hyperparameters and trimming

407 limits. The optimal models for  $\text{Sm}^{3+}$  and  $\text{HNO}_3$  were developed using trimmed spectra (580–  
408 750 nm) and RR. This removed most of the neighboring U(VI) signal from inhibiting the  $\text{Sm}^{3+}$   
409 predictions. Similarly, the  $\text{HNO}_3$  model only used the wavelength range from 400 to 480 nm, using  
410 only the Raman nitrite peak and O–H stretching band for prediction. The RR model only slightly  
411 outperformed the seven-factor PLSR model for  $\text{HNO}_3$ , but RR provided the benefit of not requiring  
412 the user to perform latent variable selection.

413 The temperature model was built using the second derivative of the wavelength range 415–500  
414 nm. This range included the Raman O–H stretching band and the highest-energy U(VI) peak. This  
415 stacked regression model comprised an RR and PLSR serving as the submodels, and their  
416 predictions were analyzed by a 1,000-tree RF regression model. The resultant model had a strong  
417 predictive capability. The calibration and CV statistics for these three models are shown in Table  
418 3 as part of the overall ensemble. One interesting observation during CV was that RMSECV was  
419 much larger than the RMSEC. This was due to Sample 16, an edge sample with U(VI) and  $\text{HNO}_3$   
420 levels at their maximum, being predicted poorly when left out of the model (Figure 5), indicating  
421 its importance to the model. Sample 16 was located at the edge of the experimental design space,  
422 which explains why it would not be well predicted when left out of the calibration set.

423 The U(VI) signal was the most influenced by changes in temperature and system acidity. The  
424 models developed to characterize temperature and acidity were used to correct the U(VI) signal.  
425 Using six D-optimal model point calibration samples (Table 2), the 510.3 nm peak intensity was  
426 plotted as a function of temperature. The peak intensity was then normalized to the maximum  
427 510.3 nm peak of that sample (i.e., at 20°C), and the temperature was normalized to the model  
428 temperature range. The empirically derived plot (Figure 3) reveals a similar exponential  
429 relationship between the intensity and temperature. The coefficients of this relationship, shown as

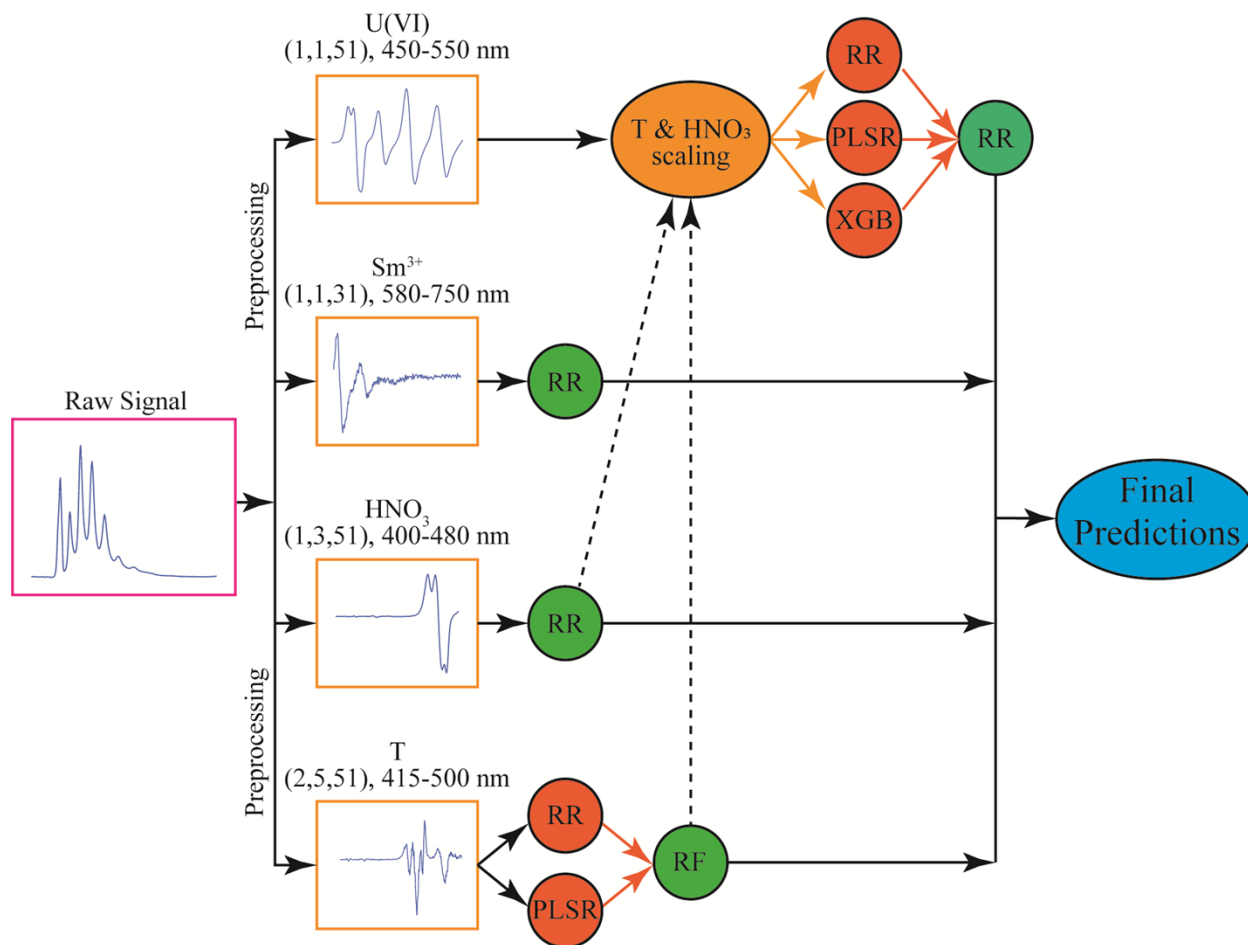
430 A and b in Figure 3, change only with acidity, not with U(VI) concentration. This relationship  
 431 allowed an additional preprocessing step for the U(VI) model where the trimmed spectra (450–  
 432 550 nm) were transformed to their respective  $I_{\max}$  values by scaling them using the relationship  
 433 shown in Figure 3, taking into consideration their temperature and acidity.  
 434



