EPA Public Access

Author manuscript

Environ Nanotechnol Monit Manag. Author manuscript; available in PMC 2020 January 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Environ Nanotechnol Monit Manag. 2019 ; 11: 1–100204. doi:10.1016/j.enmm.2018.100204.

Multivariate Calibration for Carbon Nanotubes in the Environment Using the Microwave Induced Heating Method

Yang He^a, Souhail R. Al-Abed^{b,*}, Dionysios D. Dionysiou^a

^aEnvironmental Engineering and Science program, Department of Chemical and Environmental Engineering, University of Cincinnati, 2600 Clifton Ave., Cincinnati, Ohio 45221, United States

^bNational Risk Management Research Laboratory, U.S. Environmental Protection Agency, 26 W. Martin Luther King Dr. Cincinnati, Ohio 45268, United States

Abstract

The goal of the present paper is to develop chemometrics-based multivariate calibration approaches for simultaneously determining quantity of individual carbon nanotubes (CNTs) in a multicomponent environmental matrix using a microwave induced heating method. A multifactor and multilevel experiment design was used to create 4 separate calibration datasets. Each calibration dataset contained 25 orthogonal CNTs with 2 or 3 factors (CNTs: single-walled CNTs (SWCNTs)/multi-walled CNTs (MWCNTs)/carboxylated MWCNTs (MWCNT-COOH)) and 5 levels (CNTs mass). The temperature rise (ΔT) spectral information was obtained for each sample by exposing to varying microwave conditions. This study showed the potential and applicability of partial least square regression (PLS), least square-support vector machine (LS-SVM) and artificial neural networks (ANN) in predicting quantities of SWCNTs, MWCNTs and MWCNT-COOH in environmental matrices with microwave induced temperature rises data. Our results revealed that the developed LS-SVM model presented higher R^2 and lower root mean square error of prediction (RMSEP) ($R^2 = 0.74\text{--}0.93$, RMSEP = 0.0251 mg to 0.0328 mg in 2-component systems and $R^2 = 0.64\text{--}0.95$, RMSEP = 0.0243 mg to 0.0410 mg in 3-component systems), while the ANN model was only accurate in estimating mass of SWCNT and MWCNT in a 2-component mixture ($R^2 = 0.77\text{--}0.89$, RMSEP = 0.0322 mg to 0.0503 mg). The PLS model was found not effectively interpret relationship between microwave induced temperature rises data and mass of CNTs, indicated by small R^2 (0.20–0.87) and large RMSEP (0.0209 mg –0.1021 mg).

Introduction

During the past decades, worldwide production capacity of carbon nanotubes (CNTs) has increased at least ten-fold, and the annual number of CNT-related research and applications continues to increase.¹ In spite of fascinating properties of the nanomaterials, a growing concern about the environmental occurrence of CNTs and potential hazardous effects of CNTs on the ecosystem and human health has arisen.² It has been reported that inhalation of

*Corresponding author: al-abad.souhail@epa.gov.

Conflict of interest

The authors declare no conflict of interest.

0.3–5 mg/m³ multi-walled CNTs (MWCNTs) could disrupt systemic immune function³ and 25 µg/ml single-walled CNTs (SWCNTs) could cause pulmonary damage.⁴ To address the concern regarding the CNTs, the quantitative information of the CNTs in environmental media is of great important for assessing environmental impact or biological stress of CNTs. Therefore, several methods have recently been developed for environmental quantification of the CNTs, such as ultraviolet-visible (UV-vis) spectrophotometer,^{5, 6} near-infrared fluorescence technique,^{7, 8} and programmed thermal analysis.^{9–12} It is worth noting that all methods are designed for the environmental matrix containing only single type of CNTs. However, the likelihood of presence of a mixture of different types of CNTs in the environment is not small because various types of CNTs are utilized for different purposes. For examples, SWCNTs are popular for field effect transistor architectures due to their superiority of strength and electrical conductivity to MWCNTs.¹³ The functionalized MWCNTs (e.g. carboxylate MWCNTs (MWCNT-COOH)) are suitable to be incorporated into polymer composition due to their better compatibility with polymer. The applications with various types of CNTs underscore the need for methods capable of measuring specific type of CNTs from a mixture of CNTs.

