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

This Supporting Information contains:

Pages: 10

Figures: 8

Tables: 3

Extended methods description

Steady-state excitation. An excitation spectrum is analogous to an absorbance spectrum. Excitation spectra are acquired when the emission wavelength is fixed and the excitation monochromator wavelength is scanned. The spectrum provides information about the wavelengths the molecule will absorb and emits a photon at a particular wavelength. Excitation spectra are more sensitive to specific species than an absorption spectrum, which measures all absorbing species in a solution.

Statistical model comparisons. The Tukey-Kramer method was used for the pairwise comparison of model RMSEPS for each factor, assuming the null hypothesis H_0: $\mu_i = \mu_j$. The RMSEP was separated into bias and standard error of prediction (SEP) using equations 1 and 2, respectively, and compared using a 95% t confidence interval.

$$bias = \frac{1}{n}\sum_{i=1}^{n} ei \tag{1}$$

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}(e_i - bias)^2}{n-1}} \tag{2}$$

$$e_i = (\hat{y}_i - y_i) \tag{2a}$$

At the level $\alpha = 0.05$, a type I error, falsely rejecting H_0, is made 5% of the time. To minimize this possibility, the critical t value was adjusted to account for α inflation and avoid inaccurate significant results. Prediction biases between two models were calculated at the 95% confidence interval using equation 3. The value s_e represents the standards error of the estimated difference where d_i is the difference in error between models being compared and $\bar{d}$ is the mean difference in error between the compared models. This was calculated with equation 4.

$$bias_1 - bias_2 \pm t_{n-1,0.025} \, x \, s_e \tag{3}$$

$$s_e = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n \, x \, (n-1)}} \tag{4}$$

Equations 5 and 6 were used to evaluate the 95% t confidence interval for the model SEPs. The value r represents the correlation coefficient between e_1 and e_2. The confidence interval was calculated using equation 5. If the confidence interval of the SEP ratios corresponding to two models contained the integer 1, they were not considered statistically different. The overall prediction performance of two models was considered statistically similar if the bias confidence interval contained 0 and the SEP ratio contained 1.

$$K = 1 + \frac{2(1-r^2) x \, t_{n-2,0.025}^2}{n-2} \tag{5}$$

$$L = \sqrt{K + \sqrt{(K^2 - 1)}} \tag{6}$$

$$\frac{SEP_1}{SEP_2} \, x \, \frac{1}{L} \, and \, \frac{SEP_1}{SEP_2} \, x \, L \tag{7}$$

Python Libraries.

Links to the open-source documentation for the Python packages employed are provided below.

Savitzky-Golay Filter:
https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.savgol_filter.html

Random Forrest:
https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

Stacked Regression:
https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.StackingRegressor.html

Partial Least Squares Regression:
https://scikit-learn.org/stable/modules/generated/sklearn.cross_decomposition.PLSRegression.html

Ridge Regression:
https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeCV.html

Extreme Gradient Boosted Regression:
https://xgboost.readthedocs.io/en/stable/python/python_api.html#module-xgboost.sklearn
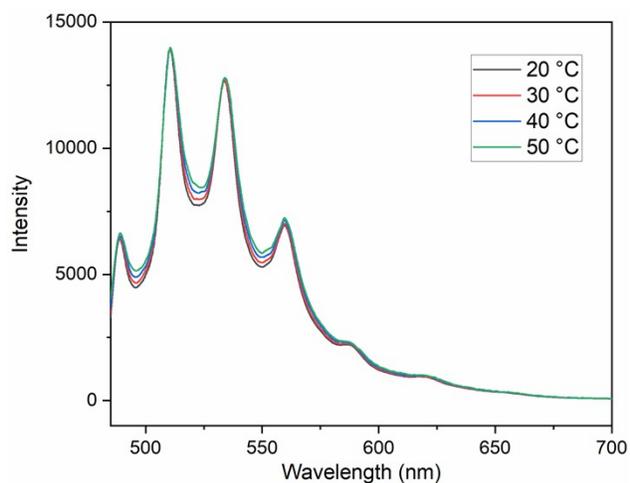
Figure S1. Normalized uranyl temperature fluorescence spectra based on the 510.3 nm peak. Sample contained 100 µg·mL⁻¹ U(VI) and 1 M HNO₃. Note the peaks sharpen slightly but have similar shape.

