Article

Hierarchical Modeling to Enhance Spectrophotometry Measurements—Overcoming Dynamic Range Limitations for Remote Monitoring of Neptunium

Hunter B. Andrews *[ID] and Luke R. Sadergaski [ID]

Radioisotope Science and Technology Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA
* Correspondence: andrewshb@ornl.gov

Abstract: A robust hierarchical model has been demonstrated for monitoring a wide range of neptunium concentrations (0.75–890 mM) and varying temperatures (10–80 °C) using chemometrics and feature selection. The visible–near infrared electronic absorption spectrum (400–1700 nm) of monocharged neptunyl dioxocation (Np(V) = NpO_2^+) includes many bands, which have molar absorption coefficients that differ by nearly 2 orders of magnitude. The shape, position, and intensity of these bands differ with chemical interactions and changing temperature. These challenges make traditional quantification by univariate methods unfeasible. Measuring Np(V) concentration over several orders of magnitude would typically necessitate cells with varying path length, optical switches, and/or multiple spectrophotometers. Alternatively, the differences in the molar extinction coefficients for multiple absorption bands can be used to quantify Np(V) concentration over 3 orders of magnitude with a single optical path length (1 mm) and a hierarchical multivariate model. In this work, principal component analysis was used to distinguish the concentration regime of the sample, directing it to the relevant partial least squares regression submodels. Each submodel was optimized with unique feature selection filters that were selected by a genetic algorithm to enhance predictions. Through this approach, the percent root mean square error of prediction values were ≤1.05% for Np(V) concentrations and ≤4% for temperatures. This approach may be applied to other nuclear fuel cycle and environmental applications requiring real-time spectroscopic measurements over a wide range of conditions.

Keywords: absorption spectroscopy; UV-Vis-NIR; partial least squares; principal component analysis; nitric acid; neptunium; temperature

1. Introduction

In optical spectroscopy applications (e.g., spectrophotometry), the linear dynamic range of a spectrometer, which is defined as the ratio between the maximum and minimum signal intensities, is an imperative variable to consider. In general, spectrophotometers are most accurate from approximately 0.4–0.9 absorbance units (Abs.) [1]. In high-absorbance situations, in which minimal light passes through the samples, the intensity may be too weak to measure. In low-absorbance situations, in which too much light passes through the sample, distinguishing the difference between the sample and reference (i.e., blank) spectrum is difficult. Overcoming dynamic range limitations is possible using multiple pathlength cuvettes and either multiple spectrometers or optical switches. Alternatively, some high-end spectrophotometers cover a wide dynamic range (~6–8 Abs.), but they are typically large, sluggish, and not amenable to remote applications with fiber optic cables. These significant caveats must be considered when deploying spectrophotometers for real-time, online monitoring applications in restrictive hot cell and control room environments. This work explores a more efficient option using hierarchical modeling to build regression

models that cover a wide range of Np(V) concentrations with a single optical pathlength and a plug-and-play spectrometer.

The timely quantification of the aqueous pentavalent neptunyl dioxocation ($^{237}NpO_2^+$) in nitric acid (HNO_3) is essential to improve Np processing efficiency, which is an ongoing effort of the ^{238}Pu Supply Program at Oak Ridge National Laboratory where ^{238}Pu is produced through irradiating ^{237}Np targets [2]. Improved analytical time will help scale up the production of heat source plutonium oxide (PuO_2) to meet NASA's projected needs [3]. Multiple analytical techniques are available for measuring Np concentration, including alpha- and gamma-ray spectroscopy, inductively coupled plasma–mass spectrometry (ICP-MS), titrimetric analysis, and optical spectroscopy. Spectrophotometry can provide rapid feedback (e.g., 10–1000 ms intervals) and is sensitive to Np valence and concentration [2]. This method can be readily employed in hazardous environments using fiber optic cables, making it ideal for remote, online monitoring applications [4,5].

The Np(V) cation is the most prevalent species in aqueous conditions in the absence of redox-active species and at acid concentrations <5 M HNO_3, making it the focal point of spectroscopic detection research [6]. The most intense $5f$–$5f$ electronic transition in the visible (vis)–near infrared (NIR) absorption spectrum of the aquo NpO_2^+ ion occurs at 979 nm ($\varepsilon = 395\ M^{-1}\cdot cm^{-1}$) [7,8]. The most intense transition originating from the $5f$ shell for the isoelectronic plutonyl ion (PuO_2^{2+}) occurs at a shorter wavelength near 830 nm [9]. The intense Np(V) 979 nm peak and a less intense band at 616 nm ($\varepsilon = 22\ M^{-1}\cdot cm^{-1}$) have been used analytically to determine the concentration of NpO_2^+ in solution using Beer's law. Additional bands are also available for analysis in the spectrum with even lower molar absorptivity values ($\varepsilon = 2.5\ M^{-1}\cdot cm^{-1}$) [10]. The 979 nm absorption band is the most widely used because the intensity and position is sensitive to the coordination environment of NpO_2^+ for most studies at relatively low Np concentrations (~1 mM) [11,12]. Quantifying Np(V) in aqueous solutions using spectrophotometry over dynamic conditions encountered during processing is challenging because of the broad range of molar extinction coefficients, nonlinear concentration responses, concentration-dependent chemical equilibria, and dynamic temperature effects [10,13,14].

Univariate regression methods (e.g., Beer's law) are not suitable for in-line analysis in this complex system [4]. Instead, multivariate chemometrics can be used to describe complex systems with convoluted or covarying spectral features [2]. For example, partial least squares regression (PLSR) correlates the entire spectrum to the dependent variables (e.g., concentration) [15]. Techniques like PLSR can be combined with multiple models to improve the measurement accuracy compared with a single model [5]. The main points of scientific advancement in this work were (1) investigating the use of principal component analysis (PCA) to distinguish the concentration regime of a given sample, (2) developing high accuracy supervised regression models to quantify Np(V) concentration and temperature, and (3) combining the established PCA classification, regression models, and a genetic algorithm for feature selection into a single hierarchical model in order to extend the dynamic range of a single spectrometer to cover a wide concentration range previously unattainable. The PLSR models accounted for temperature fluctuations, which is an important variable often ignored in laboratory studies [16]. This approach accurately quantified a wide range of Np(V) concentrations (0.0008–0.9 M) and temperatures (10–80 °C). Results indicate that hierarchical modeling can extend the dynamic range of plug-and-play spectrophotometers and quantify a wide range of concentrations using a single optical pathlength. This approach maintains the simplicity of the spectrophotometric approach while also maintaining feasibility with respect to real-world implementation in restrictive radiological environments.

