

Detection and Quantification of Bisphenol A in Surface Water Using Absorbance-Transmittance and Fluorescence Excitation-Emission Matrices (A-TEEM) Coupled with Multi-Way Techniques

Thomas Ingwani , [Nhamo Chaukura](#) , [Bhekie B Mamba](#) , [Thabo Nkambule](#) ^{*} , Adam Gilmore

Posted Date: 16 August 2023

doi: 10.20944/preprints202308.1151.v1

Keywords: method development; optimisation; and validation; parallel factor modelling; partial least squares modelling



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Detection and Quantification of Bisphenol A in Surface Water Using Absorbance-Transmittance and Fluorescence Excitation-Emission Matrices (A-TEEM) Coupled with Multi-Way Techniques

Thomas Ingwani ¹, Nhamo Chaukura ^{1,2}, Bhekhe B. Mamba ¹, Thabo T.I. Nkambule ^{1,*} and Adam M. Gilmore ^{1,3}

¹ Institute for Nanotechnology and Water Sustainability, College of Engineering, Science and Technology, University of South Africa, Johannesburg, 1709, South Africa

² Department of Physical and Earth Sciences, Sol Plaatje University, Kimberley, 8300, South Africa

³ Horiba Instruments Incorporated, Piscataway, New Jersey, 08854, USA.

* Correspondence: nkambtt@unisa.ac.za, Tel.: +27116709481

Abstract: In the present protocol, we determined the presence and concentrations of bisphenol A (BPA) spiked in surface water samples using EEM fluorescence spectroscopy in conjunction with modelling using partial least squares (PLS) and parallel factor (PARAFAC). PARAFAC modelling of the EEM fluorescence data obtained from surface water samples contaminated with BPA unraveled four fluorophores including BPA. The best outcomes for BPA concentration ($R^2 = 0.996$; Standard deviation to prediction error's root mean square ratio (RPD) = 3.41; and a Pearson's r value of 0.998). With these values of R^2 and Pearson's r , the PLS model showed a strong correlation between the predicted and measured BPA concentrations. The detection and quantification limits of the methods were 3.512 and 11.708 micro molar (μM), respectively. In conclusion, BPA can be precisely detected and its concentration in surface water predicted using the PARAFAC and PLS models developed in this study and fluorescence EEM data collected from BPA-contaminated water. It is necessary to spatially relate surface water contamination data with other datasets in order to connect drinking water quality issues with health, environmental restoration, and environmental justice concerns.

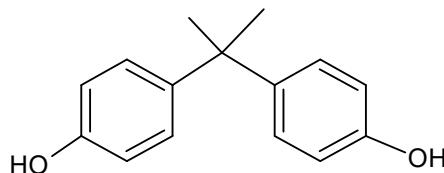
Keywords: method development; optimisation; and validation; parallel factor modelling; partial least squares modelling

1. Introduction

Bisphenol A (BPA) (Table 1), which belongs to the alkyphenol homologous series, is used in the production of polycarbonate plastics and epoxy resins [1,2] which are used to make hard plastic items like storage containers [1] and as linings to cover the interiors of metal products [2]. The aquatic, terrestrial, and atmospheric environments are all impacted by BPA due to its widespread use [2]. Using consumer products that contain BPA can also contaminate the environment and food. Given that BPA is a compound that disrupts the endocrine system and is toxic to the reproductive, developmental, and systemic systems [3] its level in aquatic systems need to be tracked and monitored. Hence, the need for cost-effective and simple-to-use technique to characterise in water.

Table 1. Bisphenol A chemical information.

Chemical structure



Molecular formula, molecular weight	C ₁₅ H ₁₆ O ₂ , 228.291
CAS number	80-05-7

High-performance liquid chromatography (HPLC) with various detection methods (e.g., UV, MS, fluorescence) [4], gas chromatography (GC) in combination with MS [5], electrochemical approaches [6], micellar electrokinetic chromatography [7,8], and adsorptive cathodic stripping voltammetry [9] methods have been previously used to analyse BPA. Although these techniques can offer further insights into the characteristics of BPA, they involve comprehensive sample pretreatment, which makes them costly, labour-intensive, and time-consuming [10]. It is therefore necessary to use a fluorescence excitation matrices (FEEM) method that has a short run time [11], is straightforward to use, is cost-effective [12], and does not require the use of extraction or reagents, such as the simultaneous absorbance and excitation-emission matrices (A-TEEM) in combination with multivariate analysis techniques. Based on a review of the literature and the distribution of species sensitivities, it was found that the measured concentration of BPA in surface water was 273 μM [13].

Due to the lack of specificity in fluorescence spectroscopy, the trilinear parallel factor (PARAFAC) model was used to detect the presence of BPA, and the partial least squares (PLS) model was used to predict the concentration of BPA in spike surface water validation samples. Since the convolution of different component fingerprints in a multi-component mix containing environmental samples results in complicated, difficult-to-interpret EEM spectra, fluorescence spectrometers and multivariate analyses are commonly used together [14].

The PARAFAC model can be used to interpret fluorescence EEMs at various excitation and emission wavelengths [15]. When the correct number of components for a PARAFAC model is specified, PARAFAC can resolve the correct emission and excitation spectra of each constituent in a mixture. It is necessary that the excitation and emission spectra be independent, linear, and that the concentrations of the analytes vary independently [14]. The component concentration scores, emission loadings, and excitation loadings obtained during PARAFAC modelling were used for the classification of components in the surface water samples spiked with BPA.

The PLS model can be used to apply linear regression in order to compare the measured and predicted analyte concentrations (X-block and Y-block data, respectively), as well as to classify EEM data [16,17].

Due to the difficulty of conducting laboratory experiments on surface water that has not been spiked and the highly variable composition of surface water as well as the fact that alkylphenols are known to be unstable in water [18], the analyte is added to surface water samples during an analysis. In order to account for matrix effects that have an impact on the analytical response, matrix matching is used in the analysis. Furthermore, the accuracy of tagging components with fingerprints is increased by analysing both the CDOM components and the BPA standards in the same solutions.

Even though fluorescence spectroscopy coupled with PARAFAC and PLS algorithms can be used to analyse BPA in water, it is still challenging to quantify BPA due to its varying fluorescence intensities and complex underlying interference in water. For instance, if BPA is found in aqueous solutions at concentrations greater than 10 μM , there may be cause for concern due to the interference that lies beneath [19]. BPA has a much lower fluorescence quantum yield of 0.002 [19], in comparison to many DOM components with fluorescence quantum yields ranging between 0.008 and 0.016 [20]. Thus, BPA is less fluorescent active than many DOM components. Many of these DOM components have surface water DOM concentrations of 1 to 20 ppm [11].

In the present study, for the quick (run time: less than 4 min), extraction- and reagent-free detection and quantitation of BPA in common surface water, we tested a possibly sensitive method at lower micro mole concentration levels. The technique should be used to measure the BPA levels in water due to how simple and inexpensive it is to use.

2. Results

The A-TEEM-PARAFAC-PLS analytical method had a short run time (< 4 minutes), high sensitivity at lower micro mole levels for BPA, required little to no sample preparation, and was easy to use. This makes method effective at monitoring BPA contamination in water.

2.1. Absorbance, excitation, and emission spectra

The absorbance spectra showed three peaks at 206, 215, and 275 nm (Figure 1a), indicating the presence of BPA introduced into the sample. As a result of the addition of BPA, the excitation spectra of the surface water sample revealed two peaks at excitation wavelengths of 203 and 218 nm (Figure 1b), which is consistent with the excitation wavelength of 203 nm at which BPA was detected in a previous study [21]. Figure 1c demonstrates that the BPA emission peak happens at a wavelength of 306 nm. The maximum BPA emission wavelength is in agreement with an earlier study [21].

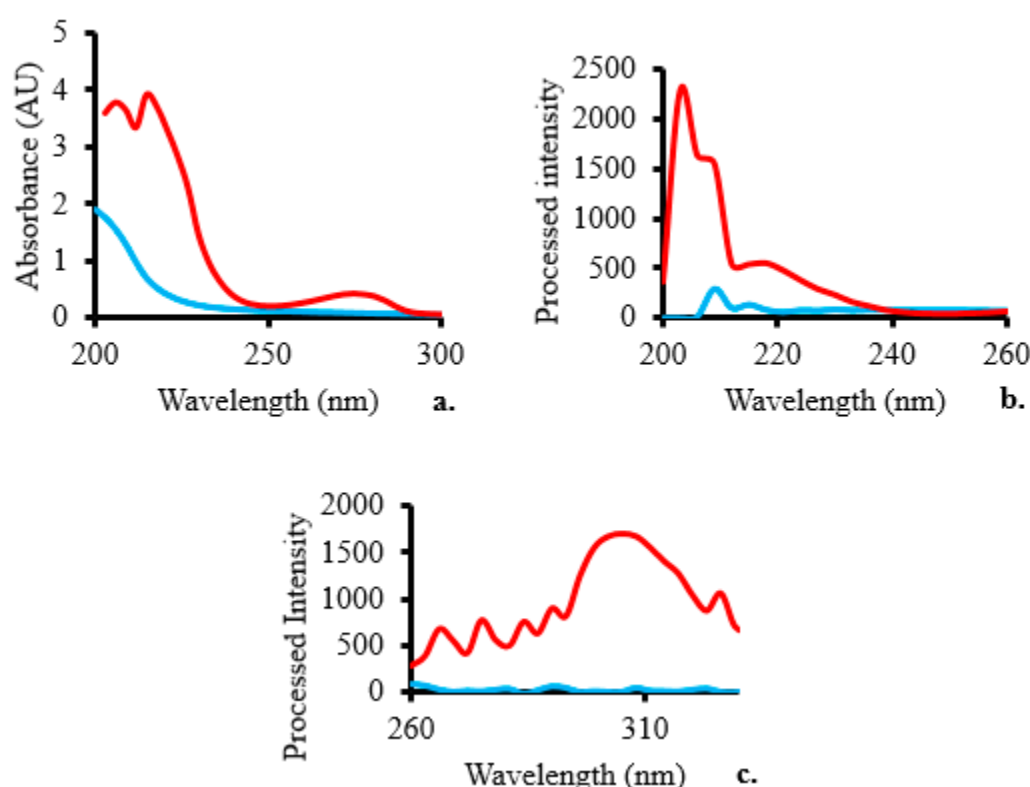


Figure 1. Comparing the baseline (blue line) and a surface water sample spiked with 100 μM BPA (red line) in terms of their respective (a) absorbance, (b) excitation, and (c) emission spectra.