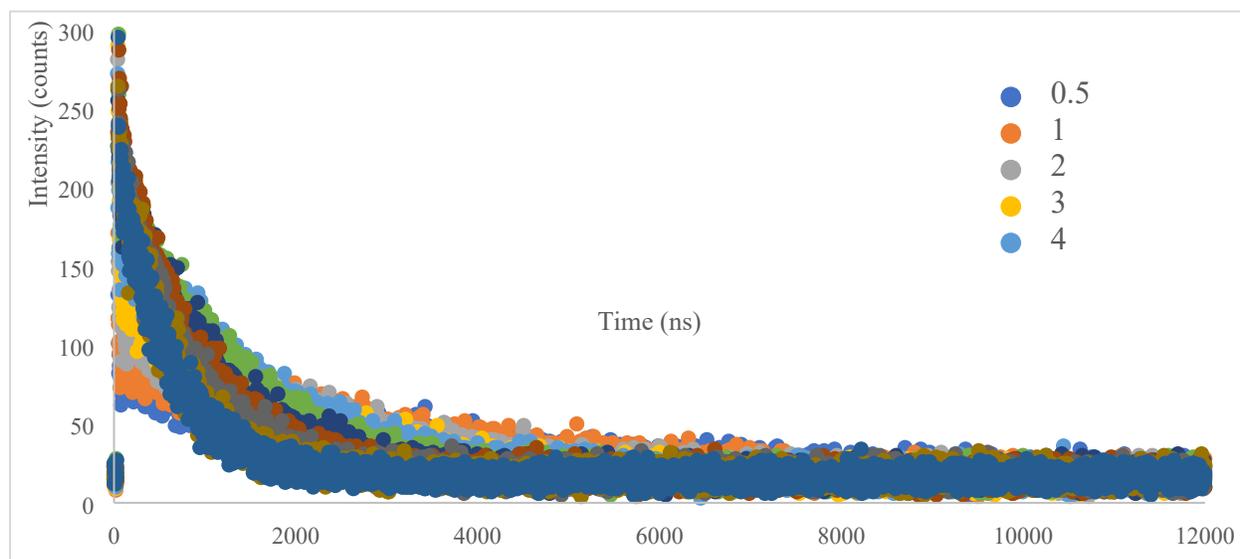**Figure S2.** Fluorescence decay curves for solutions containing 100 µg·mL⁻¹ U(VI) with varying nitric acid concentration (0.5–12 M).

Table 1. U(VI) fluorescence lifetimes with changing nitric acid concentration.