435  
 436 **Figure 3.** Temperature dependence of the 510.3 nm U(VI) fluorescence peak for six required  
 437 model points. The coefficients of the exponential fit were determined to be functions of acidity:  $A$   
 438  $= -0.0033 \times C_{HNO_3}^2 + 0.0334 \times C_{HNO_3} + 0.2038$  and  $b = 0.0152 \times C_{HNO_3}^2 - 0.1416 \times C_{HNO_3} +$   
 439  $1.5773$ .

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

446 prediction. The calibration, CV, and validation statistics shown in Table 3 were calculated using  
 447 the known temperatures and acidities for the adjustment. The final ensemble model included  
 448 combining all four models into one flow sheet where the temperature, HNO<sub>3</sub>, and Sm<sup>3+</sup>  
 449 concentrations would be predicted first, and then these predictions were used for scaling the spectra  
 450 before regression by the U(VI) model (Figure 4). The predictions of the final ensemble model are  
 451 visualized in Figure 5.



452  
 453 **Figure 4.** Overall ensemble scheme detailing how each model is connected.

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

457 species: 95, 22, 10, and 164% difference for U(VI), Sm<sup>3+</sup>, HNO<sub>3</sub>, and temperature, respectively.  
 458 Again, a Tukey-Kramer test at a 95% confidence level was performed and for all four factors the  
 459 overall ensemble model showed a statistically significant difference compared to the global and  
 460 trimmed PLSR models (Figure S8).<sup>29,30</sup>