2.2. Excitation-emission matrix signature for BPA

Figure 2a shows the fingerprint EEM of an unspiked (blank/baseline) surface water sample. The unspiked surface water sample showed fluorophores at 230-425 nm for the excitation wavelength and 325-600 nm for the emission wavelength, which indicated the presence of humic substances [22,23]. Figure 2b shows the fingerprint EEM of a surface water sample that had been spiked with BPA at a concentration of 21 μM . In the current study, BPA was found to have excitation and emission wavelengths ranging from 220 to 290 nm and 290 to 375 nm, respectively. These wavelengths match those at which BPA was detected in a previous study [24]. Low-concentration or low-quantum-yield fluorophores may not have visible peaks due to the dominating response of other fluorophores. This is especially true when the dominant wavelength regions coincide with the excitation and emission maxima of the minor ones. Therefore, it's possible that more than one fluorophore had an effect on the fluorescence spectra that were obtained here.

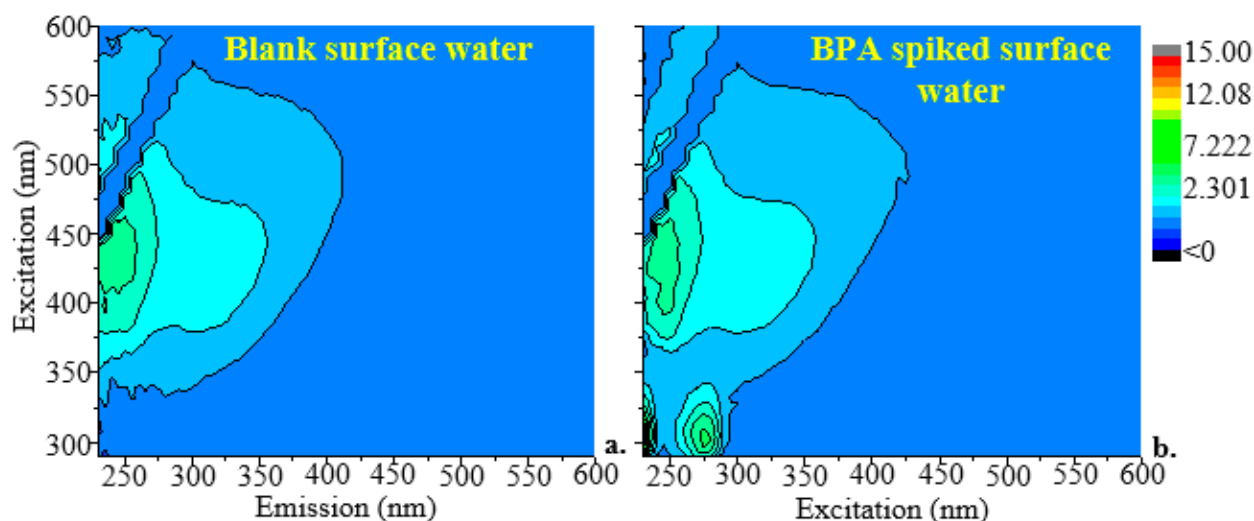


Figure 2: 2D excitation-emission matrices of (a) an unspiked surface water sample with DOC = 2.7 ppm and (b) a sample of surface water that had 21 μM of BPA added to it.

2.3. Construction and validation of the PARAFAC model

The fact that the captured variance obtained in this study using up to four components was higher than that obtained using more than four components shows that four components was the ideal number (Figure 3). An underfitted model typically has a low captured variance [25]. However, if there are too many components used, the model will be overfit. In this situation, the obtained model will capture nearly all of the variance [26]. The percent captured variance with three components can also be used with this type of data [27].

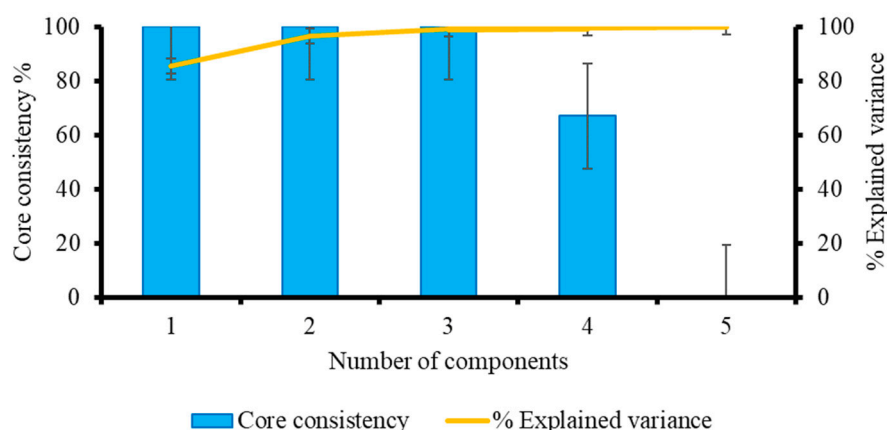


Figure 3. Combined plots for core consistency diagnostic values and explained variance for the non-negativity-constrained PARAFAC model having one to five components for the BPA dataset.

The core consistency for solutions based on a one-component model started out high (at 100%), but it abruptly decreased once the fifth component was fitted. The four components matched the number in the largest model that maintained a large core consistency [25]. The four-component model was the best fit for the dataset and generated the highest resolution of EEM spectra of components (Figure 3) for the 120 EEM dataset, with core consistency diagnostic scores for the one, two, three, four, and five components, respectively, of 100, 100, 100, 67, and <0. The core consistency value is almost 100% if the PARAFAC model is valid. Core consistency will be negative (or even very close to zero) if a trilinear model is unable to adequately describe the data or if there are too many components being fitted [28].

The results of the split-half analysis for a four-component model showed a similarity of 98.8% matching for both distinct splits, indicating that almost no difference could be made between the constituents of each separate split set. (Figure 4). At this stage, the results showed that a four-component PARAFAC model performed the best in terms of core consistency, percent explained variance, and similarity between split-half analyses.

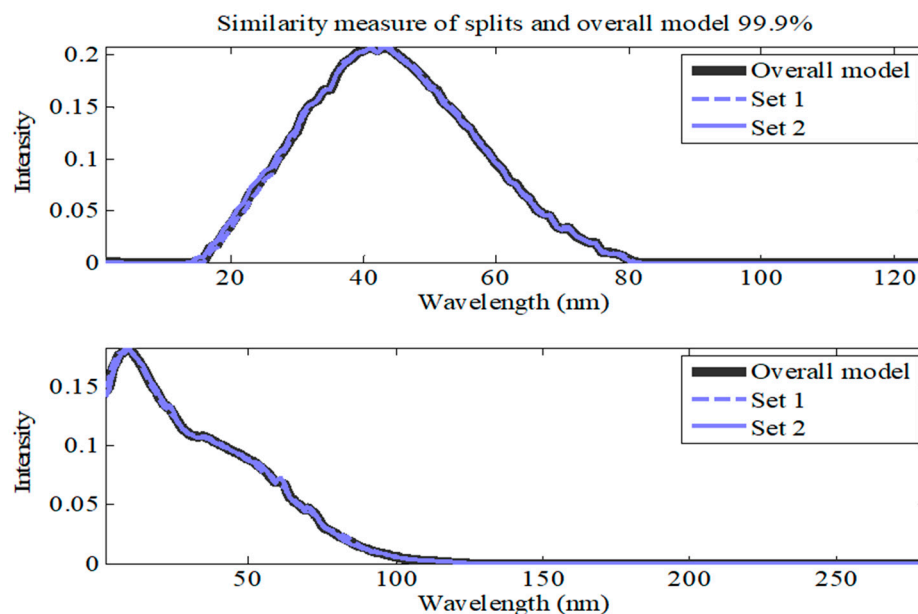


Figure 4. The PARAFAC model's split-half analysis results for surface water samples spiked with BPA at concentrations from 3 to 300 μM .

The 2-D EEM contour plots for the four components in surface water samples spiked with BPA based on PARAFAC modelling of the EEM spectral data are shown in Figure 5. Component #1 was an organic material that resembled fulvic acid-like organic matter, with maximum Ex/Em wavelengths of $\sim 260\text{--}325/440$ nm (Figure 5a). Component #2, with maximum Ex/Em wavelengths of $265\text{--}375/500$ nm, respectively, resembled humic acid-like organic matter (Figure 5b). Component #3 (Figure 5c), with maximum Ex/Em wavelengths of $\sim 275/306$ nm, respectively, is a representation of BPA. Component #4 resembled organic matter that looked like marine humic-like organic material and had maximum Ex/Em wavelengths of $\sim 350/430$ nm (Figure 5d).

