1 **Simultaneous Quantification of Aroclor Mixtures in Soil Samples**
2 **by Gas Chromatography/Mass Spectrometry with Solid Phase**
3 **Microextraction using Partial Least-Squares Regression**

4 Mengliang Zhang and Peter de B. Harrington*

5 Center for Intelligent Chemical Instrumentation, Clippinger

6 Laboratories, Department of Chemistry and Biochemistry, Ohio

7 University, Athens, Ohio 45701-2979, USA

8 **Abstract**

9 Multivariate partial least-squares (PLS) method was applied to the
10 quantification of two complex polychlorinated biphenyls (PCBs)
11 commercial mixtures, Aroclor 1254 and 1260, in a soil matrix. PCBs in
12 soil samples were extracted by headspace solid phase microextraction
13 (SPME) and determined by gas chromatography/mass spectrometry
14 (GC/MS). Decachlorinated biphenyl (deca-CB) was used as internal
15 standard. After the baseline correction was applied, four data
16 representations including extracted ion chromatograms (EIC) for
17 Aroclor 1254, EIC for Aroclor 1260, EIC for both Aroclors and two-way
18 data sets were constructed for PLS-1 and PLS-2 calibrations and
19 evaluated with respect to quantitative prediction accuracy. The PLS
20 model was optimized with respect to the number of latent variables

* Corresponding author: Tel: +01 740-994-0265/Fax: +01 740-593-
0148. *E-mail address:* peter.harrington@ohio.edu

using cross validation of the calibration data set. The validation of the method was performed with certified soil samples and real field soil samples and the predicted concentrations for both Aroclors using EIC data sets agreed with the certified values. The linear range of the method was from $10 \mu\text{g kg}^{-1}$ to $1000 \mu\text{g kg}^{-1}$ for both Aroclor 1254 and 1260 in soil matrices and the detection limit was $4 \mu\text{g kg}^{-1}$ for Aroclor 1254 and $6 \mu\text{g kg}^{-1}$ for Aroclor 1260. This holistic approach for the determination of mixtures of complex samples has broad application to environmental forensics and modeling.

Keywords: GC/MS; partial least squares; polychlorinated biphenyls; soil; solid phase microextraction.

1. Introduction

Polychlorinated biphenyls (PCBs) are persistent organic pollutants. As the only commercial PCBs products in North America, Aroclors are multi-component mixtures, which contain different relative quantities of 209 possible PCB congeners (Safe, 1984). The last two digits in Aroclor trade names reflect the percent Chlorine by mass in the mixture (e.g., 54% Chlorine by mass in Aroclor 1254) with the exception of Aroclor 1016, which is 41% Chlorine by mass (Newman et al., 1998). Aroclors had wide application from 1930 to 1975 such as in coolant, lubricant, greases, dyes and so on, and they can still be found in

the environment throughout the world which poses a threat to human health because of the high toxicity of PCBs (Safe, 1984).

PCBs were produced and disposed of in the environment as Aroclors, so the quantification of Aroclors is important. Two methods are commonly used for the quantification of Aroclors in environmental samples. One is the comprehensive congener-specific method (e.g., US EPA Method 1668c) (EPA, 2010) which requires high-resolution gas chromatography/mass spectrometry (GC/MS) to separate and determine each individual PCB congener in the sample. This method is expensive, difficult to perform, and requires expensive equipment, individual isotopic standards, and expertise that is not always available in many laboratories. A recent publication determined the concentrations of a single Aroclor in soil by summing together the concentrations of each PCB congener calculated by response factors of each PCB to PCB 209 (Bruzzoniti et al., 2012). In their method, sample preparation required 60 min and for sample analysis 65 min were required. The calculation of concentrations for each PCB congener (105 peaks and 9 response factors) was tedious and this method cannot analyze samples that contain more than one Aroclor.

The other method is a limited congener Aroclor estimation method (e.g., US EPA Method 8082a) (EPA, 2000). The Aroclors should be identified first based on the similarity of the sample chromatogram

with that of the most similar Aroclor standards. Then 3 to 5 major characteristic peaks were selected to build the calibration models and the average of the concentration results from these models was used to estimate the concentration of the Aroclor. A decision tree using PCB peak patterns (e.g., PCB118:203 for Aroclor 1260 and 1254) was required for the identification of samples with two or more Aroclors (Newman et al., 1998). The selection of characteristic peaks is experience-based, and although the full separation of all the PCB congeners was not required, this method required a relatively long GC program (≥ 45 min) to separate 3 to 5 selected characteristic peaks from the other PCB congeners (EPA, 2000).