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

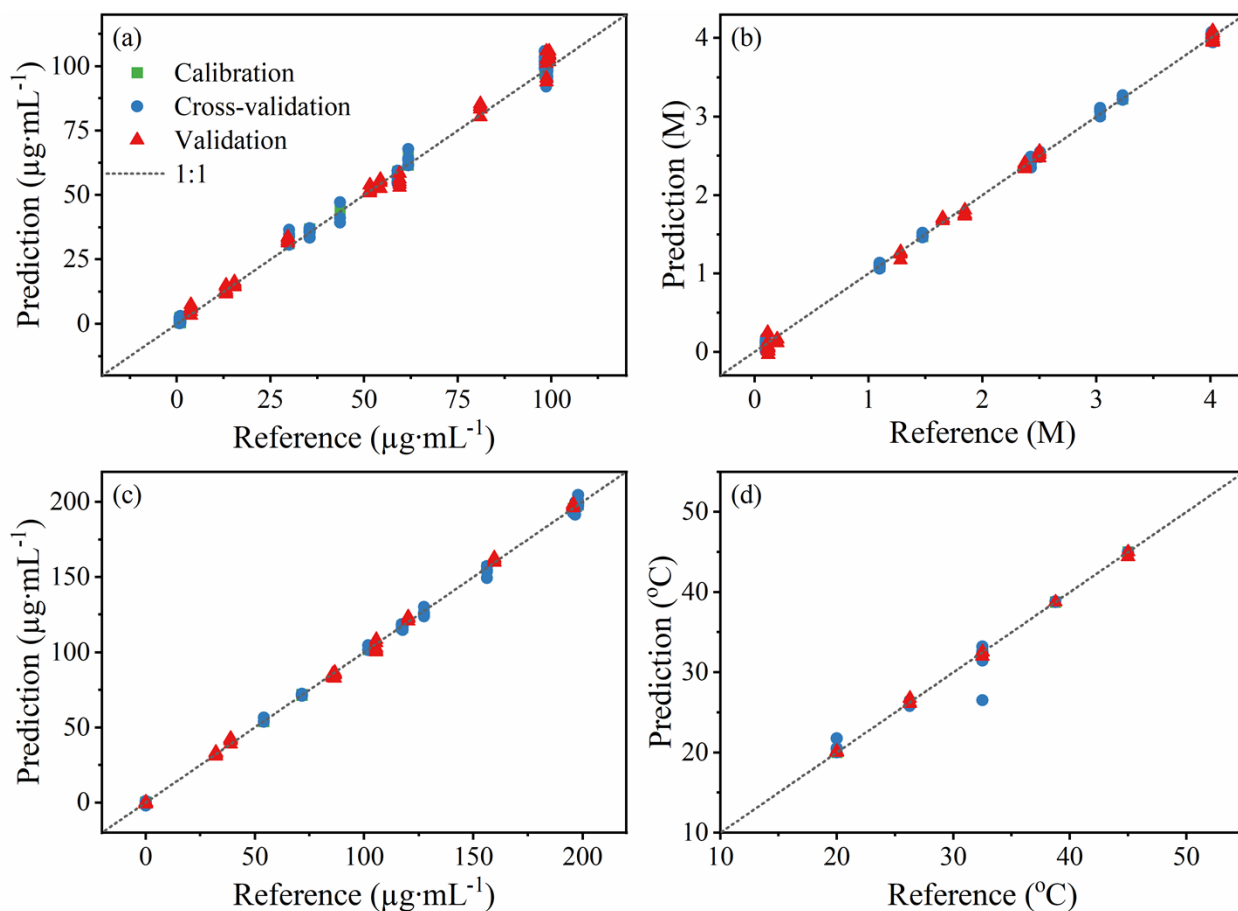
468 Another note is that these models were developed with 10 model points and 5 randomly selected  
 469 LOF points; given the complexity of the system more LOF points may improve the validation  
 470 statistics. Similarly, an alternate set of LOF points may be better suited for the calibration set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

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

Model	Global PLSR	Trimmed PLSRs	Overall ensemble
<b>Preprocessing</b>			
U(VI)		(1,1,31), 450–550 nm, 7 LVs	(1,1,51), 450–550 nm, SR1
Sm <sup>3+</sup>	(1,1,31), 410–	(1,1,31), 580–750 nm, 6 LVs	(1,1,31), 580–750 nm, RR
HNO <sub>3</sub>	750 nm, 8 LVs	(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
<b>Calibration/CV statistics</b>			
R <sup>2</sup> (U(VI))	0.9672	0.9718	0.998
RMSEC	7.0	6.5	1.7
RMSECV	14	10	2.8

$R^2$ ( $\text{Sm}^{3+}$ )	0.9988	0.9995	0.999
RMSEC	2.7	1.61	0.37
RMSECV	3.4	2.23	1.87
$R^2$ ( $\text{HNO}_3$ )	0.9976	0.9996	0.999
RMSEC	0.075	0.031	0.007
RMSECV	0.13	0.043	0.036
$R^2$ ( $^{\circ}\text{C}$ )	0.9318	0.9939	0.999
RMSEC	2.3	0.69	0.003
RMSECV	3.2	1.1	0.74
<b>Validation statistics</b>			
RMSEP (U(VI))	9.4	8.4	3.0
RMSEP%	19%	17%	6.2%
RMSEP ( $\text{Sm}^{3+}$ )	3.0	2.5	2.0
RMSEP%	3.0%	2.5%	2.0%
RMSEP ( $\text{HNO}_3$ )	0.14	0.073	0.066
RMSEP%	7.2%	3.7%	3.4%
RMSEP ( $^{\circ}\text{C}$ )	2.6	1.4	0.14
RMSEP%	21%	11%	1.1%