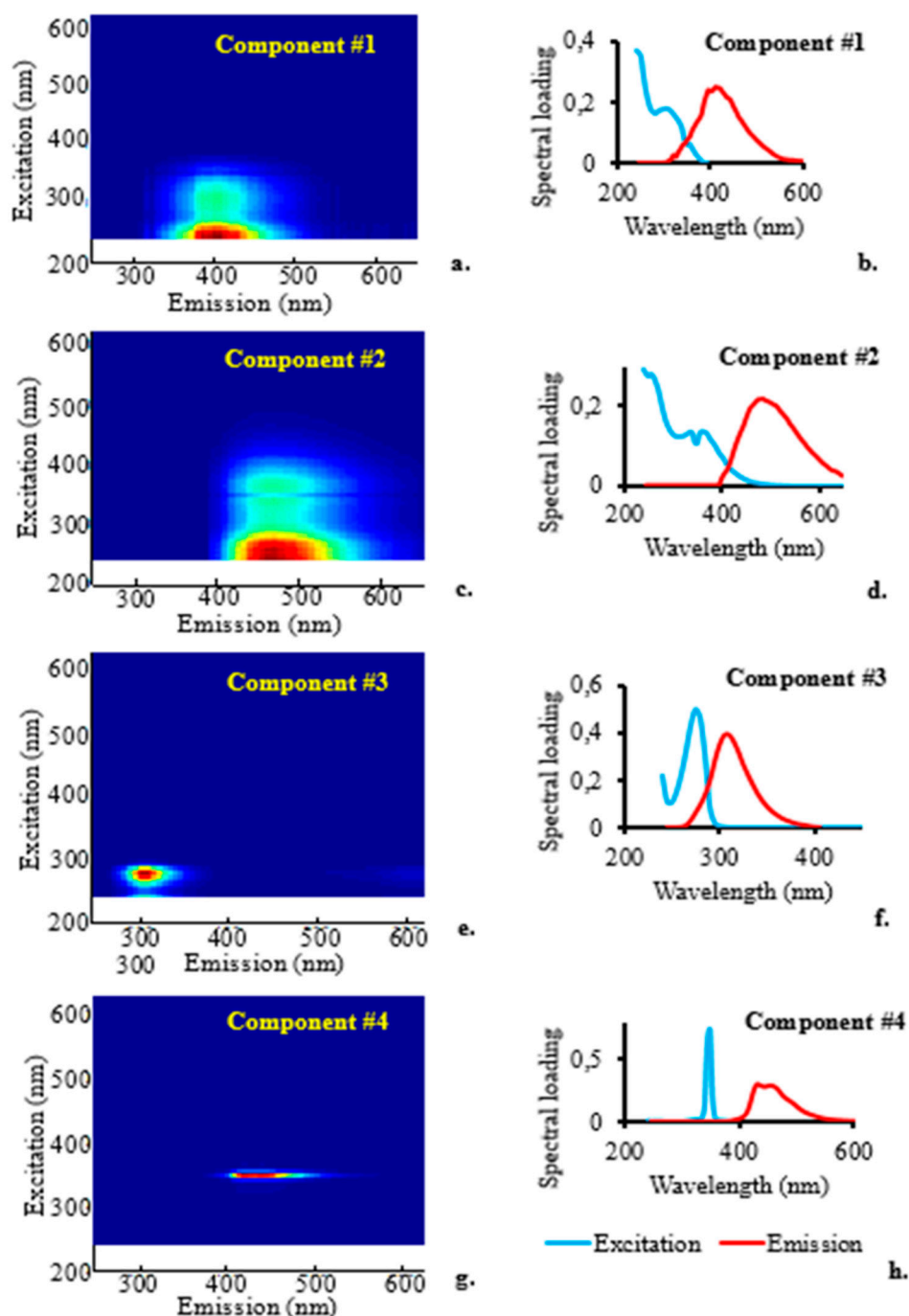


Figure 5. (a) Component #1: A 2-D contour plot of the EEM typical of the fulvic acid-like organic matter; (c) Component #2: A 2-D contour plot of the EEM typical of the humic acid-like fraction of organic matter; (e) Component #3: A 2-D contour plot of the EEM typical of the BPA; and (g) Component #4: A 2-D contour plot of the EEM typical of the marine humic-like fraction of organic matter. Excitation and emission spectral loadings of the four-components obtained by PARAFAC modelling of fluorescence spectral dataset of surface water containing BPA are also shown.

Figure 5 also displays line plots for the four components of the excitation and emission spectral loadings determined by PARAFAC analysis. The correctly validated PARAFAC spectral loadings of excitation and emission represent measures of the pure analyte spectra when the fluorescence data are trilinear and the appropriate component numbers are applied [27]. These loadings were determined by PARAFAC based solely on the excitation and emission spectra of the fluorophoric components in the investigated solutions. Each fluorophoric component was represented by a single PARAFAC component, which was made up of PARAFAC scores for relative concentrations, an

excitation loading related to the estimation of the pure excitation spectrum, and an emission loading related to measurement of the plain emission spectrum. The model was reliable since the spectra were smooth and nearly identical [29]. The smooth, comparable, wide, and frequently unimodal emission loading spectra demonstrated the PARAFAC model's viability [27,29]. On the other hand, it appears that the model was unsuccessful in identifying the excitation spectra of the pure chemical component for component #4, as indicated by the rather sharp excitation peak for that component at about 350 nm [27]. When the EEM data were analysed using the four approaches, the four-component model repeatedly reached the same conclusion.

The emission wavelengths for fulvic acid- and humic acid-like components matched those of components #1 and #2, respectively, which had previously been discovered to emit light at its maximum intensity at wavelengths of 440 and 500 nm [22,30,31]. The strongest emission was previously observed at 306 nm for BPA [21].

2.4. Construction of the PLS model

The fluorescence spectra of 120 surface water samples containing BPA at concentrations ranging from 3 to 300 μM are shown in Figure 6a. On the basis of the visual interpretation of the peaks, the figure depicts one region with a maximum emission at about 306 nm. This maximum emission wavelength is typical of BPA and agrees with an earlier study [32]. Due to overlapping emissions from various substances present in the samples, self-absorption, energy transfer, and quenching phenomena, the spiked surface water samples displayed complex fluorescence properties as multicomponent solutions [33].

Three modes make up rPLS, namely: specified, suggested, and surveyed. The RMSECV values of 10.7, 10.8, and 11.8 μM for the specified, suggested, and surveyed modes, respectively, for the selected variables, show that the cross-validation errors were minimal. The specified mode was the default mode, which constructed the PLS model exclusively using the specified number of LVs and number of components. The suggested mode ran the PLS and cross-validation on the entire dataset, determining the most appropriate number of LVs, and then proceeded with the rPLS as in the specified mode. The PLS was run in the surveyed mode from 1 LV to the most possible LVs, and the set of results with the lowest RMSECV value was returned. Figure 6b shows how the RMSECV (blue curve) and RMSEC (green curve) parameters changed depending on how many LVs were used to create the prediction model. The residual errors for the samples used to train and validate the model were measured using the RMSEC, which in the current study had a typical value of 17.434 μM .

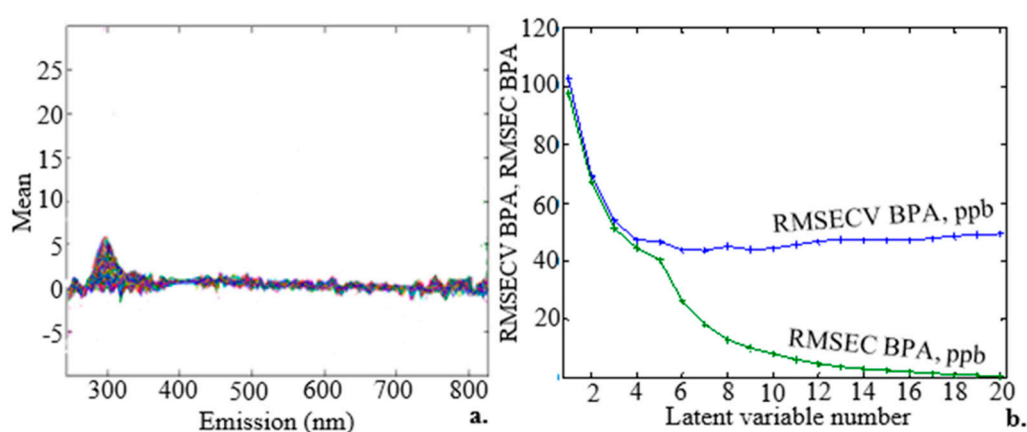


Figure 6. (a) Emission spectra for 120 samples of surface water spiked with BPA standards at various but evenly spaced concentrations ranging from 3 to 300 μM (b) The RMSECV (blue curve) and RMSEC (black curve) were created based on the quantity of LVs used to create the prediction model for the 120 EEM spectral data points of BPA. It was possible to get at least 5 LVs.

Figure 7a plots the predicted versus measured BPA concentration using 120 EEM spectra. The high value of the calibration coefficient of determination ($R^2 = 0.96$) demonstrated the significant

correlation between the predicted and measured BPA concentrations. The scores on LV1 versus the scores on LV2 plot (Figure 7b) revealed four outliers. According to the plot of Hotelling's T2 statistic versus Q residuals reduced (Figure 7c), scores were able to explain 99.35% of the total variance while residuals still held onto 0.6% of it. The figure also demonstrates the absence of notable score outliers. Figure 7d shows that no significant cases existed [34].