2. Materials and Methods

2.1. Materials

All chemicals were commercially obtained (American Chemical Society grade) and used as received unless otherwise stated. Concentrated HNO_3 (70%) was purchased from

Sigma Aldrich. All solutions were prepared using deionized water with a resistivity of 18.2 MΩ·cm at 25 °C. Oak Ridge National Laboratory provided $^{237}NpO_2$ ($t_{\frac{1}{2}} = 2.14 \times 10^6$ years) in-house.

2.2. Methodology

Samples were prepared by dissolving NpO_2 in 8 M HNO_3 and diluting to achieve a Np concentration of 210 ± 7 g L^{-1} (0.886 M Np) in 1 M HNO_3. The concentration of Np in the sample was determined using ICP-MS (iCAP Q ICP-MS, Thermo Fisher Scientific). The stock solution was used to prepare the calibration and validation samples (0.00075–0.89 M Np). Aliquots from the stock solution were taken and sequentially diluted in 1.0 M HNO_3 using a 1 mL volumetric flask (1.00 $\pm$ 0.01 mL) to achieve the desired total Np concentration. Calibration and validation samples used in this study are detailed in Table 1. Here, more validation samples were used than the calibration set. This is representative of the true application where it is desired to minimize the training set to minimize resources used. A small fraction of Np(VI) ($\leq$3%) was present in the higher Np concentration samples. This oxidation state distribution is commonly observed in process solutions. No attempts were made to adjust the valence of the Np, and the models discussed in this study are for total Np concentration. Increasing the solution temperature did not change the ratio of Np(V/VI) or alter the Np(VI) band [17].

Table 1. Calibration and validation sample information.

Sample Set	Sample Tag	Np Concentration (M)	Tested Temperatures (°C)
Calibration	C1	0.00	10.0, 20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0
	C2	0.00075	
	C3	0.0069	
	C4	0.034	
	C5	0.069	
	C6	0.34	
	C7	0.52	
	C8	0.89	
Validation	V1	0.00	15.0, 16.5, 24.0, 25.0, 35.0, 45.0, 48.8, 55.0, 56.7, 65.0, 71.6, 75.0
	V2	0.00075	
	V3	0.0044	
	V4	0.0088	
	V5	0.017	
	V6	0.069	
	V7	0.17	
	V8	0.34	
	V9	0.64	
	V10	0.89	

2.3. Spectrophotometry

A QEPro spectrometer (Ocean Insight) was used for ultraviolet (UV)-vis absorption measurements, and a NIRQuest spectrometer (Ocean Insight) was used for NIR absorption measurements. UV-vis spectra included absorbance measurements every 0.78 nm from 199 to 985 nm, and NIR spectra included absorbance measurements recorded every 1.65 nm from 897 to 1717 nm. Triplicate spectra were recorded for each sample, and each spectrum was an average of five scans. The spectrometers were referenced to pure water at 20 °C prior

to each measurement unless otherwise stated. OceanView 2.0 software was used to process each spectrum.

The stabilized incoherent light source (SLS201L, ThorLabs) was transmitted through several meters of multimode optical fiber patch cables (Ocean Insight) with a core diameter of 600 μm into and out of the glove box. Samples were introduced into the Hellma quartz micro flow cell with an optical pathlength of 1 mm and a Z height of 8.5 mm (i.e., distance from bottom of the cuvette to the light transmission point) using a syringe. A Quantum Northwest Qpod 2e temperature-controlled sample compartment holder (Avantes) with two quantum cascade laser-UV collimating lenses was used to adjust the temperature of the samples with an accuracy of 0.05 °C. Samples were equilibrated for 2 min prior to collecting each spectrum. No temperature-induced spectral variations were observed after that time.

2.4. Multivariate Data Analysis

The SciKit Learn package in Python was used for PLSR and PCA. As one of the most frequently used chemometric modeling tools, PLSR is especially useful for regressing spectra where the number of independent variables is significantly larger than the number of samples [15]. Each PLSR model is built using a signal matrix (spectra) and a response matrix (concentrations) and transforms them to a latent space. The matrix is iteratively solved for vectors, or latent variables (LVs), which explain the most covariance between the signal and the response matrices. As an unsupervised dimension-reducing method, PCA identifies orthogonal vectors that explain the variance of the independent variable without knowledge of any dependent variables [18]. A key difference between PLSR and PCA is that the former seeks to explain the most covariance between the signal and response matrices, and the latter focuses solely on the variance of the signal matrix. Simply stated, PCA explores the sample variability in an unsupervised manner whereas PLSR is designed under supervision to predict a given response. Multivariate models were developed using spectra collected on static (i.e., not flowing) samples. UV-vis and NIR spectra were fused at 870 nm, and the combined spectra were baseline adjusted by subtracting the minimum absorbance value between 930 and 935 nm generating a single spectrum ranging between 400–1700 nm. An example spectrum before and after baseline subtraction and fusion is shown in Figure S1. An in-house genetic algorithm (GA) feature selection tool was used, and further details can be found elsewhere [19–22]. Briefly, the GA iteratively forms binary arrays that, when applied to the spectra, act as filters. These filters are compared based on the corresponding model prediction performance. The best filters are retained and are also combined to form new filters, which are then evaluated in the next iteration, simulating natural selection. Overtime an optimal filter will be reached, analogous to the theory of evolution, that only permits the feature(s) required for regression into the model.

2.5. Statistics

The root mean square error (RMSE) was used as the primary metric for prediction error, defined in Equation (1) as

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

where y_i is the known concentration, $\hat{y}_i$ is the model-predicted concentration, and n is the total number of samples. The RMSE of calibration (RMSEC) is determined using the calibration set to train the data; then, the same calibration set is predicted by the model, essentially making RMSEC a measure of fit. The RMSE of cross validation (RMSECV) was determined using a fivefold approach where the calibration set was split into five random groups. Then, one group was left out as the model was built and predicted the values of the left-out group. This approach was an iterative process that proceeded until each sample was left out at least once. The deviation of predicted values was averaged into

a single RMSECV metric. The RMSE of prediction (RMSEP) measures the dispersion of samples not included in a validation set (i.e., never included in model construction) around the regression line. The RMSE values are typically discussed in terms of percentages to ease comparisons [16,20,23]. The RMSE value was divided by the median of the model concentration range to produce percent RMSEP (RMSEP%). For the percentage conversions in this study, the RMSE values were divided by 0.05 for low Np concentrations, 0.40 for high Np concentrations, and 35 for temperature models. Lower RMSE values indicate improved model performance. In the following discussion, models were rated based on their predictive error: strong (RMSEP% $\leq$ 5%), satisfactory (5 < RMSEP% $\leq$ 10%), or indicative (10 < RMSEP% $\leq$ 15%). RMSEP% values above 15% do not offer any monitoring value and are undesirable [23].