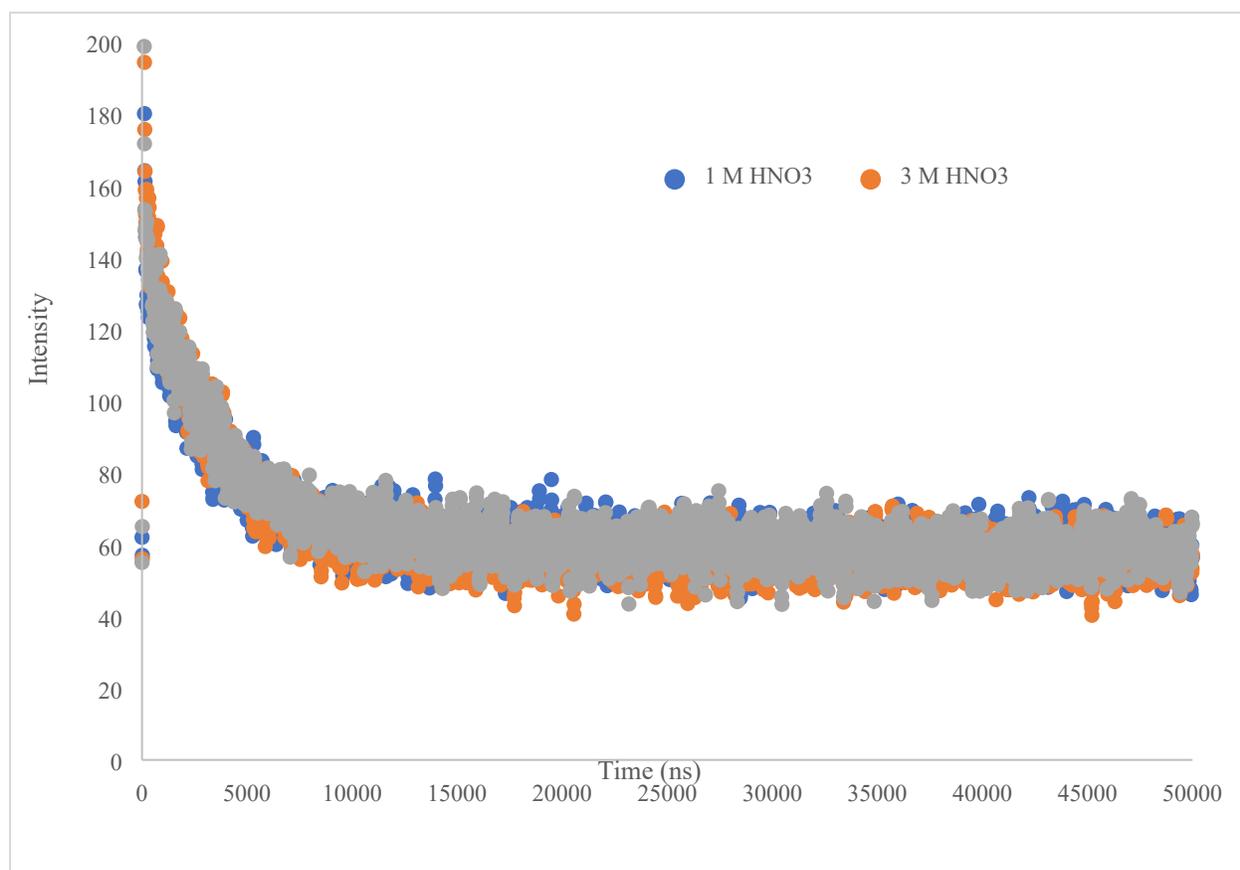| HNO$_3$ (M)	Lifetime (µs)
0.5	3.1
1	2.8
2	2.1
3	1.7
4	1.4
5	1.1
6	0.98
7	0.82
8	0.72
10	0.61
12	0.55