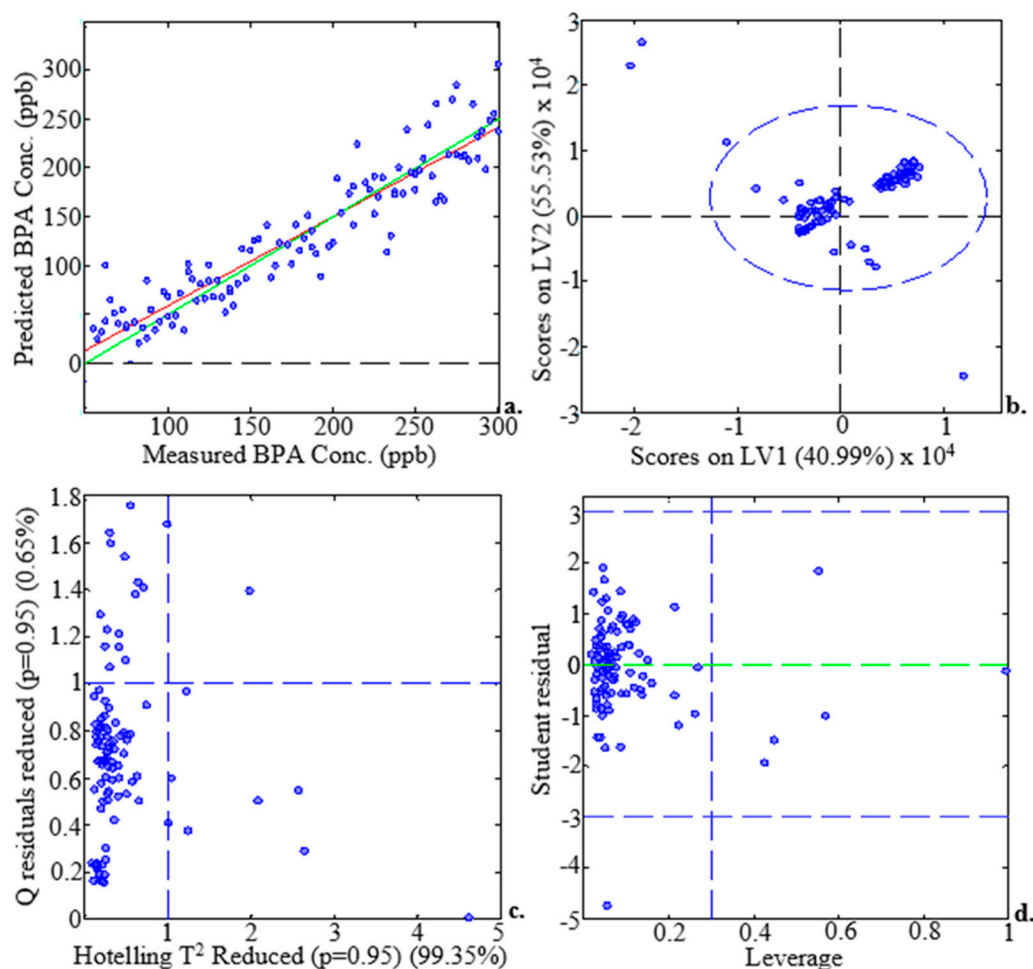


Figure 7. Samples/scores plots for the analysis of BPA concentration range (3-300 μM) depicting plots of (a) measured BPA (μM) versus predicted BPA (μM), (b) scores on LV1 versus scores on LV2, (c) Hotelling's T2 statistic versus Q residuals, and (d) leverage versus studentized residuals 1 BPA (μM).

The measured parameter values demonstrate that BPA in surface water can be analysed using the proposed method (Table 2). The RMSECV measures model performance in relation to validation samples, whereas the RMSEC measures model performance during the calibration (training) stage. The RMSEC and RMSECV values, which are 17.434 and 34.794 μM , respectively, are relatively low, indicating that the cross-validation and calibration errors are also low. Hence, the model is accurate. These variables, along with the percentage of variance explained, are indicators of how well a model performs in making predictions. The PLS model with 5 LVs was the best choice for the BPA regression analysis. The model explained 90.41% of the variance in the dependent variables and 98.84% of the variance in the predictors as a whole. Skewed estimates had no impact on the ability of the predictive model to make accurate predictions because of its low calibration, CV, and prediction biases [35]. The 120 EEM dataset had three outliers removed, and the PARAFAC model performed better as a result, according to the results. The values of R^2 Cal and R^2 CV both increased by 0.73% and 3.79%, respectively, which are small increments. This illustrated that the model was robust [36].

Table 2. PLS model performance statistics.

Parameter	Value
Number of LVs	5
RMSEC (μM)	17.434
RMSECV (μM)	34.794
Calibration Bias	1.396
CV Bias	0.33
R ² for Calibration	0.967
R ² for Cross-Validation	0.845

The plotted BPA concentration calibration curve and BPA validation curve are shown in Figure 9 as black dots and red diamonds, respectively. The performance indicators for the PLS modelling of the calibration and validation datasets are also shown in Figure 8. The high value of R2 Cal of 0.927, which denotes a strong correlation between the predicted and measured BPA concentrations, shows this.

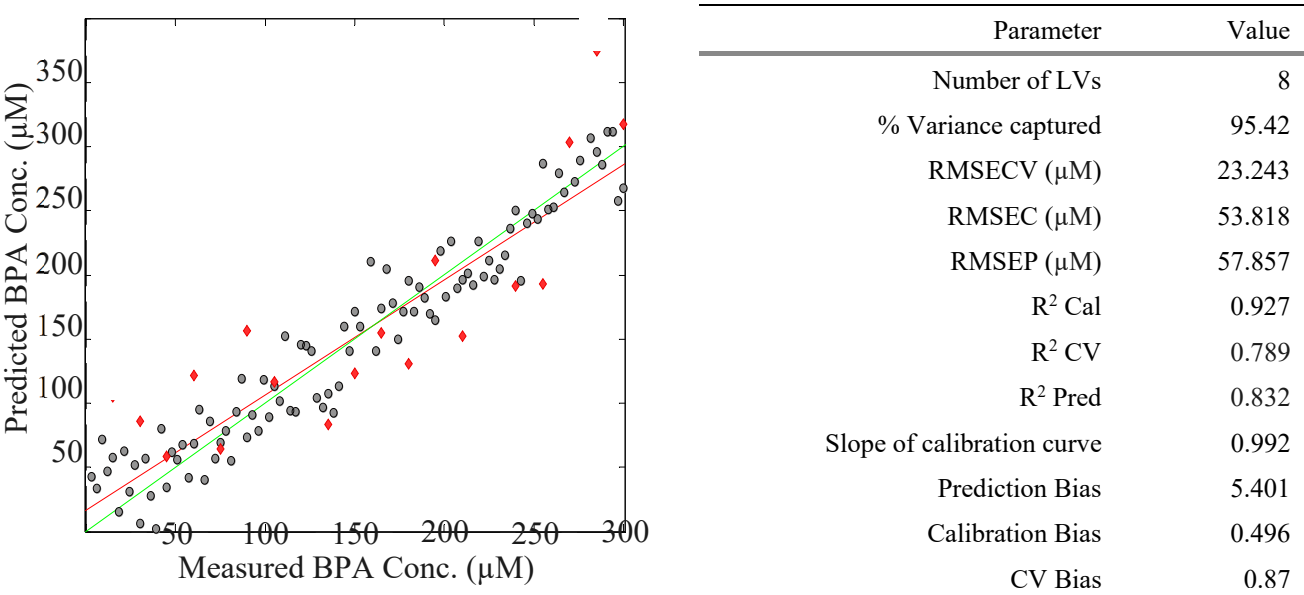


Figure 8. The plot of measured versus predicted BPA concentration (μM) for a 100-member calibration dataset (black dots) and a 16-member validation dataset (red diamonds) and the PLS model performance parameters for the analysis of surface water spiked with BPA.

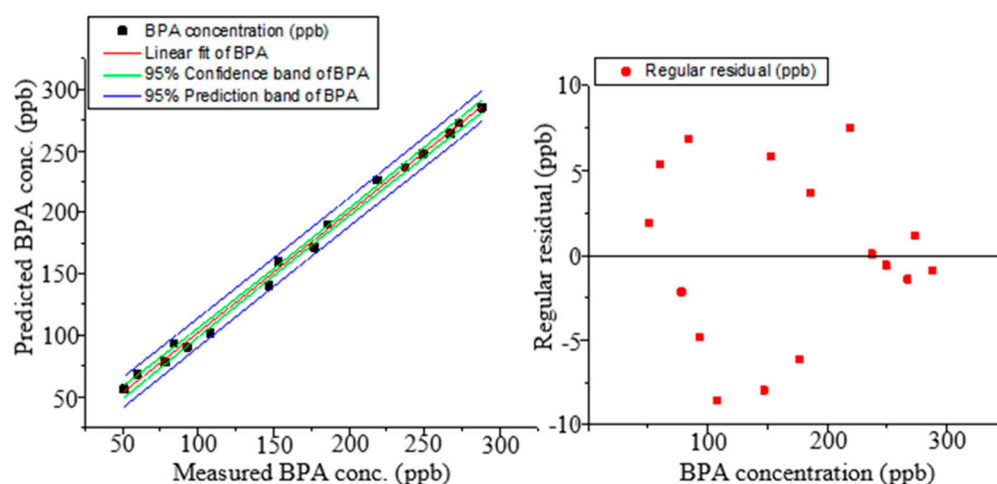


Figure 9. (a) The linear fit of the BPA validation data. Green lines show 95% confidence bands, blue lines show 95% prediction bands, and a red line shows the linear fit. (b) The plot of the regular residuals versus the independent variables for BPA.

The calibration and validation curves were found to be linear, and the high values of R^2 Cal, R^2 CV, and R^2 Pred—0.927, 0.677, and 0.832, respectively—demonstrated the robustness of the PLS model. Following the splitting of the 120 EEM data, the magnitudes of R^2 Cal and R^2 CV increased by 1.86% and 2.17%, respectively, demonstrating that the splitting of the 120-member dataset being modelled improved model performance and these results showed the robustness of the PLS model [37]. This improvement can also be attributed to the removal of outliers. Before deciding whether to include outliers in the regression analysis process or not, a very thorough analysis must be conducted. Because the RMSEP was so low (5.786 μM) and the R^2 Pred was high (0.832), the predictive power of the model and dependability were strong. The small biases obtained suggested that the A-TEEM-PLS method was correctly measuring and predicting the BPA concentration.

2.5. Validation of spiked samples

Figure 9a displays the predictions made by PLS modelling for the validation samples, i.e., surface water samples that were spiked with BPA at concentrations ranging from 50 to 270 μM . The PLS regression model generated a precise prediction model with constrained 95% prediction and 95% confidence bandwidths. Results from the model's prediction revealed a strong correlation, with the R^2 for the linear fit equalling 0.996. According to the R^2 value, 99.6% of the fluorescence spectral data fit the PLS model. A random distribution of the points along the horizontal axis of the regular residuals versus the BPA concentration and consistent residuals of ± 10 μM were observed in the residual plot (Figure 9b), indicating that the linear regression model correctly predicted the data [37].