Partial least squares (PLS) regression which is accomplished via a small set of latent variables (the components) is a powerful chemometric technique for multivariate calibration when the data are highly correlated (Li et al., 2002). As a full spectrum multivariate calibration method, PLS uses both spectral and concentration data to construct a calibration model. Only the concentrations for the analytes of interest (e.g., Aroclor 1254 or 1260) are necessary for model construction and other interferences or uncorrelated analytes (e.g., uncorrelated PCBs from other sources) have little effect on the calibration sets (Geladi and Kowalski, 1986). This advantage is important for the quantification of PCBs as Aroclors because it enables

the rapid and simultaneous determination of different Aroclors in the mixture without time-consuming separation to resolve the overlapped PCB congener peaks and with minimum sample preparation. For the same reason, experience based peak selection and tedious calculation of individual peak areas can be avoided. Another advantage of PLS is that the complexity of the model can be controlled by the number of components which is convenient to approach the maximum prediction performance and avoid overfitting (Geladi and Kowalski, 1986).

In this study, Aroclor 1254 and 1260 in soil matrices were analyzed by gas chromatography/mass spectrometry (GC/MS) combined with headspace solid phase microextraction (SPME). The multivariate PLS model was constructed for the quantification of both Aroclors using four data representations. The method was then validated with certified soil samples and real soil samples.

2. Theory

2.1. Partial least-squares (PLS) regression

PLS is one of the most widely used methods in chemometrics for multivariate calibration which was developed around 1975 by Herman Wold and then introduced into chemometrics by Svante Wold (Sjostrom et al., 1983; Wold et al., 1984; Inon et al., 2003). PLS is a method to relate a matrix **X** to a vector **Y** or to a matrix **Y**. Generally, there are three steps: 1) Calculate the mutual latent variables for both

108 the **X** matrix (e.g., two-way GC/MS data) and the **Y** matrix (e.g.,
 109 concentrations for Aroclors) based on maximum covariance; 2)
 110 Generate the diagonal regression matrix for all PLS components; 3)
 111 Predict the concentrations of calibration samples (cross validation)
 112 or/and unknown samples (test set) by retrospective calculation using
 113 the diagonal regression matrix and new **X** matrix (An et al., 2010). Fig.
 114 1 is the flowchart for the PLS method.

Fig. 1. Flowchart of PLS as a multiple linear regression method. **T** is the latent two-way GC/MS data matrix, **W** is the two-way GC/MS data loading matrix, **U** is the latent concentration matrix, **Q** represents the concentration loading matrix, **E_x**, **E_y**, and **E_d** are residual error matrices.

115 2.2. Root mean square of prediction error (RMSPE)

116 The quality of calibration by PLS can be evaluated by the RMSPE. A
 117 cross-validation method was used for the calibration data set. The
 118 prediction set contained samples that were never used for building the
 119 PLS calibration model. The RMSPE is given as

$$120 \quad RMSPE = \left[\sum_{j=1}^{n_y} \sum_{i=1}^m \frac{(\hat{y}_{ij} - y_{ij})^2}{nm} \right]^{1/2} \quad (1)$$

for which m is the number of data objects, n_y is the number of species, $y_{i,j}$ is the j th concentration for the i th data object, and $\hat{y}_{i,j}$ is the corresponding predicted concentration.

2.3. Coefficient of determination

The coefficient of determination (r^2) is calculated as follows:

$$r_j^2 = 1 - \frac{\sum_{i=1}^m (\hat{y}_{i,j} - y_{i,j})^2}{\sum_{i=1}^m (\bar{y}_{i,j} - y_{i,j})^2} \quad (2)$$

for $\bar{y}$ is the mean value of the concentration vector $\mathbf{y}_j$ for Aroclor j .

2.4. Relative predictive errors (RPEs).

The RPEs can be calculated as:

$$RPEs(\%) = 100 \times \frac{RMSPE}{s_y} \quad (3)$$

The standard deviation of $\mathbf{y}$ is given by s_y .

3. Experimental

3.1. Reagents

A mixture containing 1 mg mL⁻¹ of decachlorobiphenyl (deca-CB) and of tetrachloro-*m*-xylene (TCMX), 100 µg mL⁻¹ Aroclor 1254 and 1260 standards solutions were purchased from AccuStandard, Inc. (New Haven, CT). Potassium permanganate, potassium dichromate, the SPME fibers coated with polydimethylsiloxane (100 µm film thickness), 20-mL headspace glass vials, and crimp seals with PTFE/silicone septa were obtained from Sigma-Aldrich Co. LLC. (St.

Louis, MO). The clean soil and certified Aroclor 1254 and 1260 soil samples were purchased from RT Corp (Laramie, WY).