The limit of detection (LOD) is a typical benchmark used for quantification models. The International Union of Pure and Applied Chemistry (IUPAC) defines the LOD as the lowest concentration that can be detected with reasonable certainty for a given method [24]. Calculating the LOD for a multivariate model (e.g., PLSR models) is much more involved than for univariate models. This study used a pseudounivariate approach (LOD_{pu}) proposed by Ortiz et al. that extends the IUPAC recommendations for univariate models for use in multivariate models [25]. In this work, the multivariate model was used to estimate the concentrations of the calibration sample set. The slope of these predictions versus the known values (i.e., the slope of the parity plot) was used in place of the univariate calibration curve. The LOD_{pu} was then calculated using Equation (2):

$$LOD_{pu} = 3.3 \frac{\sqrt{\left(1 + h_{0min} + \frac{1}{N}\right) var_{pu}}}{m_{pu}},$$
$$\text{where } h_{0min} = \frac{\bar{y}_{cal}^2}{\sum_{n=1}^{N} y_n^2}, \tag{2}$$

and where m_{pu} is the slope of the produced parity plot, var_{pu} is the variance of the regression residuals, N is the number of calibration samples, $\bar{y}_{cal}$ is the mean concentration of the calibration analytes, and y_n is the centered analyte concentration of sample n.

3. Results and Discussion

3.1. Np Vis-NIR Spectra

The Np spectrum in 1 M HNO_3 is shown as a function of Np concentration in Figure 1a. A large range in molar absorptivity values are represented. The dominating band in the spectrum is centered at 979 nm for Np(V) and has a molar absorptivity of 367 $M^{-1} \cdot cm^{-1}$, which agrees well with published data. Several additional absorption bands appear at 433, 476.5, 616.4, 687.0, 914.2, 1022, 1096, and 1116 nm with molar absorptivity values of 7, 22, 5, 2.5, 9, 25, and 6, respectively [10]. The asymmetric Np(VI) peak is shown in high Np concentration samples centered near 1224 nm, but the quantity is negligible. The peak near 1616 nm is convoluted with the NIR water band response centered near 1450 nm [26]. This water band is related to the first overtone of water. The broad positive (1404 nm) and negative peaks in this region (1490 nm), which result in an isosbestic point near 1440 nm, occur because of differences in the local tetrahedral structure of water owing to its temperature [27].

Several peaks, such as the peak located at 616.4 nm, are complex and confounded by nearly overlapping absorption bands, which are not easily resolved. Figure 1b shows the spectra of a 0.0687 M Np sample in 1 M HNO_3 with temperature varied from 10–80 °C. The effect of temperature is visible in the water band (1300–1650 nm), but the inset plot provides a closer examination of how the Np(V) bands change shape and shift with temperature. These changes may appear small, but they can generate challenges when attempting to apply Beer's Law to monitor Np(V) concentrations. For example, the Np(V) 979 nm band intensity decreases by 12%, and the peak position blue-shifts by 1.9 nm from 10–80 °C [10]. These effects necessitate the use of multivariate modeling to accurately monitor such a system with a wide range of conditions.
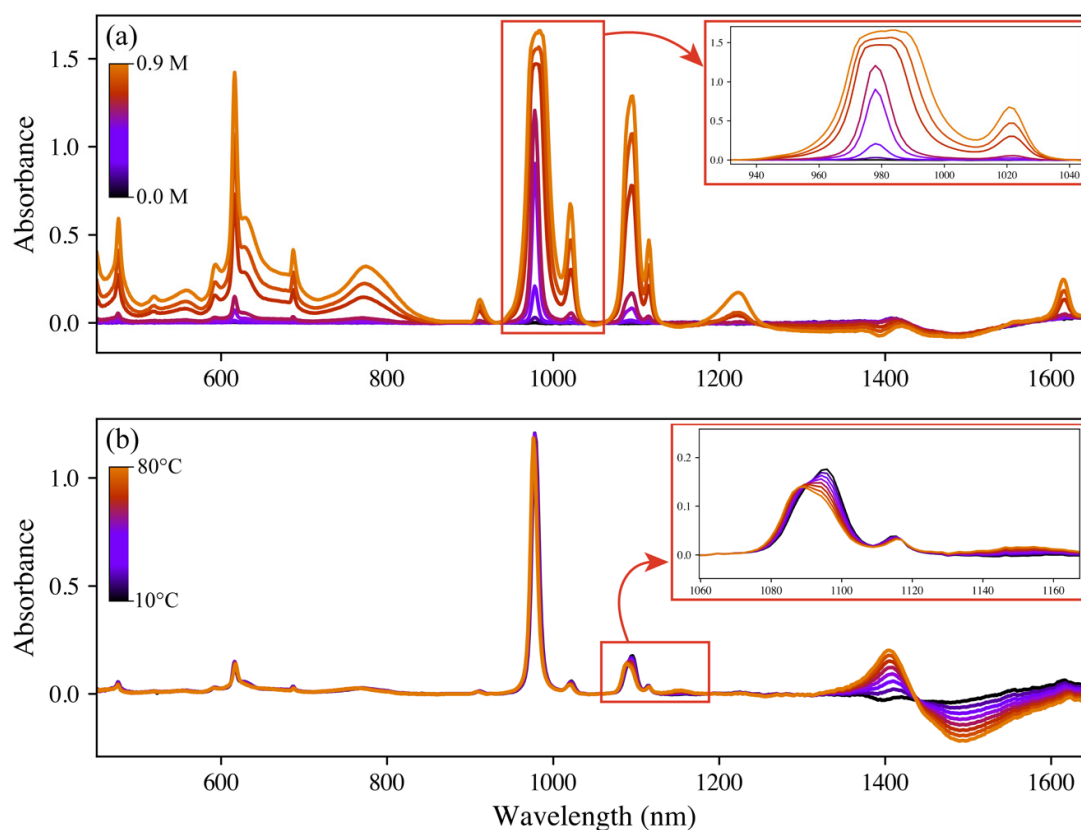
Figure 1. UV-vis-NIR absorption spectrum of (**a**) 0–0.886 M Np in 1 M HNO$_3$ at 20 °C and (**b**) 0.0687 M Np in 1 M HNO$_3$ with temperature varied from 10 °C–80 °C. All samples were blanked in water at 20 °C. The inset plot in (**a**) shows the Np(V) 979 nm peak saturating the detector at higher concentrations. The inset plot in (**b**) highlights the effect of temperature on Np(V) bands. The temperature effects on the water band can be seen in the full plot ($\geq$1300 nm).