The linear regression analysis report (Table 3) displays the linear regression parameters. The low residual sum of squares was evidence that the PLS model adequately fit the data. The low standard error of the intercept and slope values of 3.079 and 0.0167, respectively, that were obtained demonstrated the dependability of the analytical method and that the regression coefficients obtained were close to the actual coefficients. The R^2 , adj. R^2 , and Pearson's r values of 0.996, 0.996, and 0.998, respectively, demonstrated a strong correlation between the predicted and measured BPA concentrations [38]. Considering that Pearson's r value was positive, it was assumed that there was a positive correlation. The RMSE value for the prediction was 5.272 μM , while the mean absolute error (MAE) value was 4.378 μM , a difference of 0.894 μM (Table 3). Due to the fact that RMSE gives more weight to larger differences between values than MAE does, the dataset's outliers can be blamed for the significant MAE and RMSE difference (0.894 μM) [38].

Table 3. The linear regression analysis report for surface water samples spiked with BPA at concentrations of 60, 90, 120, 210, and 300 μM (from Origin V8.6).

Parameter	Value
Residual sum of squares	97.311
Pearson's r	0.998
R-Square (COD)	0.996
Adj. R-Square	0.996
RMSE	5.272
MAE	4.378
Intercept	4.219
Standard error of intercept	3.079
Slope	0.98
Standard error of Slope	0.0167

The ANOVA table (Table 4) displays how the sum of squares and mean sum of squares are distributed based on the source of variation. The sum of squares and mean sum of squares have standard errors of 389.073 and 27.791 μM , respectively, showing that the independent and dependent variables measured and predicted by the regression model are not significantly different from one another. However, the sum of squares value illustrates that the model fits the data rather poorly. On the other hand, the analytical method was deemed significant due to its high F-value of 210.474 [39]. The p-value for the entire model test is Prob > F. The p-value is shown as Prob > F in the output of a regression model with m independent variables [40]. The p-value of zero, which was less than 0.05, indicated that the model and data had a significant goodness of fit. It further demonstrated the statistical significance of the analytical technique [41]. The null hypothesis, which states that none of the measured BPA concentrations were related to the predicted BPA concentrations, was rejected because the p-value was less than 0.05. In other words, variations in the predicted BPA concentration were consistent with variations in the measured BPA concentration.

Table 4. ANOVA table for linearity of the BPA regression model .

	Degrees of Freedom	Sum of Squares	Mean Squares	F Value	Prob > F
Model	1	95914.648	95914.648	210.474	0
Error	14	389.073	27.791		
Total	15	96303.722			

2.6. Limits of detection and quantification

The calculated detection and quantification limits for BPA in surface water were 3.512 and 11.708 μM , respectively. These were the lowest BPA concentrations that the A-TEEM-PLS analytical technique could accurately identify and quantitate.

2.7. Recovery and accuracy

The accuracy was measured as a percent recovery in accordance with the ASTM [37] and ICH [39] recommendations. An average method accuracy of 97.55% was obtained from the recovery determined by analysing three different BPA concentrations in surface water (at concentrations of 50, 180, and 270 µM) (Table 5). This suggests that the A-TEEM-PLS method is highly accurate because the results were, on average, 97.55% of what the theoretical calculation would have predicted.

Table 5. Percent recoveries of BPA spiked at three different concentration levels in surface water .

Nominal conc. of BPA (µM)	Measured conc. of BPA (µM)	Percent Recovery
50	47.715	95.43
180	178.686	99.27
270	264.465	97.95

3. Materials and methods

3.1. Materials and reagents

For use in this study, BPA with a 97% purity level, AR-grade methanol, and 0.45-micron GMF filters were provided by Sigma-Aldrich®, Modderfontein, Johannesburg, South Africa. Using an Elix Integral 10 water purification system (Merck Millipore, Germiston, East Rand, Republic of South Africa), deionised water was produced on-site.

3.2. Sampling

Using the grab sampling technique [42], surface water samples were taken at the Florida stream in Johannesburg, South Africa, during the winter of 2022 (26.1739° S, 27.8971° E). Clean 1L amber glass bottles were used to collect the surface water samples. Following which, the samples were kept at 4 °C prior to being analysed for their absorbance and fluorescence properties within 48 hours of sampling. The samples were then warmed to room temperature, filtered to remove larger bacteria and particles using 0.45-micron GMF filters, and spiked with BPA right before the acquisition of absorbance and fluorescence EEM spectra.

3.3. Preparation of a stock solution and an intermediate standard solution

With a purity level of 97%, 11.77 g of BPA were weighed on a Mettler Toledo analytical balance (Switzerland) and then dissolved in 1 L of methanol to prepare a BPA stock solution with a concentration of 0.05 M. To prepare an intermediate standard solution of BPA at 1000 µM, 2 mL of the stock solution was diluted with AR-grade methanol in a 100-mL volumetric flask.

3.4. Sample preparation

Prior to use, the sample cuvette for then Aqualog® spectrometer was thoroughly cleaned by submerging it completely for 12 hours in 50% aqueous nitric acid and then rinsing it with deionised water to lessen background contributions. In order to prepare 100 calibration samples with BPA concentrations between 3 and 300 µM and 20 validation samples with BPA concentrations between 15 and 300 µM, the same quartz cuvette was filled with aliquots of surface water that had been filtered through a 0.45-micron GMF filter along with the BPA standards. The concentrations of the BPA standards in both sample types were different but evenly spaced. The total volume of the cuvette in its finished state was 4000 µL. After all the aliquots were placed inside the cuvette, the contents were vigorously shaken to ensure sample homogeneity. The final methanol concentration in the samples was limited to less than 2% in the cuvette because methanol has the potential to alter the solution's refractive index and contains a fluorescent background [11].

3.5. Total organic carbon determination

The Tekmar-Teledyne TOC analyser, which is based on UV-catalysed persulphate digestion to produce carbon dioxide, was used to measure the total organic carbon (TOC) in the surface water. Carbon dioxide was then detected by a nondispersive infrared detector.

3.6. The calibration of the A-TEEM instrument

The A-TEEM spectrometer auto-calibrates, implying that the system initialises the drives of its monochromators, finds each drive's home position, and assigns a wavelength value to this position using data from a calibration file. A sealed cuvette of Raman water was scanned, though, to check on calibration and throughput while also normalising EEM data.

3.7. Instrumentation and software

The A-TEEM spectrometer (HORIBA Aqualog® Yobin Yvon model UV-800C) was used to collect the EEM spectral data of a surface water sample that had not been spiked and surface water samples that had been spiked with BPA standard solutions. The sample queue method was employed to collect EEM data. The acquired EEM spectra served as fluorescent fingerprints. The instrumental parameters used were a fixed 5 nm optical slit width, emission wavelengths from 245.21 to 827.32 nm spaced at 8 pixels (4.66 nm) and excitation wavelengths from 200 to 800 nm spaced at 5 nm.

The signal-to-noise ratio (S/N) has a big impact on how well data is acquired. A suitable integration time of 0.5 seconds and a medium CCD gain were applied to improve the S/N. To avoid CCD saturation during data collection, the signal was kept within the linear range of the detector. Quick and simultaneous EEM spectral data acquisition was made possible by the device, which used a thermoelectrically cooled CCD spectrograph detector to capture the entire emission spectrum at each excitation increment. A saturation mask width of 16 nm was used to lessen sensitivity to erroneous CCD signal saturation alerts and Rayleigh scatter width. Silicon photodiodes were used for the reference and absorbance detectors.

3.8. Multi-way data analysis

Large data sets can be analysed using multi-way data analysis techniques by encoding the data as a multidimensional array. The Eigenvector Solo software versions 8.7 and 8.6, on which the PARAFAC and PLS multi-way analyses were based, were hyphenated to the A-TEEM spectrometer and were used to interpret all of the EEMs in the current study (Figure 10). The stand-alone algorithms were run on the PLS_Toolbox.

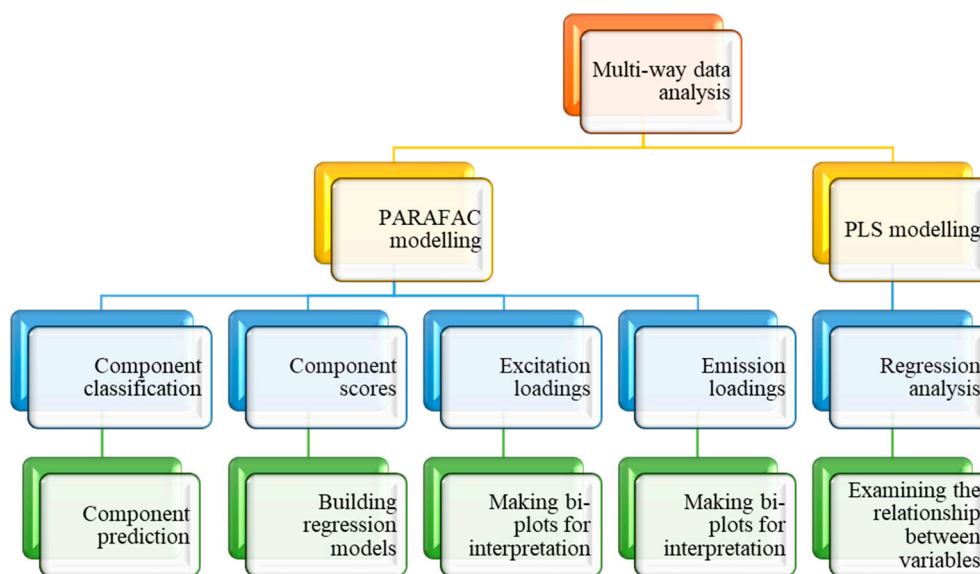


Figure 10. Workflow for multi-way EEM data analyses with Solo software.

3.8.1. Optimisation of the PARAFAC and PLS models

The experimental results were optimised due to various software and hardware conditions as well as light scattering and spectrum modifications that can result in artifacts. Optimising data acquisition improved spectral resolution. This was accomplished by using a large sample size, filtering the samples through 0.45-micron GMF filters, meticulously setting up the spectrofluorometer (Section 2.7), and preprocessing the EEM data acquired.