The quantification of target analytes in a multicomponent matrix remains challenge in many research areas owing to sample matrix interference and spectral interference.¹⁴ The matrix interference is that instrument signal intensity of one element is not only determined by its own mass/concentration, but also influenced by the overall composition of samples. The spectral interference refers to overlapping of the spectrum of different components, which makes the quantification more complicated. To minimize the sample matrix and spectral influences, some pre-treatment processes are conducted to isolate the interested compounds from a mixture before quantitative measurement, but most of these treatments are time-consuming and expensive. To address this issue, Chemometrics and multivariate calibration techniques have proven to be beneficial for simultaneous determination of specific compounds in a multi-component system without a prior separation procedure. Instead of using instrument signal under one selected condition, Chemometrics and multivariate calibration methods aim to relate spectral instrument responses of mixed samples to concentration or mass of target elements using statistical techniques.¹⁵ In recent years, the multivariate calibration approaches have been extensively employed in analytical chemistry,^{16–19} biological studies,²⁰ food science^{21–23} and quality control of drugs.²⁴ Since 2000, the Chemometrics and multivariate calibration methods have been expanded to environmental studies, mainly focusing on quantification of a wide range of pollutants in various environmental compartments.^{25–27} Among different multivariate regression methods, the factor analysis-based methods, such as partial least squares (PLS) regression and principal components regression (PCR), received considerable attention for spectroscopic data analysis in environmental research as they can deal with highly multicollinear experimental data that are common in environmental detection. PLS was employed to quantify different pesticides in river water and vegetable samples with differential pulse polarography and near infrared (NIR) spectroscopy data.^{28, 29} Since the basic concept of PLS is to establish a linear relationship between the spectral data and the property to be determined (e.g. concentration), PLS is not valid for situations where the instrument signal of a mixed sample is not a linear summation of signal of individual analyte. Thus, a support vector machine (SVM) method

was introduced to perform nonlinear regression for environmental analysis.³⁰ SVM was successfully implemented to resolve and quantify biodiesel contents in diesel fuel blends with NIR spectra³¹ and to distinguish the variety of soil with laser-induced breakdown spectroscopy.³² Moreover, artificial neural network (ANN) has been reported to be able to build highly complex nonlinear relationships between inputs and outputs without assuming a specific frequency distribution or independence of the input variables.³³ As a result, ANN model has been involved in some environmental studies as well. For example, ANN model was successfully applied to provide resolutions of a highly overlapping ternary dye mixture with UV-vis spectra,³⁴ and the quantification of anthracene, phenanthrene and naphthalene in drinking water was achieved using ANNs with fluorescence emission spectroscopy.³⁵

So far, such multivariate calibration methods have not been applied with existing quantification approaches to determine the environmental concentration of CNTs in a mixture. Among current quantification methods for CNTs, some methods are not able to generate spectral data for a sample. For instance, the thermal based quantification methods directly measure sample mass loss as the mass of CNTs when temperature increases above the degradation temperature of CNTs, so no spectral data can be obtained. Although spectral data can be obtained for samples with fluoresces or UV-vis based approaches, these techniques have some limitations for multivariate analysis in a multicomponent system. Only well dispersed SWCNTs have fluoresces signals and MWCNTs or functionalized CNTs do not. Although all types of CNTs have characteristic peaks in the UV-vis spectra, the presence of naturally occurring carbon sources in real environmental matrices has significant influence on the intensity of the CNTs peaks, making the multivariate analysis imprecise.