A large obstacle encountered when monitoring Np(V) by spectrophotometry in relevant nuclear fuel cycle applications revolves around the 979 nm band, which can easily saturate the detector at higher concentrations. This effect can be seen in the inset of Figure 1a, where high Np(V) concentration samples show a plateaued peak as opposed to the typical Gaussian shape seen in the lower Np(V) concentrations. A smaller optical pathlength cell could be used; however, this would make the detector less sensitive to lower concentrations. Another option is to develop a piecewise model that uses different spectral features depending on the Np concentration. This model would recognize which portions of the spectrum are in the ideal range for quantification and exclude regions that are not (e.g., saturated Np(V) 979 nm peak). For online monitoring applications, the model must discern which submodel to send the collected spectra to without human involvement.

3.2. Principal Component Analysis

In this work, PCA was particularly useful because it identified principal components (PCs) by examining the spectra provided to it without knowledge of Np concentrations or temperatures. With 3 PCs, 99.2% of the signal variance was explained. Through investigating loadings and score plots, the PCs were attributed to specific phenomena.

Figure 2 shows the PCA score plots colored according to the Np concentration and the sample temperature, as well as the corresponding PC loadings to investigate feature correlation. Figure 2a,b show a clear trend between PC-1 and Np concentration and an absence of a clear relation to temperature. Upon looking at the PC-1 and PC-2 loadings, both are related to Np concentration because a single PC cannot explain all the variance from the 979 nm peak saturation. PC-1 does not rely heavily on the peak absorbance value

of the 979 nm band, so it increases nearly linearly with Np concentration. Conversely, PC-2 does have a significant loading on the 979 nm band, so upon the Np concentration growing above approximately 0.1 M, the PC-2 score begins to decrease as this band saturates. This decrease resulted in the arch-shaped trend seen in Figure 2a.
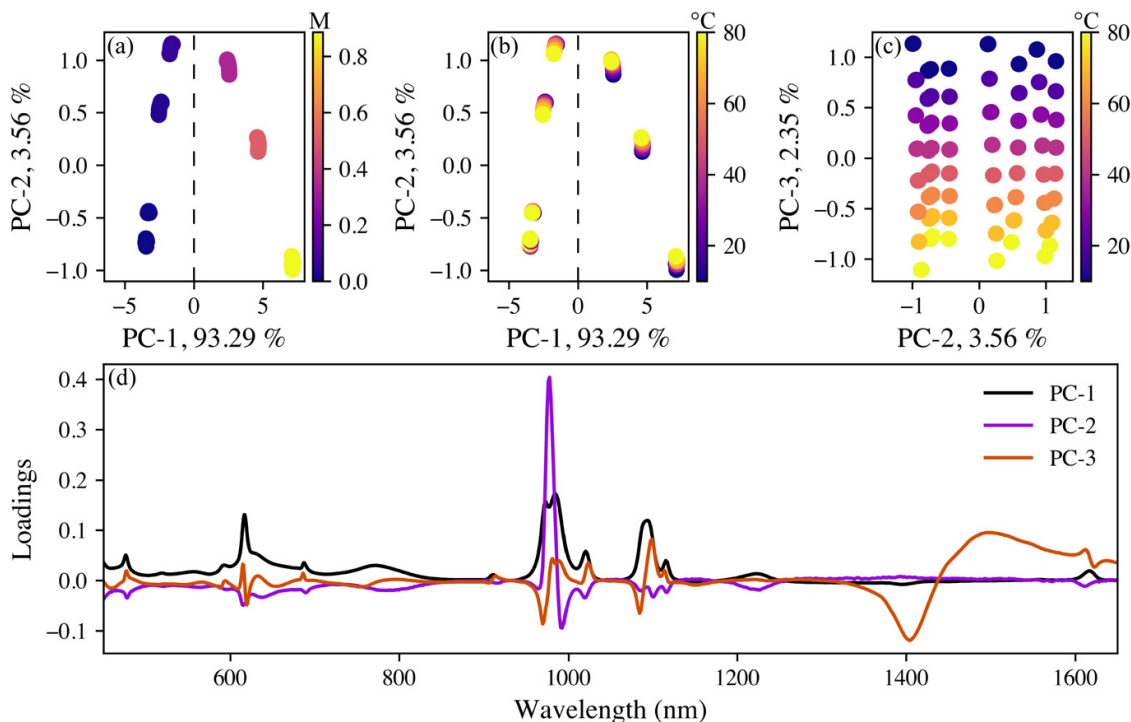

Figure 2. PCA score plots comparing principal component scores: (**a**) PC-1 vs. PC-2 colored by Np concentration; (**b**) PC-1 vs. PC-2 colored by sample temperature; and (**c**) PC-2 vs. PC-3 colored by sample temperature. In (**d**), the principal component weights vs. wavelengths are provided. As seen in (**a**), a positive PC-1 score correlates to a high Np concentration (i.e., $\geq$0.1 M).

Figure 2c shows that despite the complexity associated with explaining the concentration-related signal variance, PC-3 can describe signal changes correlated to temperature. The importance given to the NIR water band in the loadings plot verifies this correlation where the PC-3 scores become more negative as the temperature increases due to the loading being inverted to the trend seen in Figure 1b. The PC-3 loading also points to the Np band shifts as indicators. The PCA model can distinguish low and high Np spectra and temperature effects in an unsupervised approach.

3.3. Hierarchical Modeling

By applying the PC loadings shown in Figure 2d to spectra, the PC-1 scores can be used to direct data to specific submodels (i.e., PC-1 score <0 indicates low Np concentration submodel, and PC-1 score $\geq$0 indicates high Np concentration submodel). The calibration samples were split into low and high Np concentration sets using this procedure and were subsequently used to develop the PLSR submodels. The spectra sent to the high Np concentration submodel were filtered to remove the saturated NIR Np(V) bands; the low Np concentration submodel was built using the full spectrum. A fivefold cross validation was performed to select the required number of LVs for each PLSR model; seven and six LVs were chosen for the low and high Np concentration submodels, respectively (see Figure S2).

Following the LV selection, each submodel was evaluated by predicting the Np concentration and temperature values of a set of validation samples. The corresponding RMSE values for calibration, cross validation, and prediction are provided in Table 2. As seen by the RMSEC% values, both models performed at *strong* levels, but during validation tests,

the RMSEP% values for the high Np concentration submodel increased, pushing the high Np concentration submodel performance to the *satisfactory* level. The differences in the levels of prediction accuracy are shown in the parity plots in Figure 3.

Table 2. RMSE calibration, cross validation, and prediction metrics for the PLSR submodels.