In order to interpret significant results, a large sample size (i.e., 100 calibration samples and 20 validation samples) was used to allow for a more accurate measurement of the treatment effect [43].

Therefore, a large sample size could optimise the PARAFAC and PLS models because the predictive power of the model increased as more components were added to it [44].

To more effectively classify the data, data preprocessing techniques like feature extraction, dimension reduction, scaling to the reference detector, spectral correction, dark signal subtraction, blank subtraction, and normalisation [45] were used, utilising the A-TEEM spectrometer software. To enhance the performance of the classifier, preprocessing aimed identify the most informative set of features.

To maximise the number of fluorescent components in the samples analysed, the fits of one- to five-component PARAFAC models of the EEM data were investigated. Each of the five models was evaluated for appropriateness using the percent captured variation, spectral loadings visualisation, split-half analysis, and core consistency techniques as outlined in the Section 2.8.3 [27,46].

The optimisation of variable selection and latent variables (LVs) was undertaken during PLS modelling as described in Section 2.8.4. The root mean square error of cross-validation (RMSECV) optimised the number of LVs. This was necessary because an insufficient number of LVs would indicate an insignificant relationship between the two variables.

Outliers were identified during PLS modelling by examining four types of sample/score plots; namely, the predicted and measured BPA concentration, scores on LV1 versus scores on LV2, Hotelling's T2 statistic versus Q residuals reduced, and leverage versus studentized residuals plots. The outliers were tentatively removed as they can affect the overall quality of the model and this is a crucial stage in the design of an optimised model [47]. These plots operate in the manner specified in Section 2.8.4.

3.8.2. Construction of the PARAFAC model

PARAFAC modelling was undertaken for the 120 EEM dataset. To rectify any systematically biased data, get rid of interference from inner-filter effects (IFEs), Raman scatter, and Rayleigh scatter, and normalise datasets with substantial intensity variations between samples, the EEM data were preprocessed. Following the importation of the EEM data from the A-TEEM instrument to the Solo software, the first- and second-order Rayleigh scatter, primary and secondary Raman scatter, and inner filter effects were corrected using customised functions in the Aqualog® spectrometer software. EEM filtering was performed by setting the first-order Rayleigh filter to 16 nm, and the second-order Rayleigh filter to 32 nm. The filter half-width was set to 16 nm and the default Raman shift of 3382 cm⁻¹ was used to mask the Raman scatter. The fluorescence intensities of the corrected EEM were normalised into Raman Units (R.U.) based on the peak area obtained using the same spectrometer to measure the Ex/Em 350 nm/396.5 nm 2D-spectrum of a sealed water-Raman cuvette. To prevent fluorescence peaks from overlapping or approaching the Raman or Rayleigh scatter the bandwidth was varied [15].

Since concentrations, absorptivities, and other physical characteristics of chemical compounds are mostly positive [15], non-negativity-constrained PARAFAC models were built using one to five components. Thus, all problem values were equal to or greater than zero. The use of the non-negativity constraint significantly reduced the feasible space of the parameters to be evaluated. The captured variance, core consistency, split-half evaluation, and analysing the spectral loadings visually approaches were used to assess the validity of the PARAFAC models fitted.

3.8.3. PARAFAC model validation

Split-half validation analysis was used to verify consistency within the dataset and was carried out by splitting the EEM dataset (in relation to samples) into two equal separate datasets and subjecting them to PARAFAC modelling. The two halves were modelled independently, and their outcomes were compared to determine the similarities as well as with modelling results of the whole EEM dataset.

In this study, by observing that an increase obtained with more than the optimal number of components is modest compared to an increase in the explained variance obtained using up to the optimal number of components, the explained variance technique was used to determine the ideal

number of components [27]. Underfitting will occur if there are too few components extracted; this can be quickly identified by evaluating the explained variance, since a model that is not well-fit usually has a decreased explained variance. On the other hand, using an excessive number of components will overfit the model. The obtained model in this scenario will have nearly 100% explained variance.

The ideal number of components was determined to match the largest model's number while maintaining a high level of core consistency [25]. A core consistency value close to 100 indicates a well-described model, while significantly lower values point to superfluous components.

The fluorescence spectra of the samples were visually inspected for broad and frequently unimodal peaks, making the visual appearance of the loadings a helpful diagnostic [27]. The ideal number of components and, consequently, the suitability of the model had to be determined by striking a balance between split-half validation, explained variation, core consistency, and spectral loadings [26].

3.8.4. Construction of the PLS model

The spectral (X-block) calibration dataset was preprocessed for regression and classification by mean-centring and clutter removal using the full-rank extended mixture model. Concentration (Y-block) calibration data were also preprocessed by mean-centring. The emission spectra of 120 surface water samples spiked with BPA were plotted to show the maximum emission wavelength of BPA. In order to obtain accurate data analysis results by removing noise or unwanted signals, preprocessing of the emission spectral data involved the exclusion of emission data just larger than 500 nm and higher [48].

The emission spectra and concentration values for the 120 samples from the X-block were copied and loaded into a new Y-block. The SIMPS algorithm from the Solo package (version 8.6) was used to calculate the PLS model. Each value derived from the calibration dataset and sample data sets was mean-centred then autoscaled. Preprocessing of the X- and Y-block data was carried out to extract useful data. The extended full-rank mixture model, including mean-centring and clutter eradication, was used to perform additional preprocessing on the dataset for spectral calibration (X-block). Additionally, mean-centring was used as a preprocessing step for the concentration calibration data (Y-block). The confusion matrix's specifications created for the exacting requirements ($p > 0.5$) and the most likely classification method in accordance with the overall successful identifications were used to sort the data.

The model was made accurate and complete by performing variable selection to find every variable that affected the outcome. By eliminating extraneous variables that increased model complexity and decreased precision, a small number of variables were chosen using the recursive PLS (rPLS) algorithm. According to the regression coefficient calculated by the PLS model after each iteration, the variables were weighted recursively in rPLS. Initial spectral data were subjected to the variable selection procedure in order to determine the most useful variables and eliminate insignificant variables [49]. The selection process was carried out using the Solo package (version 8.6) for Eigenvector, Inc.'s PLS_Toolbox. The cross-validation algorithm was used to iterate until the minimum RMSECV and maximum correlation coefficient were attained.

The RMSECV, which frequently has a minimum whose prominence varies with the noise in the data, was used to determine the ideal number of LVs [49]. As a result, this recursive validation was used to determine the number of LVs to use in the model.

The model was calculated and evaluated to determine the data variation captured by the model. The PLS predictive method involved tracking the development of the RMSECV and the root mean square error of calibration (RMSEC) parameters, that the rPLS technique extracted and synthesised in relation to the number of estimated LVs.

To determine outliers, clusters, and recognisable patterns in the line plots of the scores, the sample/scores plots were constructed. The scatter plots were examined for data points that showed a recurring pattern, and the correlation between the two continuous variables was determined by analysing the point distribution of the scatter plots. On the plot of the measured versus predicted

BPA concentration, the form, direction, strength, and presence of outliers were evaluated. The statistical confidence region was added to the scores on the LV1 versus LV2 plot to help identify these outliers because they were outside of it [50]. The total variance in Hotelling's T2 statistic and in the lack-of-fit statistic were obtained from the plot of Hotelling's T2 versus Q residuals reduced. Any significant outliers in the upper- or lower-right corners of the plot were found using the studentized residuals versus leverage plot.

The 120 EEM dataset was split into a 20-member validation subset and a 100-member calibration subset. To assess the effectiveness of the PLS model, a PLS analysis was performed on a batch of 100 BPA standards with BPA concentrations ranging from 3 to 300 μM and 20 validation samples with BPA concentrations ranging from 15 to 300 μM . The calibration dataset was used to find or learn relationships between the traits and the target variable. The validation dataset was used to provide an impartial evaluation of a model fit on the training dataset while model hyperparameters were being adjusted [52]. BPA concentrations in spiked surface water samples were predicted based on two-dimensional (2-D) EEMs derived from the unfolding of three-dimensional (3-D) EEMs using principal component analysis.

The performance parameters of the PLS model (RMSECV, RMSEC, coefficient of correlation for prediction (R^2 Pred), cross-validation (R^2 CV), calibration (R^2 Cal), and the root mean square error of prediction (RMSEP or SEP)) were assessed to evaluate the effectiveness of the regression model. A number of figures of merit, including adjusted R-squared (adj. R^2), standard error, linear fit slope, and intercept parameters, were all established and assessed for the performance of the model in order to further validate it.

3.8.5. PLS model validation

The model was shown to be accurate by predicting BPA concentrations in surface water samples that had been spiked. PLS model performance parameters derived from the PLS analysis were used to evaluate the model's validity. Included in the parameters were the RMSEP, RMSECV, RMSEC, R^2 Pred, R^2 CV, and R^2 Cal.

3.8.6. Validation of spiked surface water samples

A measured versus predicted BPA concentration plot and a regular residual versus independent variables plot for BPA were constructed using the EEM data from surface water samples spiked with the BPA standards at concentrations of 15 to 300 μM . The Origin V8.6 software was used to construct the plots. The predicted and measured BPA concentration plot was used to generate regression parameters, which were then compiled in a report along with other goodness-of-fit parameters. The accuracy of the model was evaluated by comparing the predicted and measured values using RMSE, MAE, Person's r value, adj. R^2 , residual sum of squares, standard error of the intercept, intercept, and slope [37]. The difference between the measured and estimated values assesses how outliers affect the data, with RMSE giving outliers a higher priority than MAE. The standard p-value was set at 0.05, and the confidence and prediction intervals were both set at 95%. The plot of the regular residual versus independent variables provided information on the residual distribution. Degrees of freedom, squared sum of residuals, squared mean, F-value, and probability > F or p-value statistical parameters were compiled in an ANOVA table.

An F-value was used to determine the statistical significance of the test [53]. Prob > F is the p-value for the entire model test. The p-value, as Prob > F-value, was used to ascertain the relationship between the predicted and measured BPA concentrations and test the hypothesis that none of the independent variables are related to the dependent variables.