The standard solutions for Aroclor 1254 and 1260 at concentrations of 0.1, 0.3, 1, 3, and 10 $\mu\text{g mL}^{-1}$ were prepared by dilution of 100 $\mu\text{g mL}^{-1}$ Aroclor 1254 and 1260 stock solutions with methanol. The internal standard solution containing 10 $\mu\text{g mL}^{-1}$ deca-CB and TCMX was prepared by dilution of aliquots of the 1 mg mL^{-1} stock solution with hexane. The standard soil samples for calibration were prepared by thoroughly mixing 50 μL of standard solutions with 0.5 g masses of clean soil and evaporating the solvent in a hood at room temperature until the soil became free-flowing and powdery. The concentrations of Aroclors in the prepared soil samples were calculated as follows:

$$C_{\text{soil}} = \frac{C_{\text{solution}} \times V_{\text{solution}}}{M_{\text{soil}}} \quad (4)$$

for which C_{solution} is the concentration of Aroclor standard solutions; V_{solution} is 50 μL and M_{soil} is the mass of the soil which was precisely weighed to 0.5 g in our experiment. To determine the two Aroclors simultaneously by PLS, the calibration matrix for a 6×6 mixture designed of Aroclor 1254 and Aroclor 1260 was constructed. A saturated potassium dichromate solution was prepared by dissolving an excess of potassium dichromate in 6.0 M sulfuric acid.

3.2. Instruments

Analyses were conducted on a Thermo Finnigan PolarisQ quadrupole ion trap mass spectrometer/Trace GC system with a Triplus AS2000 autosampler (San Francisco, CA, USA) which was controlled using the XCalibur software version 2.0.7 provided by Thermo. Analytes were separated using a 30 m×0.25 mm×0.1 µm 5% diphenyl/95% dimethylpolysiloxane cross-linked capillary column (SHRXI-5MS, Shimadzu Scientific Instruments Inc. Columbia, MD). The temperature program was 50 °C, hold for 1 min, ramp at 20 °C min⁻¹ to 280 °C, hold for 10 min. Helium was used as the carrier gas, with a constant flow rate of 1 mL min⁻¹. The transfer line and ion source temperatures were both maintained at 280 °C. Mass spectra were obtained with electron ionization mode (40 eV) in the range from mass-to-charge ratio (m/z) 140 to 550 starting after a 4 min solvent delay. All the data were processed using MATLAB R2012b (MathWorks Inc., Natick, MA).

3.3. Sample preparation

Extraction of PCBs from soil samples under optimized conditions were conducted using headspace SPME with 2 mL of saturated potassium dichromate solution and a 100 µm PDMS fiber at 100 °C for 30 min (Zhang et al., 2014). An aliquot of 0.5 g of soil sample was placed in a 20-mL SPME vial with the addition of 20 µL of internal standard solution. After the solvent was evaporated in the hood at room temperature, 2 mL of the saturated potassium dichromate

solution was added to the vial and the vial was sealed by a headspace aluminum cap with a polytetrafluoroethylene (PTFE)/silicone septum. After the sample vial was incubated at 100 °C for 0.5 min, a PDMS fiber was exposed to the headspace for 30 min with the agitation setting on for 10 s and then off for 10 s sequentially. The fiber was then immediately inserted into the GC injector for desorption at 280 °C for 5 min after which the analytes were analyzed by GC/MS.

3.4 Data Format.

The RAW two-way GC/MS data (retention time × mass-to-charge ratio) were initially acquired, and were converted to the network common document format (CDF) with the XCalibur Software 'File Converter Tool'. The CDF files were read directly into MATLAB using the netcdf tools.

Each two-way data was binned by retention time from 4.10 to 22.00 min with a 0.01 min increment and binned by mass-to-charge ratios from 140 to 550 Th with a 1 Th increment. Therefore, each two-way GC/MS object comprised 1801 × 411 data points in which 1801 rows corresponded to the retention times and 411 columns corresponded to mass-to-charge ratios. The extracted ion chromatograms (EIC) data were extracted from the two-way data at a retention time window between 11.0 and 16.0 min for five PCB molecular ions (m/z 292, 326, 360, 394, and 430). Therefore, each

EIC data comprised 499×5 data points. Each two-way and EIC data object was unfolded to a vector by concatenating the rows of the two-way object along the retention time way. For the calibration set, 36 samples were collected, so the two-way data calibration set ($36 \times 1801 \times 411$) was unfolded to 36×740211 and the EIC data calibration set ($36 \times 499 \times 5$) was unfolded to 36×2495 .

4. Results and discussion

The optimization of the SPME-GC/MS method was discussed in another publication (Zhang et al., 2014). Aroclor 1260 standard was used to evaluate the optimization of the SPME and instrument conditions. The peak areas of hexa-chlorinated biphenyls (hexa-CBs) were representative for the comparison of the effects of different conditions because hexa-CBs are one of the major PCBs in both Aroclor 1254 (26.9 weight %) and Aroclor 1260 (46.9 weight %) (Frame et al., 1996). Fig. 2 comprises the chromatograms obtained from Aroclor 1254 and 1260. Because of the high similarity of both Aroclors, none of the PCB peaks can be used as characteristic peaks for either Aroclor. However the ratios of different PCBs in each Aroclor vary. For this case, the multivariate PLS method is suitable.