Submodel		C	C%	CV	CV%	P	P%
					RMSE		
Low Np concentration	Conc.	0.0004	0.78%	0.0034	6.89%	0.0004	0.85%
	T	0.1471	0.42%	0.3838	1.10%	0.9892	2.83%
High Np concentration	Conc.	0.0032	0.80%	0.0088	2.20%	0.0220	5.49%
	T	0.1157	0.33%	0.3964	1.13%	1.7967	5.13%

Note: Conc.: concentration, T: temperature.
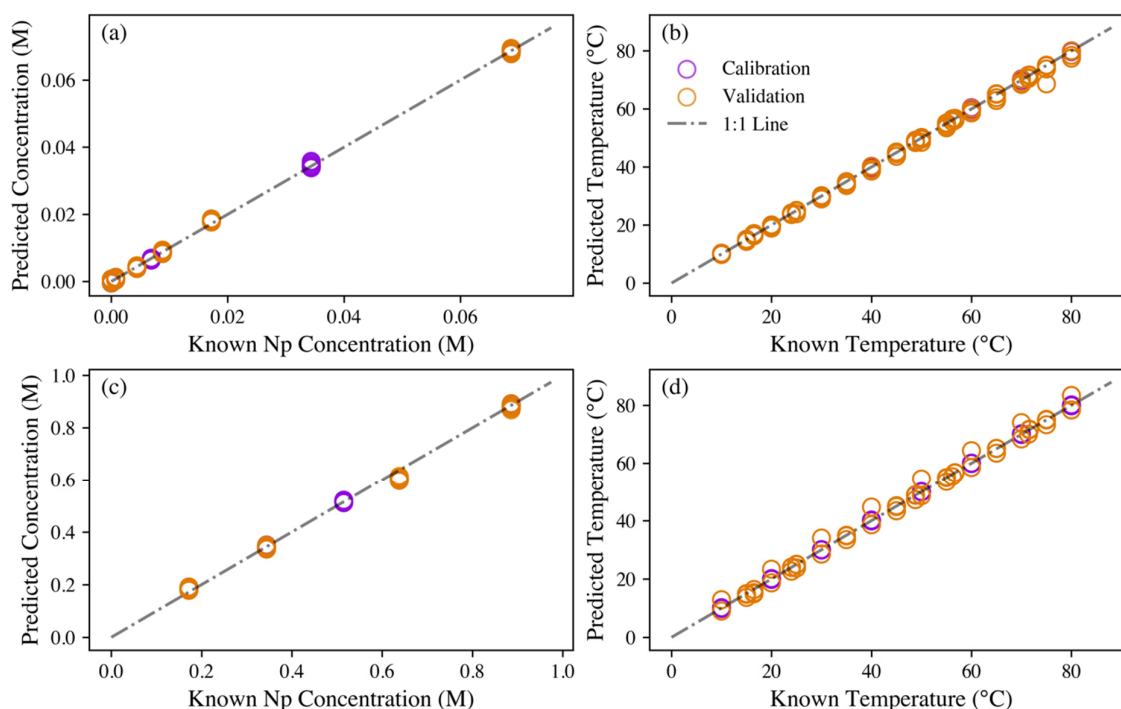
Figure 3. Parity plots comparing known and (**a**,**b**) low Np concentration submodels and (**c**,**d**) high Np concentration submodels' predicted values for Np concentration and temperature. The 1:1 line signifies a perfect prediction; the closer a marker to the 1:1 line, the better the prediction.

The low Np concentration submodel performed well for Np concentration and temperature predictions with only slight variance seen in the validation set temperature predictions. The high Np concentration submodel showed moderate variance around the 1:1 line for Np concentration prediction, but the temperature prediction showed a larger scatter. Inspection of the PLSR regression coefficients (not shown) indicated that both submodels rely heavily on the NIR water band for temperature, but the low Np concentration submodel also significantly weighs the 979 nm band when predicting temperature. This submodel shows some of the most prominent features related to temperature changes. The prominent features explain the slightly reduced temperature prediction accuracy of the high Np concentration submodel because this band is unavailable to be regressed. Although these current models perform adequately, a *strong* performance is desired to meet program requirements. Although the temperature sensitivity of the NIR water band provides a straightforward model, this sensitivity may become detrimental to the model if the absorbance blank is measured slightly off-temperature.

Removing the reliance on the NIR water band would result in a more robust monitoring model. This model is more resistant to differences in blank temperatures, which would be inevitable in process operations. Temperature differences relative to the blank in the water spectrum <1300 nm are either negligible or essentially nonexistent. Weak NIR absorption bands are centered near the 960 and 1190 nm water bands, but no band is evident in the visible region from approximately 400 to 950 nm [26,28]. The absence of these water bands allows the model to measure temperature solely based on Np spectral variations. The previously detailed submodels were reconstructed with the spectrum upper limit set to 1300 nm to not regress the water band features. The result was a slight increase in the high Np concentration submodel RMSEP%, to 6.50%. Both models show a sizable increase in temperature RMSEP% values, to 7.52% and 22.9% for the low and high Np concentration submodels, respectively. The parity plots for these models are shown in Figure S3. These values indicate that rather than simply cropping the spectra before modeling, proper feature selection would be better suited towards filtering the spectra prior to the PLSR modeling, or an alternative model for temperature is needed.