Figure S3. Decay curves for samples containing 200 µg·mL^{-1} Sm(III) and nitric acid.
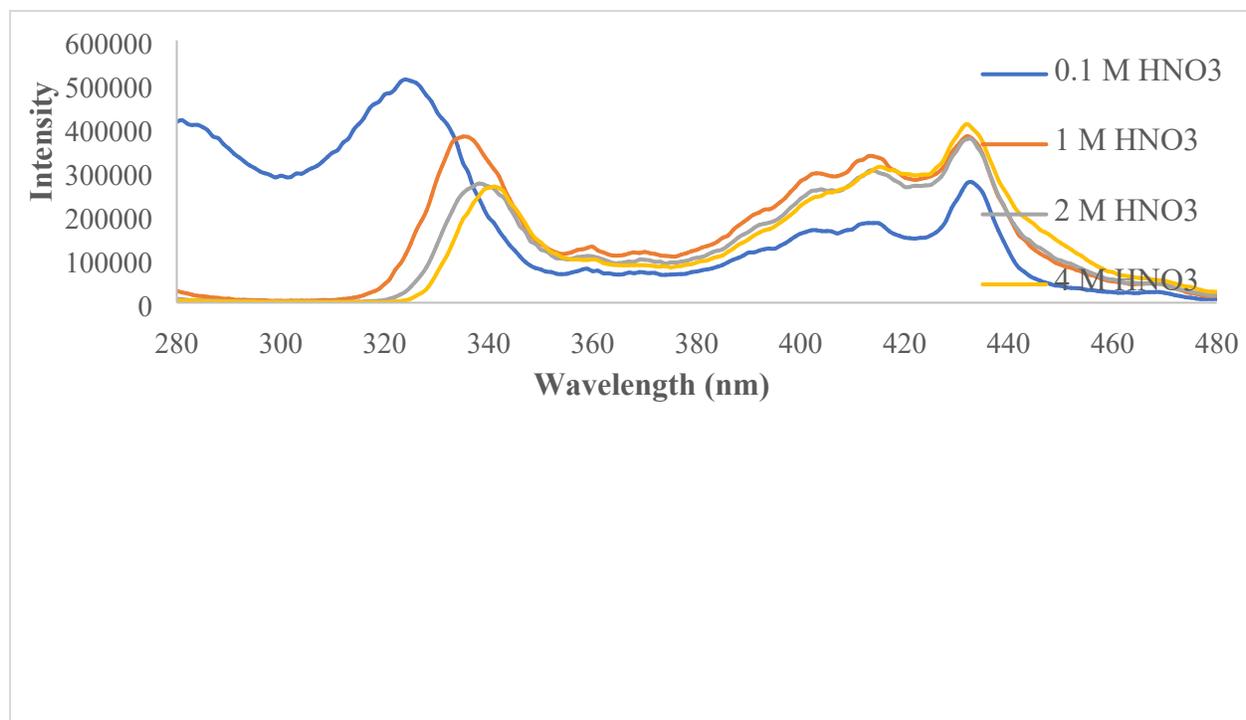
Figure S4. Excitation spectra of solutions containing 100 µg·mL^{-1} U(VI) and HNO$_3$ (0.1–4 M).
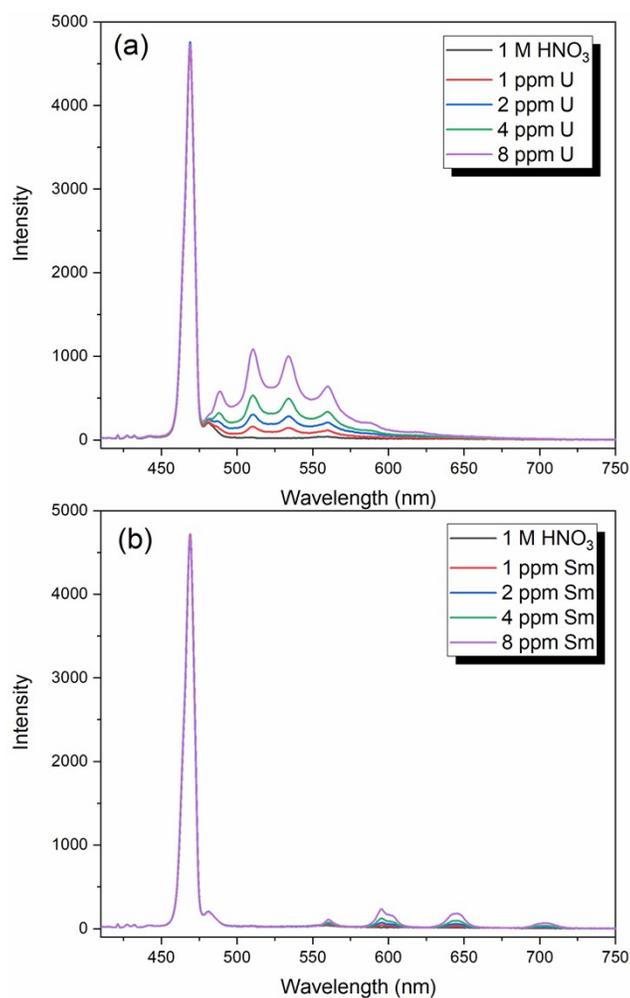
Figure S5. Laser-induced fluorescence spectroscopy spectra (λ_{ex} = 405 nm) of solutions containing 1–8 µg·mL^{-1} uranium(VI) (a) and 1–8 µg·mL^{-1} Sm(III) (b) with constant acid concentration (1.0 M HNO$_3$).

Table S2. D-optimal design matrix for sample temperatures.