A recently developed microwave induced heating method is most likely to be applied with multivariate calibration methods because a temperature rise spectrum (ΔT , temperature difference of a sample before and after exposure to the microwave energy) for a CNT loaded sample can be constructed with varying microwave conditions. Moreover, the microwave measurement has been demonstrated not to be affected by the presence of natural and artificial carbon sources in sample because of selective microwave heating of CNTs, which makes less matrix interferences for multivariate analysis.³⁶ The detailed mechanism and advantages of the microwave method have been discussed in our previous study.³⁶ One of potential challenges in integrating the microwave method with the multivariate calibration methods is that the temperatures rises of a sample measured at different microwave conditions may be highly correlated. Another problem could be that although the microwave responses of different types CNTs are different, the spectrum of each type of CNTs heavily overlaps with each other. Additionally, the of single type of CNTs is linearly related to the mass of CNTs, but it is not clear whether the relationship between temperature rises spectrum and mass of CNTs is linear or not when different types of CNTs are present in one sample.

Therefore, the present study aims to exam the application of multivariate calibration techniques to determine quantities of individual CNTs (SWCNTs, MWCNT and MWCNT-COOH) in environmental matrices with the microwave induced temperature rise spectral data. The PLS, SVM and ANN models were used with calibration datasets to build models

for predicting CNTs mass. The goodness of fit of developed models was evaluated and compared in such metrics as R^2 , RMSECV and the prediction capacity of models was assessed by RMSEP in separated testing datasets.

Experimental

Samples Preparation and Microwave Induced Heating System

A full factorial experimental design was employed to prepare environmental samples with various types of CNTs. Four calibration sets, containing various types of CNTs at various mass levels in different environmental samples, were designed for multivariate analysis (Table 1). Each calibration dataset consisted of 25 samples with 2 or 3 factors (e.g. SWCNT/MWCNT or SWCNT/MWCNT/MWCNT-COOH) and 5 levels (CNTs mass in environmental sample). It was an orthogonal experiment design to ensure that each type of CNTs can be quantified independently of other types of CNTs. The detailed experimental design was presented in supplementary information (SI), Table S1. All samples were prepared in duplicate and each sample was measured three times. The average temperature rises were used to construct the spectrum for each sample. In addition to calibration dataset, an independent testing dataset was prepared for corresponding calibration dataset. Each testing dataset consisted of 5 samples, used to assess the prediction ability of resulted models.

The calibration samples were prepared by adding varying aliquot of 0.3 mg/L CNT dispersion to environmental matrix (0.1 g soil or 5 ml sludge), resulting in the mass of each type of CNTs at five levels (0, 0.075, 0.15, 0.225 and 0.3 mg). The testing samples were prepared in the same way and the detailed design was summarized in Table S2. To eliminate matrix influences, all environmental samples used in this study were exposed to elevated temperature (500 °C for 2 hours) to remove organic carbon before use. The detail description and operation of the microwave induced heating system were discussed in our previous paper.³⁶ Briefly, in this study, all samples were exposed to different microwave powers of 40, 60, 80, 100, 120, 133 and 149 W for 2 min, and the temperature rises ($T = \text{temperature after the microwave irradiation} - \text{temperature before the microwave irradiation}$) were recorded during each run. Each calibration dataset contained mass of each type of CNTs and microwave induced temperature rises at different microwave conditions.

Model Development

In this section, PLS, SVM and ANN models were constructed with different calibration datasets, respectively. In each calibration dataset, the microwave induced temperature rises were used as predictors and mass of CNTs was used as response variable. During model training, a cross validation (CV) procedure was used to optimize model parameters and detect overfitting. The goodness-of-fit of resulted PLS, SVM and ANN models was compared based on root mean square error in cross validation (RMSECV) and correlation coefficient (R^2) of CV. The prediction performances of the models were assessed according to root mean square error in prediction (RMSEP) for separate testing datasets that were not involved in the model building step. The equations for RMSECV and RMSEP are defined as Equation 1 (Eq.1):

$$\text{RMSECV or RMSEP} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{Eq.1}$$

Where y_i is the observed values and $\hat{y}_i$ is value of the y variable estimated by CV for RMSECV or the predicted values of y variable for RMSEP. In general, the criterion for a precise model is a higher R^2 value and lower RMSECV and RMSEP values.