3.4. Feature Selection for Model Optimization

Previous studies have shown that the use of feature selection via a GA can result in significant improvements for PLSR models. In this work, the GA was run on the PCA-split low and high Np concentration calibration sets to develop independent Np concentration and temperature filters to optimize prediction performance. The GA was completed 5 times with 150 generations each, and the optimal filter for each submodel and variable was selected. The changes in explained variance, the number of estimators, and the RMSE record for each of the final GA filters are shown in Figures S4–S7. The GA selected features are shown in Figure 4.
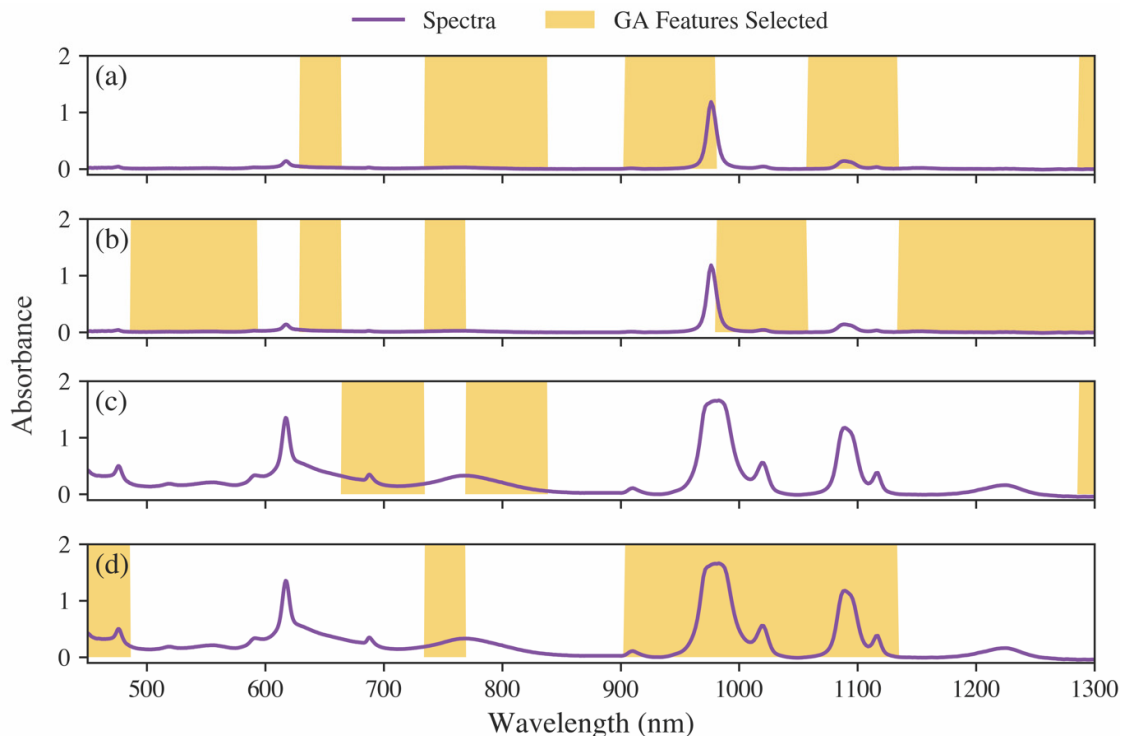

Figure 4. Features selected by the GA for the low Np concentration submodel to regress (**a**) Np concentration and (**b**) temperature, and the high Np concentration submodel to regress (**c**) Np concentration and (**d**) temperature.

Several interesting takeaways can be concluded from the GA filter results. First, the low Np concentration filter selects the 979 nm NIR Np(V) features as expected because this

peak is intense and not saturated in this regime, whereas the high Np concentration filter does not retain this peak due to the saturation issues. Secondly, the temperature filters both select at least portions of the NIR Np(V) bands (979, 1022, and/or 1096 nm), indicating the bands are useful for regressing temperature effects despite saturation—likely due to the peak shift phenomenon seen in Figure 1. This behavior is difficult to model linearly across the Np concentration range; a single PLSR for temperature on the entire calibration set was built but could not properly regress the behavior. Thirdly, the sharp 616 nm band was not selected, although it appeared as a distinguishing feature.

These GA-selected filters improved prediction accuracy with up to an 87% reduction in RMSEP (average percent difference equal to −75%). Parity plots for this new model are shown in Figure 5. The GA filtered models' RMSEP% values were 0.14% and 4.00% for low Np concentration and temperature, respectively, and 1.05% and 2.96% for high Np concentration and temperature, respectively. These results indicated a *strong* level of prediction for each submodel. Because these models are independent of the NIR water band centered near 1450 nm, the high dependency on the blank temperature was minimized.
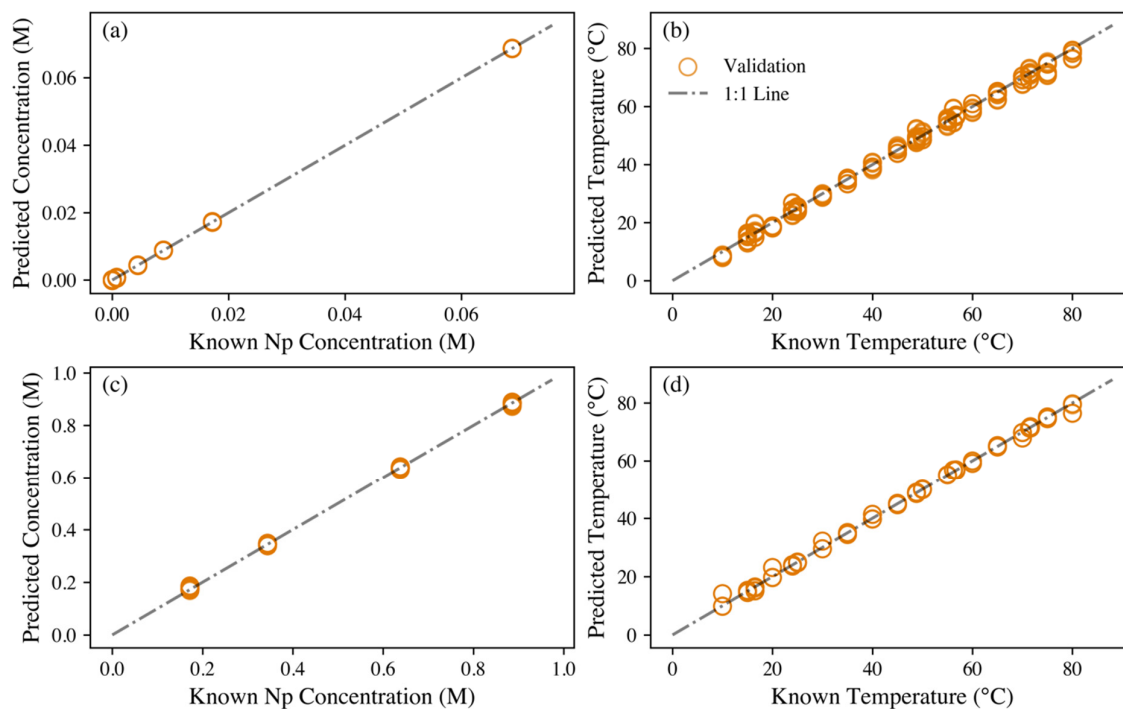

Figure 5. Parity plots comparing known and predicted values for the GA filtered models: (**a**) low Np concentration submodel concentration predictions, (**b**) low Np concentration submodel temperature predictions, (**c**) high Np concentration submodel concentration predictions, and (**d**) high Np concentration submodel temperature predictions. The 1:1 line signifies a perfect prediction; the closer a marker is to the 1:1 line, the better the prediction is.