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



479  
 480 **Figure 5.** Predicted (a) U(VI), (b) HNO<sub>3</sub>, (c) Sm<sup>3+</sup> concentrations, and (d) temperature as  
 481 determined by the overall ensemble model compared to normalized reference (known) values.

### 482 3.6 Optimizing LOF points

483 To find the minimal number of samples in the training set and determine which LOF points  
 484 improved the calibration, each combination of 1–5 LOF points was tested in the training set, and  
 485 the remaining LOF points were treated as the validation set. This power set resulted in 4,944  
 486 combinations that were evaluated in two ways including which LOF combinations best improved  
 487 the overall RMSEP and which LOF combinations best improved the U(VI) RMSEP. The top  
 488 overall combination and U(VI) combinations for each level of LOF inclusion are detailed in Table  
 489 S3 in the SI. The top U(VI) model used only four LOF points (1, 14, 19, 20) and resulted in

490 RMSEP% values of 5.2%, 1.9%, 3.0%, and 2.3% for U(VI), Sm<sup>3+</sup>, HNO<sub>3</sub>, and temperature,  
491 respectively. This combination improved U(VI) prediction significantly with the trade-off of  
492 losing some temperature accuracy; however, given that U(VI) is the primary analyte of interest,  
493 this combination of model and LOF points would be best used to recalibrate a model in a new  
494 environment such as a hot cell.

495 Another benefit of considering all the top combinations was that the most important LOF points  
496 in the design space could be inferred. For overall prediction accuracy, Samples 1, 19, 20, 23, and  
497 24 were found to be the most beneficial. For U(VI) prediction accuracy, Samples 1, 14, 19, 20,  
498 and 24 were found to be the most beneficial. Interestingly, samples 1, 19, 20, and 24 are  
499 represented in both lists; Samples 1, 20, and 24 are all plane points with the maximum U(VI)  
500 concentration, and sample 19 is an interior point nearly at the center of the overall design space.

#### 501 **4. CONCLUSIONS**

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

509 The work presented here was essential to demonstrate key concepts related to minimizing the  
510 number of samples in the training set and building stacked regression models to account for  
511 temperature fluctuations, prior to characterizing a more complicated system. To demonstrate

512 greater applicability in real-world processing solutions, future work will include additional factors  
513 in the training set to account for self-absorption, quenching, and additional peak overlap  
514 effects.<sup>11,43</sup> These species may include fission products (e.g., Zr, Mo, Ru); lanthanides (e.g., Eu<sup>3+</sup>,  
515 Ce<sup>3+</sup>, Nd<sup>3+</sup>); corrosion products (e.g., Ni<sup>2+</sup>, Cr<sup>3+</sup>, Fe<sup>3+</sup>) and phosphorous-containing breakdown  
516 products. This method will also be applied to a more challenging validation set that includes a flow  
517 loop demonstration.

## 518 ASSOCIATED CONTENT

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

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

## 522 AUTHOR INFORMATION

### 523 Corresponding Author

524 \*Telephone: +1 (865) 574-1167; email address: [sadergaskilr@ornl.gov](mailto:sadergaskilr@ornl.gov)

### 525 Author Contributions

526 The manuscript was written using contributions of all authors. All authors have given approval to  
527 the final version of the manuscript.

### 528 Funding Sources

529 Funding for this work was provided by the National Nuclear Security Administration's (NNSA)  
530 Office of Defense Nuclear Nonproliferation Research and Development mission under contract

531 DE-AC05-00OR22725. This work used resources at the Radiochemical Engineering  
532 Development Center operated by the US Department of Energy's Oak Ridge National  
533 Laboratory (ORNL).