4.1. PLS method

Orthogonal baseline correction was applied prior to data processing which reduced the artifact peaks (e.g., PDMS peaks) and corrected the

226 baselines for all the TIC chromatograms (Zhang and Harrington, 2013).
 227 The molecular ion of deca-CB (m/z 498) in each sample was extracted
 228 from TIC and the peak area of deca-CB was integrated between 19.3
 229 to 19.6 min. Then the intensities at different masses in each
 230 chromatogram were normalized to the area of its deca-CB peak.

Fig. 2. GC/MS total ion current (TIC) chromatograms of Aroclor 1254 (A) and Aroclor 1260 (B) for $300 \mu\text{g kg}^{-1}$ soil sample after headspace SPME extraction. On the right of each TIC chromatogram is zoom-in retention time window of PCB peaks for each Aroclor.

231 The calibration set comprised 36 standard mixtures for which the
 232 concentrations were 10, 30, 100, 300, and $1000 \mu\text{g kg}^{-1}$ for each

Aroclor. The concentration matrix (36×2) for the calibration set for which each column contains the concentrations of the corresponding Aroclor is given in the supplementary material. Other than the original two-way GC/MS data set, extracted ion chromatograms (EICs) at retention time windows (RTWs) between 11.0 to 16.0 min were used to construct smaller two-way GC/MS data sets that were selective for the PCBs. The molecular ions of tetra-chlorinated biphenyls (tetra-CB, m/z 292), penta-chlorinated biphenyls (penta-CB, m/z 326), hexa-CB (m/z 360), hepta-chlorinated biphenyls (hepta-CB, m/z 394), and octa-chlorinated biphenyls (octa-CB, m/z 430) were used to create the EIC two-way data. Three EIC data sets were investigated: 1) the EIC-A54 data set includes m/z 292, 326, and 360 for tetra-, penta-, and hexa-CBs that account for 94% of the PCBs in Aroclor 1254; 2) the EIC-A60 data set includes m/z 360, 394, and 430 for tetra-, penta-, and hexa-CBs that account for 90% of the PCBs in Aroclor 1260; 3) the EIC-Both data set includes all five PCBs mentioned above which represent more than 98% of the PCBs in both Aroclor 1254 and Aroclor 1260 (Frame et al., 1996).

To select the number of components for the PLS-2 model and compare the different data sets, a cross-validation method was used that leaves out one sample at a time (i.e., one design point). Fig. 3 is a plot of the RMSPE with respect to the number of PLS components for

the EIC-Both data. The summed prediction error for both Aroclor 1254 and 1260 decreased for the first 8 principal components and increased thereafter. Therefore number of components was set to 8 for PLS-2 model.

Fig. 3. Plot of the RMSPE with respect to the number of PLS components for the EIC-Both data.

From Fig. 3, it is apparent that the optimal component numbers for the two Aroclors are different (e.g., 7 for Aroclor 1254 and 9 for Aroclor 1260). To investigate the effect of the number of PLS components, PLS (PLS-1) was used because it models each Aroclor separately.

The results for the optimal component numbers, linear regression coefficients, RMSPE and RPEs for univariate PLS (PLS-1) and multivariate PLS (PLS-2) using different data representations are

reported in Table 1. For Aroclor 1254, the coefficient of determination (r^2), RMSPE, and RPEs results for the EIC-A60 data are the worst among the different data representations. This result is reasonable because hepta-CB (m/z 394) and octa-CB (m/z 430) only account for less than 4% of the PCBs in Aroclor 1254 and those PCB signals can only be detected in calibration samples at the higher concentrations, so they result in larger errors. For similar reasons, the results for Aroclor 1260 using EIC-A54 data sets are slightly worse than the other data representations because tetra-CB (m/z 292) and penta-CB (m/z 326) only account for less than 10% of the PCBs in Aroclor 1260 and cannot be detected in calibration samples at the lower concentrations.

Table 1

Comparison of PLS-1 and PLS-2 using calibration set by cross validation

	EIC-Both ^a				EIC-A60			
	A1254		A1260		A1254		A1260	
	PLS-1	PLS-2	PLS-1	PLS-2	PLS-1	PLS-2	PLS-1	PLS-2
	Num_comp ^b							
r^2	5	8	7	8	9	8	7	8
RMSPE ($\mu g\ kg^{-1}$)	0.989	0.988	0.992	0.984	0.908	0.894	0.993	0.989
RPEs (%)	0.0378	0.0373	0.0327	0.0348	0.1078	0.1178	0.0289	0.0306
	10.52	10.37	9.09	9.68	29.96	32.73	8.04	8.49
	EIC-A54				Two-way			
	A1254		A1260		A1254		A1260	
	PLS-1	PLS-2	PLS-1	PLS-2	PLS-1	PLS-2	PLS-1	PLS-2
	Num_comp							
r^2	5	8	10	8	4	8	10	8
RMSPE ($\mu g\ kg^{-1}$)	0.990	0.988	0.990	0.983	0.993	0.991	0.991	0.987
RPEs (%)	0.0363	0.0377	0.0348	0.0355	0.0297	0.0337	0.0331	0.0335
	10.08	10.46	9.68	9.85	8.25	9.36	9.19	9.30

^aEIC-Both denotes extracted ion chromatogram (EIC) data sets which includes m/z 292, 326, 360, 394, and 430; EIC-A60 denotes EIC data sets which includes m/z 360, 394, and 430; EIC-A54 denotes EIC data sets which includes m/z 292, 326, and 360; Two-way denotes two-way data sets.