3.5. Final PLSR Model Evaluation

The final model design is provided in Figure 6. As one final comparison, a global PLSR model was built on the complete calibration data set to evaluate the level of improvement offered through the hierarchical design. The global PLSR model was developed using the same general procedure outlined previously. The final model was built with four LVs because the addition of further LVs caused an increase in RMSECV values. The calibration and validation parity plots are shown in Figure S8. The RMSEP% values for this global PLSR model were 22.06% and 49.79% for Np concentration and temperature, respectively, at Np concentrations below 0.1 M and 6.48% and 27.46% for Np concentration and temperature, respectively, at Np concentrations above 0.1 M. This poor performance, particularly with low Np concentration samples (see inset plot in Figure S8), is a direct

result of the previously discussed issues; the PLSR model cannot cope with the nonlinearity of the saturating 979 nm peak and uses the vis Np(V) bands primarily for quantification. Unfortunately, these peaks are very weak at low Np concentrations, so the model suffers. Regarding temperature, improved prediction accuracy was seen in the hierarchical model following the GA feature selection process. This comparison exemplifies the superior performance of the hierarchical model with feature selection.
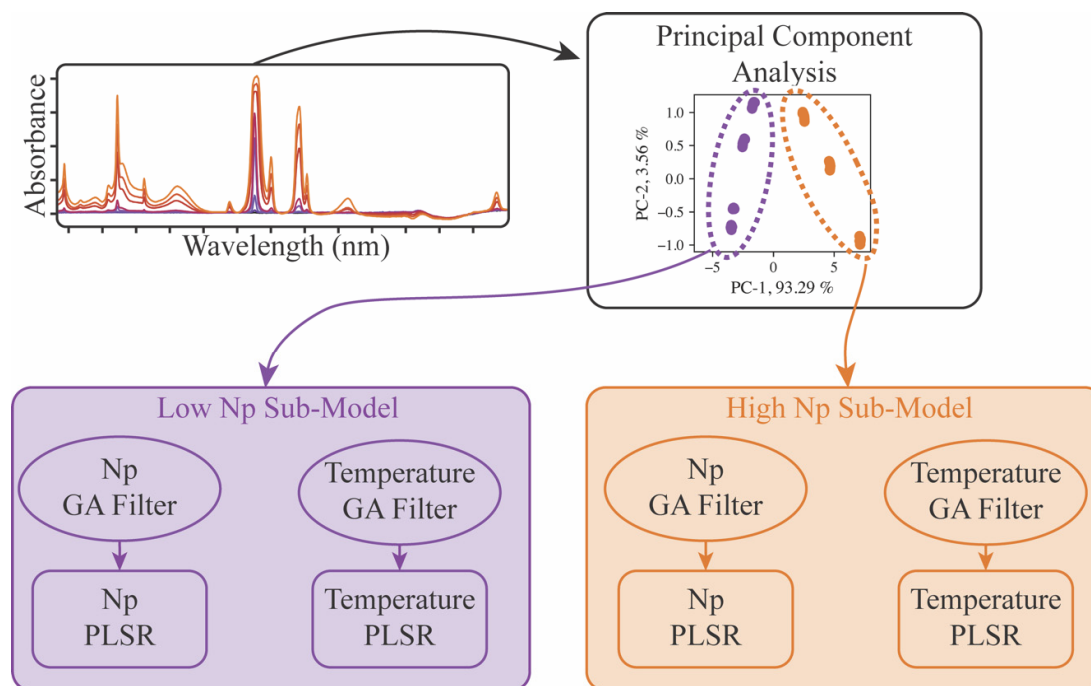

Figure 6. Overview of final hierarchical model design. Each GA filter prior to PLSR modeling is unique.

The LOD_{pu} was calculated for the final low Np concentration submodel to evaluate the lowest concentration that the overall hierarchical model would be able to predict with reasonable sensitivity. Using Equation (2), the LOD_{pu} was 0.232 mM Np. To measure lower concentrations, an alternative measurement system (e.g., larger pathlength or a more sensitive spectrometer) would be needed, but a hierarchical model similar to the one discussed in this work (Figure 6) could describe a lower concentration range.

4. Conclusions

This study demonstrated the design of a hierarchical model to quantify a wide range of Np concentrations (0.00075–0.89 M) and varying temperatures (10–80 °C) relevant to remote monitoring systems in a single model and without the need to complicate the experimental setup. This approach takes advantage of multiple Np(V) absorption bands with molar absorption coefficients covering two orders of magnitude. This new approach pushes the burden onto the data analytics, which is relatively cheap and efficient, and helps minimize the equipment footprint. Results indicated that spectrophotometry and multivariate chemometrics have great potential to improve the timeliness and performance of important aqueous Np processes—for instance, Np cation exchange column runs for the Oak Ridge National Laboratory [238]Pu Supply Program. Future work will explore quantifying multiple Np oxidation states in the regression model.

large noise fluctuations; Figure S2: Cross validation latent variable selection for the (a) low–Np concentration submodel and (b) high–Np concentration submodel. RMSECV is shown in percent value for better visualization. LVs selected are circled; Figure S3: Parity plots comparing known and predicted values for Np concentration and temperature: (a and b) low–Np concentration submodel and (c and d) high–Np concentration sub-model (wavelengths <1300 nm). The 1:1 line signifies a perfect prediction; the closer a marker is to the 1:1 line, the better the prediction is; Figure S4: Generation record for the low–Np concentration submodel Np concentration GA filter: (top) change in explained variance, (middle) change in the number of estimators, and (bottom) change in the best and average RMSE vs. generation; Figure S5: Generation record for the low–Np concentration submodel temperature GA filter: (top) change in explained variance, (middle) change in the number of estimators, and (bottom) change in the best and average RMSE vs. generation; Figure S6: Generation record for the high–Np concentration submodel Np concentration GA filter: (top) change in explained variance, (middle) change in the number of estimators, and (bottom) change in the best and average RMSE vs. generation; Figure S7: Generation record for the high–Np concentration submodel temperature GA filter: (top) change in explained variance, (middle) change in the number of estimators, and (bottom) change in the best and average RMSE vs. generation; Figure S8: Global PLSR model predictions vs. known values for (a) Np concentration and (b) temperature. The Np concentration predictions appear to be fit well around the 1:1 line, but the inset figure shows that at low Np concentrations, the model performs inadequately.