#### 534 **Notes**

535 The authors declare no competing financial interest.

#### 536 **ACKNOWLEDGMENT**

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

#### 539 **REFERENCES**

- 540 1. J-Y. C. Colle, D. Manara, T. Geisler and R. J. M. Konings, Advances in the application of Raman  
541 spectroscopy in the nuclear field. *Spectrosc. Eur.* 2020, **32** (6), 9–13.
- 542 2. D. Kirsanov, A. Rudnitskaya, A. Legin and V. Babain, UV-VIS spectroscopy with chemometric  
543 data treatment: An option for on-line control in nuclear industry. *J. Radioanal. Nucl. Chem.* 2017,  
544 **312**, 461–470.
- 545 3. L. R. Sadergaski, D. W. DePaoli and K. G. Myhre, Monitoring the Caustic Dissolution of  
546 Aluminum Alloy in a Radiochemical Hot Cell Using Raman Spectroscopy. *Appl. Spectrosc.* 2020,  
547 **74** (10), 1252–1262.
- 548 4. G. J. Lumetta, J. R. Allred, S. A. Bryan, G. B. Hall, T. G. Levitskaia, A. M. Lines and S. I. Sinkov,  
549 Simulant Testing of a Co-Decontamination (CoDCon) Flowsheet for a Product with a Controlled  
550 Uranium-to-Plutonium Ratio. *Sep. Sci. Technol.* 2019, **54** (12), 1977–1984.
- 551 5. B. H. Deniau, P. Decambox, P. Mauchien and C. Moulin, Time-Resolved Laser-Induced  
552 Spectrofluorometry of  $\text{UO}_2^{2+}$  in Nitric Acid Solutions. Preliminary Results for On-Line Uranium  
553 Monitoring Applications. *Radiochim. Acta* 1993, **61**, 23–28.

- 554 6. R. A. Peterson, E. C. Buck, J. Chun, R. C. Daniel, D. L. Herting, E. S. Ilton, G. J. Lumetta and S.  
555 B. Clark, Review of the Scientific Understanding of Radioactive Waste at the U.S. DOE Hanford  
556 Site. *Environ. Sci. Technol.* 2018, **52**, 381–396.
- 557 7. H. Fujimori, T. Matsui, and K. Suzuki, K. Fluorometry of  $\text{UO}_2^{2+}$  Ion in Nitric Acid Solutions. *J*  
558 *Nucl. Sci. Technol.* 1986, **23**, 1069–1074.
- 559 8. C. Moulin, P. Decambox, L. Couston and D. Pouyat, Time-Resolved Laser-Induced Fluorescence  
560 of  $\text{UO}_2^{2+}$  in Nitric Acid Solutions. Comparison between Nitrogen and Tripled Nd-YAG laser. *J*  
561 *Nucl. Sci. Technol.* 1994, **31**, 691–699.
- 562 9. T. Matusi, H. Fujimori and K. Suzuki, Effects of Coexisting Ions upon  $\text{UO}_2^{2+}$  Fluorescence in Fuel  
563 Reprocessing Solutions. *J Nucl. Sci. Technol.* 1988, **25**, 868–874.
- 564 10. C. Moulin, P. Decambox and P. Mauchien, Direct Uranium(VI) and Nitrate Determinations in  
565 Nuclear Reprocessing by Time-Resolved laser-Induced Fluorescence. *Anal. Chem.* 1996, **68**,  
566 3204–3209.
- 567 11. T. Matsui, T. Kitamori, H. Fujimori, K. Suzuki and M. Sakagami, In-Line Uranium Monitoring in  
568 Reprocessing Waste Solution by Time-Resolved Laser-Induced Fluorometry. *J Nucl. Sci. Technol.*  
569 1992, **29**, 664–670.
- 570 12. N. S. Smith, G. S. Cerefice and K. R. Czerwinski, Fluorescence and absorbance spectroscopy of  
571 the uranyl ion in nitric acid for process monitoring applications. *J. Radioanal. Nucl. Chem.* 2013,  
572 **295**, 1553–1560.
- 573 13. L. M. Colletti, R. Copping, K. Garduno, E. J. W. Lujan, A. K. Mauser, A. Mechler-Hickson, I.  
574 May, S. D. Reilly, D. Rios, J. Rowley and A. B. Schroeder. The application of visible absorption  
575 spectroscopy to the analysis of uranium in aqueous solutions. *Talanta*, 2017, **175**, 390–405.
- 576 14. T. Saito, N. Aoyagi and T. Kimura. Time-resolved laser-induced fluorescence spectroscopy  
577 combined with parallel factor analysis: a robust technique for  $\text{UO}_2^{2+}$ . *J. Radioanal. Nucl. Chem.*  
578 2015, **303**, 1129–1132.