^bNum_comp denotes number of components used in PLS models.

278

279 The optimal component numbers for PLS-2 models using different
280 data representations are consistent but those for PLS-1 models using
281 different data representations varies. Comparing the r^2 , RMSPE, and
282 RPEs results between PLS-1 and PLS-2 models, overall the PLS-1
283 results are better than PLS-2 except that for Aroclor 1254 using the
284 EIC-Both data representation, the prediction errors of the PLS-1 model
285 are slightly larger than the results of the PLS-2 model.

286 It is worth mentioning that any single EIC (i.e., m/z 292 for tetra-
287 CB) data may be used to construct the multivariate calibration.
288 However some masses result in substantial errors (see supplementary
289 material). It is risky to use only one mass for the whole mixture
290 because the selected PCBs could be relatively low in intensity
291 compared with other PCBs and they also could be highly overlapped
292 between different Aroclors. Moreover it would be difficult to predict
293 which mass is the best to use.

294 4.2. Validation of method by certified soil and real soil samples

295 Certified soil samples that contained Aroclor 1254 at $5.93 \mu\text{g g}^{-1}$
296 (prediction interval $2.45\text{--}9.41 \mu\text{g g}^{-1}$) and Aroclor 1260 at $1.50 \mu\text{g g}^{-1}$
297 (prediction interval $0.65\text{--}2.34 \mu\text{g g}^{-1}$) were analyzed. The certified soil
298 samples were diluted with certified clean soils in accord with the
299 calibration range and mixed to furnish 8 soil samples with different

concentrations of Aroclor 1254 and 1260. Each soil sample was analyzed by SPME-GC/MS in five replicates. The same PLS-1 and PLS-2 models using their optimal component numbers for the 4 different data representations were applied with a validation set of certified Aroclor soil samples. The prediction rates are given in Table 2 and the full details of samples and prediction results are reported in the supplementary material.

Table 2

Comparison of prediction rates between PLS-1 and PLS-2 for certified soil samples (number of samples = 8).

Data representation ^a		Prediction Rate (%)	
		PLS-1	PLS-2
EIC-Both	A1254	100	100
	A1260	100	100
EIC-A60	A1254	62.5	62.5
	A1260	100	100
EIC-A54	A1254	100	100
	A1260	100	87.5
Two-way	A1254	75	62.5
	A1260	100	75

^aEIC-Both denotes extracted ion chromatogram (EIC) data sets which includes *m/z* 292, 326, 360, 394, and 430; EIC-A60 denotes EIC data sets which includes *m/z* 360, 394, and 430; EIC-A54 denotes EIC data sets which includes *m/z* 292, 326, and 360; Two-way denotes two-way data sets.

For the PLS-1 model, 100% prediction rates were achieved for both Aroclor 1254 and 1260 by using EIC-Both and EIC-A54 data representations, and for Aroclor 1260 by using EIC-A60 and two-way data representations. The predicted concentrations for three samples of Aroclor 1254 using EIC-A60 and two-way data are outside of the

prediction intervals. The large prediction error for Aroclor 1254 using EIC-A60 data is also found in Table 1 (RPEs = 29.96%). Both results indicate that the EIC-A60 data representation is not suitable for predicting the concentration of Aroclor 1254. The concentrations for two samples of Aroclor 1254 using two-way data did not fall into the prediction interval ranges. An explanation could be that the two-way data is more sensitive to background variations that arose from the soil samples. Although the baseline correction was applied to all the data sets prior to data processing, the soils used for the background correction may not have matched well with the soil from the real and certified soil samples. However, for the other three EIC data sets, the selected PCB ions were extracted from two-way data sets, so they are more selective and more resistant to background peaks.

For the PLS-2 model, all the concentrations obtained for Aroclor 1254 using EIC-A54 data sets, for Aroclor 1260 using EIC-A60 data sets and for both Aroclors using EIC-Both data sets are in agreement with the certified reference values and fell inside the certified range. Predicted concentrations for the EIC-A60 data for three samples of Aroclor 1254 were outside the prediction intervals and one sample for the EIC-A54 data for Aroclor 1260 was outside of the prediction interval range, which indicates that EIC-A60 and EIC-A54 data sets may not be suitable for Aroclor 1254 and Aroclor 1260, respectively.

These results agree with the PLS cross-validation results (Table 1). All the soil samples at low concentrations ($50 \mu\text{g kg}^{-1}$ for either Aroclor 1254 or 1260 or both) failed to predict accurately using the two-way data sets which might be caused by background variation as well. Overall, the prediction results for certified soil samples from the PLS-1 models are better than those results from the PLS-2 models. However, PLS-2 models built from EIC-Both, EIC-A54, and EIC-A60 data representations have demonstrated the capability to predict the concentrations of Aroclor 1254 or/and Aroclor 1260.