Author Contributions: The manuscript was written through contributions of all authors. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available upon reasonable request to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Harris, D.C. Chapter 18: Fundamentals of Spectrophotometry. In *Quantitative Chemical Analysis*, 7th ed.; W. H. Freeman and Company: New York, NY, USA, 2007.
2. Sadergaski, L.R.; Myhre, K.G.; Delmar, L.H. Multivariate chemometric methods and Vis-NIR spectrophotometry for monitoring plutonium-238 anion exchange column effluent in a radiochemical hot cell. *Talanta Open* **2022**, *5*, 1000120. [CrossRef]
3. Sadergaski, L.R.; DePaoli, D.W.; Myhre, K.G. Monitoring the caustic dissolution of aluminum in a hot cell by Raman spectroscopy. *Appl. Spectrosc.* **2020**, *74*, 1252–1262. [CrossRef] [PubMed]
4. Kirsanov, D.; Rudnitskaya, A.; Legin, A.; Babain, V. UV-Vis spectroscopy with chemometric data treatment: An option for on-line control in nuclear industry. *J. Radioanal. Nucl. Chem.* **2017**, *312*, 461–470. [CrossRef]
5. Lascola, R.; O'Rourke, P.E.; Kyser, E.A. A Piecewise Local Partial Least Squares (PLS) Method for the Quantitative Analysis of Plutonium Nitrate Solutions. *Appl. Spectrosc.* **2017**, *71*, 2579–2594. [CrossRef]
6. Ikedo-Ohno, A.; Hennig, C.; Rossberg, A.; Funke, H.; Scheinost, A.C.; Bernhard, G.; Yaita, T. Electrochemical and Complexation Behavior of Neptunium in Aqueous Perchlorate and Nitrate Solutions. *Inorg. Chem.* **2008**, *47*, 8294–8305. [CrossRef] [PubMed]
7. Matsika, S.; Pitzer, R.M.; Reed, D.T. Intensities in the Spectra of Actinyl Ions. *J. Phys. Chem. A* **2000**, *104*, 11983–11992. [CrossRef]
8. Matsika, S.; Pitzer, R.M. Electronic Spectrum of the NpO_2^{2+} and NpO_2^+ Ions. *J. Phys. Chem. A* **2000**, *104*, 4064–4068. [CrossRef]

9. Eisenstein, J.C.; Pryce, M.H.L. Interpretation of the Solution Absorption Spectra of the $(PuO_2)^{++}$ and $(NpO_2)^+$ Ions. *J. Res. Natl. Bur. Stand. A* **1966**, *70*, 165–173. [CrossRef]
10. Sadergaski, L.R.; Morgan, K. Applying Two-Dimensional Correlation Spectroscopy and Principal Component Analysis to Understand How Temperatures Affects the Neptunium(V) Absorption Spectrum. *Chemosensors* **2022**, *10*, 475. [CrossRef]
11. Maiwald, M.M.; Skerencak-Frech, A.; Panak, P.J. The complexation and thermodynamics of neptunium(V) with acetate in aqueous solution. *New J. Chem.* **2018**, *42*, 7796–7802. [CrossRef]
12. Maiwald, M.M.; Sittel, T.; Fellhauer, D.; Skerencak-Frech, A.; Panak, P.J. Thermodynamics of neptunium(V) complexation with sulfate in aqueous solution. *J. Chem. Thermodyn.* **2018**, *116*, 309–315. [CrossRef]
13. Chatterjee, S.; Bryan, S.A.; Casella, A.J.; Peterson, J.M.; Levitskaia, T.G. Mechanisms of neptunium redox reactions in nitric acid solutions. *Inorg. Chem. Front.* **2017**, *4*, 581–594. [CrossRef]
14. Edelstein, N.M. Reanalysis of the Aqueous Spectrum of the Neptunyl(V) $[NpO_2^+]$ Ion. *J. Phys. Chem. A* **2015**, *119*, 11146–11153. [CrossRef] [PubMed]
15. Dupont, F.M.; Elbourne, A.; Cozzolino, D.; Chapman, J.; Truong, V.K.; Crawford, R.J.; Latham, K. Chemometrics for environmental monitoring: A review. *Anal. Methods* **2020**, *12*, 4597–4620. [CrossRef]
16. Sadergaski, L.R.; Andrews, H.B. Simultaneous quantification of uranium(VI), samarium, nitric acid, and temperature with combined ensemble learning, laser fluorescence, and Raman scattering for real-time monitoring. *Analyst* **2022**, *147*, 4014–4025. [CrossRef]
17. Ban, Y.; Hakamatsuka, Y.; Tsutsui, N.; Urabe, S.; Hagiya, H.; Matsumura, T. Spectroscopic study of Np(V) oxidation to Np(VI) in 3 mol/dm^3 nitric acid at elevated temperatures. *Radiochim. Acta* **2014**, *102*, 775–780. [CrossRef]
18. Bro, R.; Smilde, A.K. Principal component analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [CrossRef]
19. Andrews, H.B.; Myhre, K.G. Quantification of lanthanides in a molten salt reactor surrogate off-gas stream using laser-induced breakdown spectroscopy. *Appl. Spectrosc.* **2022**, *76*, 877–886. [CrossRef]
20. Andrews, H.B.; Sadergaski, L.R.; Cary, S.K. Pursuit of the Ultimate Regression Model for Samarium(III), Europium(III), and LiCl using Laser-Induced Fluorescence, Design of Experiments, and a Genetic Algorithm for Feature Selection. *ACS Omega* **2023**, *8*, 2281–2290. [CrossRef]
21. Leardi, R.; Boggia, R.; Terrile, M. Genetic algorithms as a strategy for feature selection. *J. Chemom.* **1992**, *6*, 267–281. [CrossRef]
22. Leardi, R.; Gonzalez, A.L. Extraction of representative subsets by potential functions method and genetic algorithms. *J. Chemom. Intell. Lab. Syst.* **1998**, *40*, 33–52.
23. Andrews, H.B.; Sadergaski, L.R. Leveraging visible and near-infrared spectroelectrochemistry to calibrate a robust model for Vanadium (IV/V) in varying nitric acid and temperature levels. *Talanta* **2023**, *259*, 124554. [CrossRef]
24. McNaught, A.D.; Wilkinson, A. (Eds.) *IUPAC Compendium of Chemical Terminology (the "Gold Book")*, 2nd ed.; Blackwell Scientific Publications: Oxford, UK, 1997.
25. Ortiz, M.C.; Sarabia, L.A.; Herrero, A.; Sánchez, M.S.; Sanz, M.B.; Giménez, D.; Meléndez, M.E. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. *Chemom. Intell. Lab. Syst.* **2003**, *69*, 21–33. [CrossRef]
26. Toney, G.K.; Delmau, L.H.; Myhre, K.G. Chemometrics and Experimental Design for the Quantification of Nitrate Salts in Nitric Acid: Near-Infrared Spectroscopy Absorption Analysis. *Appl. Spectrosc.* **2021**, *75*, 1155–1167.
27. Segtnan, V.H.; Sasic, S.; Isaksson, T.; Ozaki, Y. Studies on the Structure of Water Using Two-Dimensional Near-Infrared Correlation Spectroscopy and Principal Component Analysis. *Anal. Chem.* **2001**, *73*, 3153–3161. [CrossRef] [PubMed]
28. Chang, K.; Shinzawa, H.; Chung, H. Concentration determination of inorganic acids that do not absorb near-infrared (NIR) radiation through recognizing perturbed NIR water bands by them and investigation of accuracy dependency on their acidities. *Microchem. J.* **2018**, *139*, 443–449. [CrossRef]