- 579 15. I. N. Izosimov, N. G. Firsin, N. G. Gorshkov and S. N. Nekhoroshkov. Detection of lanthanides  
580 and actinides in solutions based on laser-induced luminescence and chemiluminescence. *Hyperfine*  
581 *Interact.* 2014, **227**, 271–281.
- 582 16. T. Kimura, R. Nagaishi, T. Ozaki, M. Arisaka and Z. Yoshida, Uranium(VI) Speciation at Elevated  
583 Temperatures and Pressures by Time-resolved Laser-induced Fluorescence Spectroscopy. *J Nucl.*  
584 *Sci. Technol.* 2002, **3**, 233–239.
- 585 17. C. Talbot-Eeckelaers, S. J. A. Pope, A. J. Hynes, R. Copping, C. J. Jones, R. J. Taylor, S. Faulkner,  
586 D. Sykes, F. R. Livens and I. May. Luminescence from Neptunyl(VI) Species in Solution. *J. Am.*  
587 *Chem. Soc.* 2007, **129**, 2442–2443.
- 588 18. R. G. Denning, Electronic Structure and Bonding in Actinyl Ions and their Analogs. *J. Phys. Chem.*  
589 *A* 2007, **111**, 4125–4143.
- 590 19. S. Ganesh, F. Khan, M. K. Ahmed, P. Velavendan, N. K. Pandey, K. Mudali and S. K. Pandey,  
591 Determination of ultra traces amount of uranium in raffinates of Purex process by laser fluorimetry.  
592 *J. Radioanal. Nucl. Chem.* 2012, **292**, 331–334.
- 593 20. D. Y. Chung, E. H. Lee, J. H. Yoo and T. Kimura, Laser-Induced Fluorescence Studies on Sm(III)  
594 and Cm(III) Complexes in the HNO<sub>3</sub>/DHDECMP Extraction System. *J. Nucl. Sci.* 2002, **3**, 332–  
595 335.
- 596 21. W. T. Carnall, P. R. Fields and K. Rajnak, Electronic Energy Levels in the Trivalent Lanthanide  
597 Aquo Ions. I. Pr<sup>3+</sup>, Nd<sup>3+</sup>, Pm<sup>3+</sup>, Sm<sup>3+</sup>, Dy<sup>3+</sup>, Ho<sup>3+</sup>, Er<sup>3+</sup>, and Tm<sup>3+</sup>. *J. Chem. Phys.* 1968, **49**, 4424–  
598 4442.
- 599 22. K. B. Hong, K. W. Jung and K.-H. Jung, Application of Laser-Induced Fluorescence for  
600 Determination of Trace Uranium, Europium and Samarium. *Talanta* 1989, **36**, 1095–1099.
- 601 23. S. Goderski, M. Runowski, P. Wozny, V. Lavin and S. Lis, Lanthanide Upconverted  
602 Luminescence for Simultaneous Contactless Optical Thermometry and Manometry – Sensing

- 603 under Extreme Conditions of Pressure and Temperature. *CS Appl. Mater. Interfaces* 2020, **12**,  
604 40475–40485.
- 605 24. U. K. Maity, P. Manoravi, M. Joseph and N. Sivaraman, Laser-induced breakdown spectroscopy  
606 for simultaneous determination of lighter lanthanides in actinide matrix in aqueous medium.  
607 *Spectrochim. Acta B: At. Spectrosc.* 2022, **190**, 106393.
- 608 25. A. M. Lines, G. L. Nelson, J. M. Casella, J. M. Bello, S. E. Clark and S. A. Bryan, Multivariate  
609 Analysis to Quantify Species in the Presence of Direct Interferents: Micro-Raman Analysis of  
610 HNO<sub>3</sub> in Microfluidic Devices. *Anal. Chem.* 2018, **90** (4), 2548–2554.
- 611 26. A. J. Casella, T. G. Levitskaia, J. M. Peterson and S. A. Bryan, Water O–H Stretching Raman  
612 Signature for Strong Acid Monitoring via Multivariate Analysis. *Anal. Chem.* 2013, **85** (8), 4120–  
613 4128.
- 614 27. P. Tse, J. Shafer, S. A. Bryan, G. L. Nelson and A. M. Lines. Measuring Nd(III) Solution  
615 Concentration in the Presence of Interfering Er(III) and Cu(II) Ions: A Partial Least Squares  
616 Analysis of Ultraviolet-Visible Spectra. *Appl. Spectrosc.* 2022, **76**, 173–183.
- 617 28. P. Tse, J. Shafer, S. A. Bryan and A. M. Lines, Quantification of Raman-Interfering Polyoxoanions  
618 for Process Analysis” Comparison of Different Chemometric Models and a Demonstration on Real  
619 Hanford Waste. *Environ. Sci. Technol.* 2021, **55**, 12943–12950.
- 620 29. L. R. Sadergaski, G. K. Toney, L. H. Delmau and K. G. Myhre. Chemometrics and Experimental  
621 Design for the Quantification of Nitrate Salts in Nitric Acid: Near-Infrared Spectroscopy  
622 Absorption Analysis. *Appl. Spectrosc.* 2021, **75** (9), 1155–1167.
- 623 30. L. R. Sadergaski, T. J. Hagar and H. B. Andrews. Design of Experiments, Chemometrics, and  
624 Raman Spectroscopy for the Quantification of Hydroxylammonium, Nitrate, and Nitric Acid. *ACS*  
625 *Omega*, 2022, **7**, 7287–7296.
- 626 31. V. Czitrom. One-Factor-at-a-Time Versus Designed Experiments. *Am. Stat.* 1999, **53** (2), 126–131.