3.8.7. Determination of limits of detection and quantification of the A-TEEM-PLS method

The slope of the calibration graph and the standard error of the intercept of the graph were used to calculate the LOD of the A-TEEM-PLS method using Equation 1 [11]:

$$\text{LOD} = (3 \times \text{Standard error of intercept} + \text{Intercept}) / \text{Slope} \quad (1)$$

Based on the slope of the calibration graph and the standard error of the intercept of the graph, the LOQ of the A-TEEM-PLS method was calculated using Equation 2 [11]:

$$\text{LOQ} = (10 \times \text{Standard error of intercept} + \text{Intercept}) / \text{Slope} \quad (2)$$

3.8.8. Accuracy and recovery of the method

The accuracy of the A-TEEM-PLS method was assessed in a single experimental run, with identical solutions, and by a single analyst. The percent recoveries for the new method (Equation 3) [54] were used to gauge accuracy:

$$\% \text{ Recovery} = (\text{obtained result}) / (\text{expected result}) \times 100 \quad (3)$$

The recoveries were determined by the acquisition of EEM data from surface water samples spiked with BPA at three different concentrations (60, 180, and 270 μM). The results were assessed for compliance with international method validation guidelines (SANCO/12495/2011), which require the mean recovery to range between 70 and 120% [55].

3.8.9. The robustness of the model

Because EEMs and component outliers like Raman and Rayleigh scattering have a significant impact on PARAFAC and PLS predictions [27,29], the 3-D EEM dataset underwent preprocessing to eliminate these interferences as described in Section 2.7.1 to strengthen the models. The width, smoothness, and similarity of the loading spectrum were used to evaluate the robustness of the PARAFAC model. The model was deemed to be robust if the loading spectra were smooth, comparable, broad, and frequently unimodal.

The magnitudes of the R2 Cal, R2 CV, and R2 Pred parameters were examined in order to gauge the robustness of the PLS model. Tests for nonlinearity effects and outlier removal were carried out to validate the model. The magnitudes of the coefficients of determination for the calibration (R2 Cal) and cross-validation (R2 CV) parameters were evaluated to determine the effects of nonlinearity. As part of the evaluation of the removal of outliers, it was determined how much the correlation coefficients had changed after the outliers were eliminated. LVs were used in the construction of a robust PLS model in order to find pertinent features and/or eliminate irrelevant variables so as to increase prediction accuracy and reduce model complexity.

4. Conclusions

The A-TEEM-PARAFAC-PLS analytical method is effective at monitoring BPA contamination in water because of its quick run time (less than 4 minutes), high sensitivity at lower micro mole levels for BPA, little to no sample preparation, and ease of use. Because it is challenging to detect aqueous fluorescent substances using optical spectroscopy, handling interferences and practical outlier control are two major advantages of using multi-way-based algorithms in optical data analysis. In fact, when using multi-way data, instrument selectivity is not necessary [56]. The obtained results demonstrated that the data from the spectra, extracted and analysed using PARAFAC and PLS algorithms, could be usefully exploited in the development of PARAFAC and rather robust regression models that demonstrated the ability to identify BPA in surface water and predict the BPA content of surface water.

Future research can look into combining the Solo regression and discrimination models (method models) developed in this study and multiblock models (absorbance and EEM/PEM data concatenation) to predict BPA concentrations in real surface water using the multi-model predictor tool. The current study could benefit greatly from this improvement in order to move it toward a more advanced stage of predicting BPA concentrations in real surface water analyses, which should at the very least be aimed at achieving a reliable classification as "quality control."

Also, A-TEEM spectroscopic methods will need to be developed in the future in conjunction with other chemometric tools, such as decomposition tools (e.g., multivariate curve resolution, SIMPLISMA (purity), and others), and regression tools (e.g., artificial neural network, designed

experiment MLR, locally weighted regression, and others), in order to track and quantify water quality-related parameters like microplastics, turbidity, chemical oxygen demand, oxidation-reduction potential, etc.

Authorship: Conceptualisation and study design, investigation, formal analysis, collection of data, first draft writing, editing, and review, T.I.; Writing: editing and proofreading, guidance, N.C.; Funding acquisition, supervision, B.B.M.; Supervision, funding acquisition, T.I.T.N.; Methodology, supervision, editing, A.M.G.

Funding: The authors are appreciative of the financial support secured from the University of South Africa.

Declaration of the Institutional Review Board: Not applicable

Data availability: Following a valid request, the corresponding author will provide the dataset used and analysed during this study.

Acknowledgements: We would like to thank everyone who contributed to this study. We acknowledge Nontokozo Magwaza for providing general support as well as W. Moyo, Institute of Nanotechnology and Water Sustainability, University of South Africa, for operating the Aqualog® spectrometer.

Interest-based conflict: The authors say they have no competing interests.

References

1. Iyagawa, S.; Sato, T.; Iguchi, T. Bisphenol A. In *Handbook of Hormones*, Academic Press, London, United Kingdom, 2021, 1003-1004.
2. Ugoeze, K.C.; Amogu, E.O.; Oluigbo, K.E.; Nwachukwu, N. Environmental and public health impacts of plastic wastes due to healthcare and food products packages: A Review. *J. Environ. Sci. Pub. Health*. 2021 5 1-31. DOI: 10.26502/jesph.96120114.
3. Ohore, O.E.; Zhang, S. Endocrine disrupting effects of bisphenol A exposure and recent advances on its removal by water treatment systems. A review, *Sci. Afr.* 5 2019, 1-12. <https://doi.org/10.1016/j.sciaf.2019.e00135>.
4. Nerín, C.; Fernández, C.; Domeño, C.; Salafranca J. Determination of potential migrants in polycarbonate containers used for microwave ovens by high-performance liquid chromatography with ultraviolet and fluorescence detection. *J. Agric. Food Chem.* 2003, 51, 5647-5653. <https://doi.org/10.1021/jf034330p>.
5. Caban, M.; Stepnowski, P. The quantification of bisphenols and their analogues in wastewaters and surface water by an improved solid-phase extraction gas chromatography/mass spectrometry method. *ESPR*. 2020, 27, 28829-28839. <https://doi.org/10.1007/s11356-020-09123-2>.
6. Zhou, W.; Sun, C.; Zhou, Y.; Yang, X.; Yang, W. A facial electrochemical approach to determinate bisphenol A based on graphene-hypercrosslinked resin MN202 composite. *Food Chem.* 2014, 158, 81-87. <https://doi.org/10.1016/j.foodchem.2014.02.123>.
7. El-Awady, M.; Pyell, U. Sweeping as a multistep enrichment process in micellar electrokinetic chromatography: The retention factor gradient effect. *J. Chromatogr. A*. 2013, 1297, 213-225. <https://doi.org/10.1016/j.chroma.2013.04.069>.
8. Katayama, M.; Matsuda, Y.; Shimokawa, K.I.; Ishikawa, H.; Kaneko, S. Preliminary monitoring of bisphenol A and nonylphenol in human semen by sensitive high performance liquid chromatography and capillary electrophoresis after proteinase K digestion. *Anal. Lett.* 2003, 36, 2659-2667. <https://doi.org/10.1081/AL-120024640>.
9. Hu, S.; He, Q.; Zhao, Z. Determination of trace amounts of estriol and estradiol by adsorptive cathodic stripping voltammetry. *Analyst*. 1992, 117, 181-184. <https://doi.org/10.1039/AN9921700181>.
10. Pan, Y.; Li, H.; Zhang, X.; Li, A. Characterization of natural organic matter in drinking water: Sample preparation and analytical approaches. *Trends Environ. Anal. Chem.* 2016, 12, 23-30. <https://doi.org/10.1016/j.teac.2016.11.002>.
11. Gilmore, A.M.; Chen, L. Optical early warning detection of aromatic hydrocarbons in drinking water sources with absorbance, transmission and fluorescence excitation-emission mapping (A-TEEM) instrument technology. In *Next-Generation Spectroscopic Technologies XII* (10983, 2019, 45-52). SPIE. <https://doi.org/10.1117/12.2522498>.
12. Gilmore, A.M. Horiba Advanced Techno Co Ltd and Horiba Instruments Inc. Determination of water treatment parameters based on absorbance and fluorescence U.S. Patent 9,670,072 (2017).