Run	Temp. (°C)	Space type	Build type
1	20	Vertex	Model
2	26.25	AxialCB	Lack of fit
3	32.5	Center	Model
4	38.75	AxialCB	Lack of fit
5	45	Vertex	Model

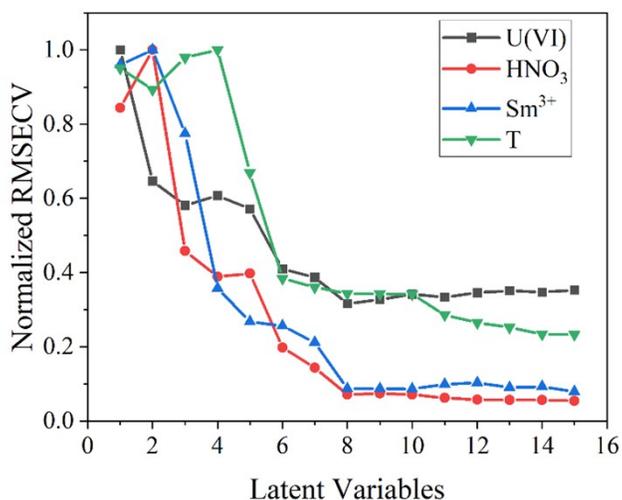*Note*: Abbreviations used in this table are temperature (Temp.).

Figure S6. Normalized RMSECV vs. the number of factors (i.e., latent variables) included in the initial partial least squares regression (PLSR) model. Here, all factors were modeled simultaneously using the PLS-2 approach.
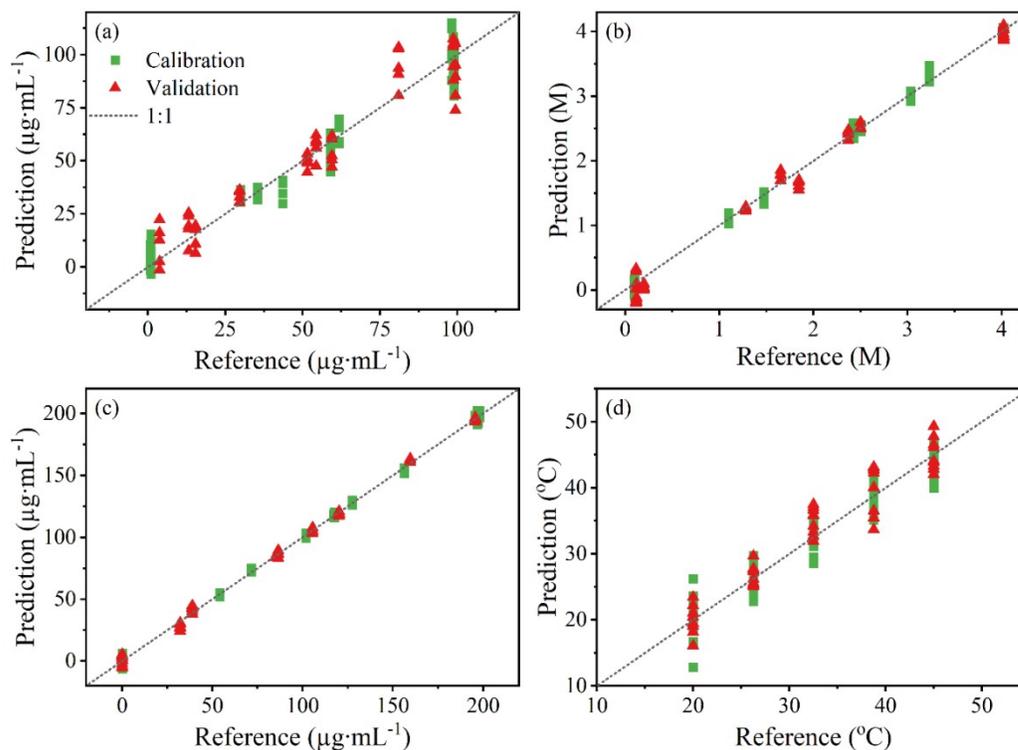