The real soil samples were collected at the Portsmouth Gaseous Diffusion Plant, Ohio. The samples were split for analysis between our laboratory and GEL Laboratories, LLC (GEL), Charleston, South Carolina. Five replicates were performed on each soil sample and the same soil samples were analyzed by GEL using EPA 8082 method. As given in Table 3, the estimated concentrations for both Aroclor 1254 and 1260 using EIC-A54, EIC-A60, and EIC-Both data sets are in agreement with the results from GEL using EPA 8082 method by PLS-2 models. All samples were outside the prediction intervals using the two-way data by PLS-2 model, which again might be caused by novel peaks arising from the soil background. PLS-1 models had better predictions for both Aroclors using two-way data but worse performance using EIC-Both and EIC-A60 data. The full prediction

results are given in supplementary material. Together with the prediction results for certified soil samples (Table 2), PLS-1 models were better for the EIC-A54 and the two-way data representations and PLS-2 models were better for the EIC-Both and the EIC-A60 data representations.

Table 3

Comparison of prediction rates between PLS-1 and PLS-2 for real soil samples (number of samples = 3).

Data representation ^a		Prediction Rate (%)	
		PLS-1	PLS-2
EIC-Both	A1254	100	100
	A1260	33	100
EIC-A60	A1254	67	100
	A1260	100	100
EIC-A54	A1254	100	100
	A1260	33	33
Two-way	A1254	100	33
	A1260	100	0

^aEIC-Both denotes extracted ion chromatogram (EIC) data sets which includes *m/z* 292, 326, 360, 394, and 430; EIC-A60 denotes EIC data sets which includes *m/z* 360, 394, and 430; EIC-A54 denotes EIC data sets which includes *m/z* 292, 326, and 360; Two-way denotes two-way data sets.

Therefore, the EIC-A54 data sets predicted well for the Aroclor 1254 with PLS-2 regression; EIC-A60 data sets predicted well for the Aroclor 1260 with both PLS-1 and PLS-2 regression; the EIC-Both data predicted the concentrations of both Aroclors well with PLS-2 models; and the comprehensive two-way data predicted both Aroclors poorly with the PLS-2 regression but predicted Aroclor1260 well with the PLS-

1 regression. In practice, PLS-2 regression is more parsimonious and simpler because only one calibration model is used.

4.3. Detection limit determination

Several approaches have been reported for the determination of detection limit in multivariate calibration procedures (Boque and Rius, 1996). In this study, the detection limit has been calculated using three times the standard deviation of the predicted concentrations for blank samples. Four blank soil samples with internal standard were treated the same as those described in section 3.3. *Sample preparation.* The EIC-Both data sets for blank samples were constructed. From the PLS-2 modeling using EIC-Both data for each Aroclor, the concentrations for blank samples were predicted. Then three times the standard deviation of the predicted concentrations for each Aroclor was taken as detection limit. The detection limits for Aroclor 1254 and 1260 are $4 \mu\text{g kg}^{-1}$ and $6 \mu\text{g kg}^{-1}$, respectively.

5. Conclusion

In this study, a multivariate linear regression method, PLS, has been developed and used to determine Aroclor 1254 and 1260 mixtures in soil samples. The use of the PLS method enabled the analysis of multiple Aroclors in soil without the complex sample preparation and tedious instrumental separation. Four GC-MS data representations, SIC-A54, SIC-A60, SIC-Both, and two-way data were

evaluated for predicting the Aroclor concentrations by PLS-1 and PLS-2 models. The selected ion two-way (SIC) data sets were found to be more selective and resistant to background variance. The method has been validated by commercial certified soil samples and real on-site soil samples.

The predicted concentrations of Aroclor 1254 by SIC-A54 data sets, the predicted concentrations of Aroclor 1260 by SIC-A60 data sets and the predicted concentrations of both Aroclors by SIC-Both data sets were all in the prediction intervals for the certified soil and real soil samples. The main advantage of our method over previous reported approaches is the high sample throughput, with a soil sample being processed about every 35 min. Because the SPME method requires only 0.5 g soil, the difficulties of sampling, homogenizing, shipping, and storing the soil samples for laboratory analysis are greatly reduced. The uses of nonorganic solvents and short analysis time also increase the applicability of the developed method for on-site analysis (Zhang et al., 2013).

Furthermore, instead of profiling a few marker compounds found in complex mixtures (e.g., benzene, toluene, ethylbenzene, and xylene in gasoline), a holistic approach to mixture characterization has been furnished. For environmental samples (e.g., soil) mixtures of complex samples can be differentiated and determined, which has broad

application in environmental forensic applications, site remediation,
and modeling pollutant dynamics and sources.