- 627 32. B. Smucker, M. Krzywinski and N. Altman, Optimal experimental design. *Nat. Methods* 2018, **15**,  
628 559–560.
- 629 33. A. Zahran, C. M. Anderson-Cook and R. H. Myers, Fraction of Design Space to Assess Prediction  
630 Capability of Response Surface Designs. *J. Qual. Tech.* 2003, **35** (4), 377–386.
- 631 34. P. Shan, Y. Zhao, Q. Wang, X. Sha, X. Lv, S. Peng and Y. Ying. Stacked ensemble extreme  
632 learning machine coupled with Partial Least Squares-based weighting strategy for nonlinear  
633 multivariate calibration. *Spectrochim. Acta A: Mol. and Biomol. Spectrosc.*, 2019, **215**, 97–111.
- 634 35. R. Lascola, P. E. O'Rourke and E. A. Kyser. A Piecewise Local Partial Least Squares (PLS)  
635 Method for the Quantitative Analysis of Plutonium Nitrate Solutions. *Appl. Spectrosc.* 2017, **7**  
636 (12), 2579–2594.
- 637 36. M. Mancini, V.-M. Taavitsainen and G. Toscano. Comparative study between Partial Least  
638 Squares and Rational function Ridge Regression models for the prediction of moisture content of  
639 woodchip samples using a handheld spectrophotometer. *J. Chemometr.* 2021, **35**, e3337. DOI:  
640 10.1002/cem.3337.
- 641 37. H. B. Andrews and K. G. Myhre. Quantification of Lanthanides in a Molten Salt Reactor Surrogate  
642 Off-Gas Stream Using Laser-Induced Breakdown Spectroscopy. *Appl. Spectrosc.* 2022,  
643 doi:10.1177/00037028211070323.
- 644 38. L. Wei, Z. Yuan, Y. Zhong, L. Yang, X. Hu and Y. Zhang. An improved gradient boosting  
645 regression tree estimation model for soil heavy metal (arsenic) pollution monitoring using  
646 hyperspectral remote sensing. *Appl. Sci.* 2019, **9** (9), 1943.
- 647 39. L. Wei, Z. Yuan, M. Yu, C. Huang and L. Cao. Estimation of Arsenic Content in Soil Based on  
648 Laboratory and Field Reflectance Spectroscopy. *Sensors* 2019, **19** (18), 3904,  
649 doi:10.3390/s19183904.
- 650 40. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P.  
651 Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M.

652 Perrot and É. Duchesnay. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 2011, **12**,  
653 2825–2830.

654 41. T. Chen and C. Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd*  
655 *SIGKDD Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, August  
656 13–17, 2016; ACM, New York, 2016, 785–794.

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

659 43. L. R. Sadergaski, K. G. Myhre, L. H. Delmau, Multivariate chemometric methods and Vis-NIR  
660 spectrophotometry for monitoring plutonium-238 anion exchange column effluent in a  
661 radiochemical hot cell. *Talanta Open* 2022, **5**, 100120.

662

663

664

665

666

667

668

669