Figure S7. Parity plots from initial PLS model for (a) U(VI) concentration, (b) Sm^{3+} concentration, (c) nitric acid concentration, and (d) temperature and eight latent variables.
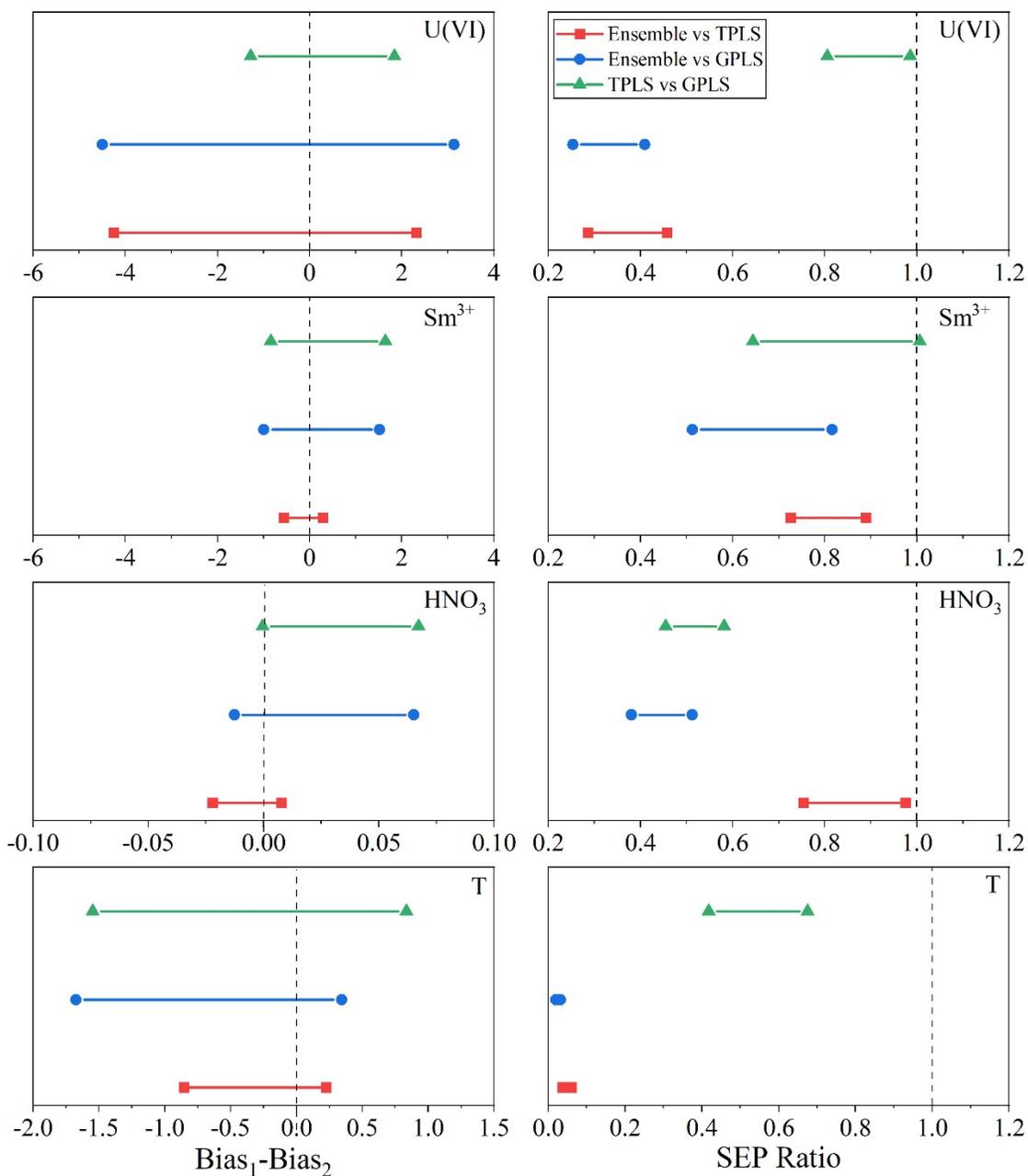
Figure S8. Bias and standard error of prediction (SEP) 95% confidence bands from the Tukey Kramer tests. All models were determined to be significantly different from one another, with the exception of the global PLSR (GPLS) and trimmed PLSR (TPLS) models for Sm^{3+}.

Table S3. Root-mean-square error of the prediction values for each species with various combinations of lack-of-fit (LOF) points included in the calibration set.

Combination	LOF points	Temp. (°C)	HNO$_3$ (M)	Sm(III) (µg·mL^{-1})	U(VI) (µg·mL^{-1})
1 LOF	14	0.67	0.083	2.22	3.89
2 LOF	14,20	0.71	0.067	2.36	3.30
3 LOF	1,13,19	0.62	0.061	1.94	3.01
4 LOF	1,14,19,20	0.29	0.061	1.91	2.56
5 LOF	1,14,19,20,24	0.18	0.070	2.03	2.58