Acknowledgements

The research is funded by a grant from US Department of Energy,
Office of Environmental Management, Portsmouth/Paducah Project
Office.[#] The Center for Intelligent Chemical Instrumentation and
Department of Chemistry and Biochemistry at Ohio University are
acknowledged for the financial support. We thank Dr. Natalie A. Kruse
and Jennifer R. Bowman for providing the real soil samples. Xinyi
Wang is thanked for her comments and suggestions.

*[#]The project was supported by US Department of Energy. Any
opinions, findings, and conclusions or recommendations expressed in
this material are those of the author(s) and do not necessarily reflect
the views of the US Department of Energy Office of Environmental
Management Portsmouth/Paducah Project Office, or of the Voinovich
School of Leadership and Public Affairs at Ohio University.*

References

An, L., Deng, J., Zhou, L., Li, H., Chen, F., Wang, H., Liu, Y.T., 2010.
Simultaneous spectrophotometric determination of trace amount of
malachite green and crystal violet in water after cloud point

445 extraction using partial least squares regression. J. Hazard. Mater.
446 175, 883-888.

447 Boque, R., Rius, F.X., 1996. Multivariate detection limits estimators.
448 Chemometr. Intell. Lab. 32, 11-23.

449 Bruzzoniti, M.C., Maina, R., Tumiatti, V., Sarzanini, C., Rivoira, L., De
450 Carlo, R.M., 2012. Fast low-pressure microwave assisted extraction
451 and gas chromatographic determination of polychlorinated
452 biphenyls in soil samples. J. Chromatogr. A. 1265, 31-38.

453 EPA, 2000. Polychlorinated Biphenyls (PCBs) by Gas Chromatography.
454 EPA method 8082A. U.S. Environmental Protection Agency,
455 Washington, DC.

456 EPA, 2010. Method 1668C: Chlorinated biphenyl congeners in water,
457 soil, sediment, biosolids, and tissue by HRGC/HRMS. EPA Method
458 1668C. U.S. Environmental Protection Agency, Washington, DC.

459 Frame, G.M., Wagner, R.E., Carnahan, J.C., Brown, J.F., May, R.J.,
460 Smullen, L.A., Bedard, D.L., 1996. Comprehensive, quantitative,
461 congener-specific analyses of eight aroclors and complete PCB
462 congener assignments on DB-1 capillary GC columns. Chemosphere
463 33, 603-623.

464 Geladi, P., Kowalski, B.R., 1986. Partial Least-Squares Regression - a
465 Tutorial. Anal. Chim. Acta 185, 1-17.

- 466 Inon, F.A., Garrigues, J.M., Garrigues, S., Molina, A., de la Guardia, M.,
467 2003. Selection of calibration set samples in determination of olive
468 oil acidity by partial least squares-attenuated total reflectance-
469 Fourier transform infrared spectroscopy. *Anal. Chim. Acta* 489, 59-
470 75.
- 471 Li, B.B., Morris, J., Martin, E.B., 2002. Model selection for partial least
472 squares regression. *Chemometr. Intell. Lab. 64*, 79-89.
- 473 Newman, J.W., Becker, J.S., Blondina, G., Tjeerdema, R.S., 1998.
474 Quantitation of Aroclors using congener-specific results. *Environ.*
475 *Toxicol. Chem.* 17, 2159-2167.
- 476 Safe, S., 1984. Polychlorinated biphenyls (PCBs) and polybrominated
477 biphenyls (PBBs): biochemistry, toxicology, and mechanism of
478 action. *Crit. Rev. Toxicol.* 13, 319-395.
- 479 Sjostrom, M., Wold, S., Lindberg, W., Persson, J.A., Martens, H., 1983.
480 A Multivariate Calibration-Problem in Analytical-Chemistry Solved
481 by Partial Least-Squares Models in Latent-Variables. *Anal. Chim.*
482 *Acta* 150, 61-70.
- 483 Wold, S., Ruhe, A., Wold, H., Dunn, W.J., 1984. The Collinearity
484 Problem in Linear-Regression - the Partial Least-Squares (PLS)
485 Approach to Generalized Inverses. *Siam J. Sci. Stat. Comp.* 5, 735-
486 743.

487 Zhang, M., Harrington, P.d.B., 2013. Automated pipeline for classifying
488 Aroclors in soil by gas chromatography/mass spectrometry using
489 modulo compressed two-way data objects. *Talanta* 117, 483-491.

490 Zhang, M., Harrington, P.d.B., Kruse, N.A., Bowman, J.R., Lammert,
491 S.A., Lee, E.D., Jackson, G.P., 2013. Development of an expedited
492 field study method for PCBs in sediments and soils using portable
493 GC/MS. 61st ASMS Conference on Mass Spectrometry and Allied
494 Topics, Minneapolis, MN.