13. Zheng, C.; Liu, J.; Ren, J.; Shen, Fan, J.; Xi, R.; Chen, W.; Chen, Q. Occurrence, distribution and ecological risk of bisphenol analogues in the surface water from a water diversion project in Nanjing, China. *IJERPH*. 2019, 16, 1-11. <https://doi.org/10.3390/ijerph16183296>.
14. Nikolajsen, R.P.H.; Booksh, K.S.; Hansen, A.M.; Bro, R. Quantifying catecholamines using multi-way kinetic modelling. *Anal. Chim. Acta*. 2003, 475, 137-150. [https://doi.org/10.1016/S0003-2670\(02\)01256-4](https://doi.org/10.1016/S0003-2670(02)01256-4).
15. Bro, R. PARAFAC. Tutorial and applications. *Chemom. Intell. Lab. Syst.* 1997, 38, 149-171. [https://doi.org/10.1016/S0169-7439\(97\)00032-4](https://doi.org/10.1016/S0169-7439(97)00032-4).
16. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 2001, 58, 109-130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
17. Siqueira, L.F.; Júnior, R.F.A.; de Araújo, A.A.; Morais, C.L.; Lima, K.M. LDA vs. QDA for FT-MIR prostate cancer tissue classification. *Chemom. Intell. Lab. Syst.* 2017, 162, 123-129.
18. Derco, J.; Dudáš, J.; Valičková, M.; Sumegová, L.; Murínová, S. Removal of Alkylphenols from Industrial and Municipal Wastewater. *Chem. Biochem. Eng. Q.* 2017, 31, 173-178. <https://doi.org/10.15255/CABEQ.2016.1021>.
19. Bocharnikova, E.N.; Tchaikovskaya, O.N.; Bazyl, O.K.; Artyukhov, V.Y.; Mayer, G.V. Theoretical study of bisphenol A photolysis. In *Advances in Quantum Chemistry*, Academic Press, London, United Kingdom. 2020, 81, 191-217. <https://doi.org/10.1016/bs.aiq.2019.12.001>.
20. Wünsch, U.J.; Murphy, K.R.; Stedmon, C.A. Corrigendum: Fluorescence quantum yields of natural organic matter and organic compounds: Implications for the fluorescence-based interpretation of organic matter composition. *Front. Mar. Sci.* 2016, 3, 9. DOI:10.3389/fmars.2015.00098.
21. Xia, Y. Correlation and association analyses in microbiome study integrating multiomics in health and disease. *PMBTS* 2020, 171, 309-491. <https://doi.org/10.1016/bs.pmbts.2020.04.003>.
22. Hudson, N.; Baker, A.; Reynolds, D. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters-a review. *River Res. Appl.* 23 (2007) 631-649. <https://doi.org/10.1002/rra.1005>.
23. Del Olmo, M.; Zafra, A.; Jurado, A.B.; Vilchez, J.L. Determination of bisphenol A (BPA) in the presence of phenol by first-derivative fluorescence following micro liquid-liquid extraction (MLLE). *Talanta*. 2000, 50, 1141-1148. [https://doi.org/10.1016/S0039-9140\(99\)00176-9](https://doi.org/10.1016/S0039-9140(99)00176-9).
24. Sethuraman, S.; Rajendran, K. Multicharacteristic behavior of tyrosine present in the microdomains of the macromolecule gum arabic at various pH conditions, *ACS omega*. 2018, 3, 17602-17609. <https://doi.org/10.1021/acsomega.8b02928>.
25. Brereton, R.G. Multilevel multifactor designs for multivariate calibration. *Analyst*. 1997, 122, 1521-1529. DOI: 10.1039/A703654J.
26. Lia, F.; Formosa, J.P.; Zammit-Mangion, M.; Farrugia, C. The first identification of the uniqueness and authentication of Maltese extra virgin olive oil using 3D-fluorescence spectroscopy coupled with multi-way data analysis. *Foods*. 2020, 9, 1-14. <https://doi.org/10.3390/foods9040498>.
27. Andersen, C.M.; Bro, R. Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *J. Chemom.* 2003, 17, 200-215. DOI: 10.1002/cem.790.
28. Bro, R. Interactive introduction to multi-way analysis in MATLAB: Chapter 2: Basic PARAFAC modeling. The N-way on-line course on PARAFAC and PLS 2015.
29. Murphy, K.R.; Stedmon, C.A.; Graeber, D.; Bro, R. Fluorescence spectroscopy and multi-way techniques. *PARAFAC. Anal. Methods*. 2013, 5, 6557-6566. DOI: 10.1039/c3ay41160e.
30. Sierra, M.M.D.; Giovanela, M.; Parlanti, E.; Soriano-Sierra, E.J. Fluorescence fingerprint of fulvic and humic acids from varied origins as viewed by single-scan and excitation/emission matrix techniques. *Chemosphere*. 2005, 58, 715-733. <https://doi.org/10.1016/j.chemosphere.2004.09.038>.
31. Liu, H.; Pu, Y.; Qiu, X.; Li, B.; Sun, B. Zhu, X.; Liu, K. Humic Acid Extracts Leading to the Photochemical Bromination of Phenol in Aqueous Bromide Solutions: Influences of Aromatic Components, Polarity and Photochemical Activity. *Molecules*. 2021, 26, 1-12. <https://doi.org/10.3390/molecules26030608>.
32. Hong, Y.J.; Nam, C.J.; Song, K.B.; Cho, G.S.; Uhm, H.S.; Choi, D.I.; Choi, E.H. Measurement of hydroxyl radical density generated from the atmospheric pressure bioplasma jet. *JINST*. 2012, 7, 1-12. DOI 10.1088/1748-0221/7/03/C03046.
33. Polat, E.; Gunay, S. A New Robust Partial Least Squares Regression Method Based on a Robust and an Efficient Adaptive Reweighted Estimator of Covariance. *REVSTAT-STA J.* 2019, 17, 449-474. <https://doi.org/10.57805/revstat.v17i4.276>.
34. Zhang, Z. Residuals and regression diagnostics: focusing on logistic regression. *Ann. Transl. Med.* 2016, 4, 1-8. doi: 10.21037/atm.2016.03.36.
35. Xu, T.; Valocchi, A.J.; A Bayesian approach to improved calibration and prediction of groundwater models with structural error. *Water Resour. Res.* 2015, 51, 9290-9311. <https://doi.org/10.1002/2015WR017912>.
36. Refat, B.; Yu, P. Evaluation of prediction of indigestible fiber fraction (iNDF) of whole-crop barley silage by using non-destructive spectroscopic techniques as a fast-screening method: comparison between FTIR vs. NIR. *Can. J. Plant Sci.* 2022, 102, pp.1130-1138. <https://doi.org/10.1139/cjps-2022-0054>.

37. ASTM, Standard Practice for Validation of Empirically Derived Multivariate Calibrations Methods. Document E2617-17. ASTM International, West Conshohocken, USA. 2018, 10.1520/E2617-17.2.
38. Shen, J.; Valagolam, D.; McCalla, S. Prophet forecasting model: A machine learning approach to predict the concentration of air pollutants (PM_{2.5}, PM₁₀, O₃, NO₂, SO₂, CO) in Seoul, South Korea. *PeerJ*. 2020, 8, 1-18. <https://doi.org/10.7717/peerj.9961>.
39. Darpo, B.; Nebout, T.; Sager, P.T. Clinical evaluation of QT/QTc prolongation and proarrhythmic potential for nonantiarrhythmic drugs: The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use E14 guideline. *J. Clin. Pharmacol.* 2006, 46, 498-507. <https://doi.org/10.1177/0091270006286436>.
40. Zulfikar, R.; STp, M.M. Estimation model and selection method of panel data regression: an overview of common effect, fixed effect, and random effect model, INA-Rxiv 9qe2b, Center for Open Science 2019, 1-10.
41. Ziliak, S. P values and the search for significance. *Nat. Methods*. 2017, 14, 3-4.
42. Kianpoor Kalkhajeh, Y.; Jabbarian Amiri, B.; Huang, B.; Henareh Khalyani, A.; Hu, W.; Gao, H.; Thompson, M.L. Methods for sample collection, storage, and analysis of freshwater phosphorus. *Water*. 2019, 11, 1-24. <https://doi.org/10.3390/w11091889>.
43. Biau, D.J.; Kernéis, S.; Porcher, R. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. *Clin. Orthop. Relat. Res.* 2008, 466, 2282-2288. <https://doi.org/10.1007/s11999-008-0346-9>.
44. Mundfrom, D.J.; Shaw, D.G.; Ke, T.L. Minimum sample size recommendations for conducting factor analyses. *Int. J. Test.* 2005 5, 159-168. https://doi.org/10.1207/s15327574ijt0502_4.
45. Subasi, A. Practical machine learning for data analysis using python. Academic Press. 2020.
46. Bro, R.; Kiers, H.A. A new efficient method for determining the number of components in PARAFAC models. *J. Chemom.* 17 (2003) 274-286. DOI: 10.1002/cem.801.
47. Stedmon, C.A.; Bro, R. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *LIMNOL. OCEANOGR. METH.* 2008, 6, 572-579. <https://doi.org/10.4319/lom.2008.6.572>.
48. Gautam, R.; Vanga, S.; Ariese, F.; Umaphathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ tech. instrum.* 2015, 2, 1-38. DOI 10.1140/epjti/s40485-015-0018-6.
49. Deng, B.C.; Yun, Y.H.; Liang, Y.Z.; Cao, D.S.; Xu, Q.S.; Yi, L.Z.; Huang, X. A new strategy to prevent over-fitting in partial least squares models based on model population analysis. *Anal. Chim. Acta.* 2015, 880, 32-41. <https://doi.org/10.1016/j.aca.2015.04.045>.
50. Rinnan, A.; Andersson, M.; Ridder, C.; Engelsen, S.B. Recursive weighted partial least squares (rPLS): an efficient variable selection method using PLS. *J. Chemom.* 2014, 28, 439-447. <https://doi.org/10.1002/cem.2582>.
51. Häggblom, K.E. Basics of Multivariate Modelling and Data Analysis. Disponible en. 2018. <http://www.users.abo.fi/khaggblo/MMDA/MMDA6.pdf>.
52. Sarigiannis, D.; Parnell, T.; Pozidis, H. Weighted sampling for combined model selection and hyperparameter tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, 34, 5595-5603. DOI: <https://doi.org/10.1609/aaai.v34i04.6012>.
53. Chen, J.; LeBoeuf, E.J.; Dai, S.; Gu, B. Fluorescence spectroscopic studies of natural organic matter fractions. *Chemosphere*. 2003, 50, 639-647. [https://doi.org/10.1016/S0045-6535\(02\)00616-1](https://doi.org/10.1016/S0045-6535(02)00616-1).
54. Steiner, D.; Krska, R.; Malachová, A.; Taschl, I.; Sulyok, M. Evaluation of matrix effects and extraction efficiencies of LC-MS/MS methods as the essential part for proper validation of multiclass contaminants in complex feed. *J. Agric. Food Chem.* 2020, 68, 3868-3880. <https://doi.org/10.1021/acs.jafc.9b07706>.
55. Cassel, C.; Hackl, P.; Westlund, A.H. Robustness of partial least-squares method for estimating latent variable quality structures. *J. Appl. Stat.* 1999, 26, 435-446. <https://doi.org/10.1080/02664769922322>.
56. Bro, R. Multivariate calibration: what is in chemometrics for the analytical chemist? *Anal. Chim. Acta.* 2003, 500, 185-194. [https://doi.org/10.1016/S0003-2670\(03\)00681-0](https://doi.org/10.1016/S0003-2670(03)00681-0).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.