495 Zhang, M., Jackson, G.P., Kruse, N.A., Bowman, J.R., Harrington,
496 P.d.B., 2014. Determination of Aroclor 1260 in Soil Samples by Gas
497 Chromatography with Mass Spectrometry with Solid-Phase
498 Microextraction. *J. Sep. Sci.* doi: 10.1002/jssc.201400102.

499

Table 1

Comparison of PLS-1 and PLS-2 using calibration set by cross validation

	EIC-Both ^a				EIC-A60			
	A1254		A1260		A1254		A1260	
	PLS-1	PLS-2	PLS-1	PLS-2	PLS-1	PLS-2	PLS-1	PLS-2
	Num_comp ^b							
	5	8	7	8	9	8	7	8
r^2	0.989	0.988	0.992	0.984	0.908	0.894	0.993	0.989
$RMSPE (\mu g\ kg^{-1})$	0.0378	0.0373	0.0327	0.0348	0.1078	0.1178	0.0289	0.0306
$RPEs (\%)$	10.52	10.37	9.09	9.68	29.96	32.73	8.04	8.49
	EIC-A54				Two-way			
	A1254		A1260		A1254		A1260	
	PLS-1	PLS-2	PLS-1	PLS-2	PLS-1	PLS-2	PLS-1	PLS-2
	Num_comp							
	5	8	10	8	4	8	10	8
r^2	0.990	0.988	0.990	0.983	0.993	0.991	0.991	0.987
$RMSPE (\mu g\ kg^{-1})$	0.0363	0.0377	0.0348	0.0355	0.0297	0.0337	0.0331	0.0335
$RPEs (\%)$	10.08	10.46	9.68	9.85	8.25	9.36	9.19	9.30

^aEIC-Both denotes extracted ion chromatogram (EIC) data sets which includes m/z 292, 326, 360, 394, and 430; EIC-A60 denotes EIC data sets which includes m/z 360, 394, and 430; EIC-A54 denotes EIC data sets which includes m/z 292, 326, and 360; Two-way denotes two-way data sets.

^bNum_comp denotes number of components used in PLS models.

Table 2

Comparison of prediction rates between PLS-1 and PLS-2 for certified soil samples (number of samples = 8).

Data representation ^a		Prediction Rate (%)	
		PLS-1	PLS-2
EIC-Both	A1254	100	100
	A1260	100	100
EIC-A60	A1254	62.5	62.5
	A1260	100	100
EIC-A54	A1254	100	100
	A1260	100	87.5
Two-way	A1254	75	62.5
	A1260	100	75

^aEIC-Both denotes extracted ion chromatogram (EIC) data sets which includes m/z 292, 326, 360, 394, and 430; EIC-A60 denotes EIC data sets which includes m/z 360, 394, and 430; EIC-A54 denotes EIC data sets which includes m/z 292, 326, and 360; Two-way denotes two-way data sets.

Table 3

Comparison of prediction rates between PLS-1 and PLS-2 for real soil samples (number of samples = 3).

Data representation ^a		Prediction Rate (%)	
		PLS-1	PLS-2
EIC-Both	A1254	100	100
	A1260	33	100
EIC-A60	A1254	67	100
	A1260	100	100
EIC-A54	A1254	100	100
	A1260	33	33
Two-way	A1254	100	33
	A1260	100	0

^aEIC-Both denotes extracted ion chromatogram (EIC) data sets which includes m/z 292, 326, 360, 394, and 430; EIC-A60 denotes EIC data sets which includes m/z 360, 394, and 430; EIC-A54 denotes EIC data sets which includes m/z 292, 326, and 360; Two-way denotes two-way data sets.

Fig. 1. Flowchart of PLS as a multiple linear regression method. **T** is the latent two-way GC/MS data matrix, **W** is the two-way GC/MS data loading matrix, **U** is the latent concentration matrix, **Q** represents the concentration loading matrix, E_x , E_y , and E_d are residual error matrices.

Fig. 2. GC/MS total ion current (TIC) chromatograms of Aroclor 1254 (A) and Aroclor 1260 (B) for 300 $\mu\text{g kg}^{-1}$ soil sample after headspace SPME extraction. On the right of each TIC chromatogram is zoom-in retention time window of PCB peaks for each Aroclor.

Fig. 3. Plot of the RMSPE with respect to the number of PLS components for the EIC-Both data.

Figure captions:

Fig. 1. Flowchart of PLS as a multiple linear regression method. T is the latent two-way GC/MS data matrix, W is the two-way GC/MS data loading matrix, U is the latent concentration matrix, Q represents the concentration loading matrix, E_X , E_Y , and E_d are residual error matrices.

Fig. 2. GC/MS total ion current (TIC) chromatograms of Aroclor 1254 (A) and Aroclor 1260 (B) for 300 $\mu\text{g kg}^{-1}$ soil sample after headspace SPME extraction. On the right of each TIC chromatogram is zoom-in retention time window of PCB peaks for each Aroclor.

Fig. 3. Plot of the RMSPE with respect to the number of PLS components for the EIC-Both data.

Graphical abstract