PLS

The PLS regression is the most commonly used technique for predicting concentration of interested analytes in a multicomponent system, as it can handle strongly collinear data with only a limited number of observations. Briefly, the PLS model is to establish linear relationship between analytical signal data (X) and property to be determined (Y) by simultaneously decomposing X and Y into two separate sets of scoring and loading.¹⁵ The PLS model can be expressed by an outer relation and an inner relation.¹⁹ The outer relations, describing the X and Y matrices individually, are given as Eq.2 and Eq.3:

$$X = T \cdot P^T + E \quad \text{Eq.2}$$

$$Y = U \cdot C^T + F \quad \text{Eq.3}$$

Where T and U are scoring matrices of X and Y, P^T and C^T are the loading matrices of X and Y and E and F are the error matrices of the X and Y matrices, respectively. The inner relation links X with Y as the following equations (Eq.4 and Eq.5):

$$Y = T \cdot C^T + G \quad \text{Eq.4}$$

$$T = X \cdot W^* \quad \text{Eq.5}$$

Eq.5 shows how X scoring estimates the linear combination of variables with the weight coefficient W^* . Therefore, Y can be calculated according to Eq.6:

$$Y = X \cdot W^* \cdot C^T + G = XB + G \quad \text{Eq.6}$$

Where G is the error matrix and B is the PLS regression coefficient. When the modeling of the PLS with the first component has been completed, then the process with further one continues based on the error matrices (G) until approximately 99% of the variance is explained¹⁹. The use of fewer components than the number of X variables can effectively

reduce the dimension of the regression¹⁹. In the paper, the PLS regression was performed with RStudio with “pls” package. A leave-one-out CV procedure was used to determine the optimal number of PLS components producing the smallest RMSECV.

Least Squares-Support Vector Machine (LS-SVM)

Generally, SVM model develops a nonlinear kernel-based function approximation by mapping the input vectors into high dimensional feature spaces where a special hyperplane is created, and then a regression model is built in the hyperplane.³⁷ The outstanding performance of SVM originates from its capacity of learning in the high-dimensional hyperplane with fewer calibration data. In principle, SVM is to find best relationship that can minimize a cost function, subjecting to a set of inequality constraints.³⁰ However, finding the optimal SVM model can be computationally slow due to sparse solutions. As a simplification, LS-SVM is developed based on the traditional SVM by Suykens *et al* by change the inequality constraints to equality ones.³⁸ A multi-response LS-SVM regression (MLSSVR) algorithm was proposed by Xu *et al.*, and the detailed algorithm of MLSSVR is clearly described in their publication.³⁹ In this paper, the MLSSVR toolbox was implemented with MATLAB to develop the calibration models.³⁹ Briefly, the MLSSVR is an extension of standard LS-SVM by replacing the input vector with a matrix. Similar to the determination of the number of components in PLS, regularized parameters (γ and λ) and hyperparameters (p) of radial basis kernel function were optimized by searching for a smallest RMSECV through a 10-fold CV procedure. γ was optimized in the range of 2^{-5} – 2^{15} , 2^{-10} – 2^{10} for λ and 2^{-15} – 2^3 for p with proper increments. The RMSECV for MLS-SVM is an average of the two or three RMSCEVs of the components determined.

ANN

Basically, an ANN model is composed of a number of neurons (nodes) that are arranged and connected into layers.³³ The first layer is the input layer with one node representing one predictor variable and the last layer is the output layer with one node representing one desired response variable. The hidden layer is between the input and output layers, consisting of a number of nodes for learning the relationship of input and output variables. The connection between the nodes is defined as weight (ω_{ij}),⁴⁰ expressed as Eq.7:

$$\Delta\omega_{ij(n+1)} = \Delta\eta e_j y_j + \alpha \Delta\omega_{ij(n)} \quad \text{Eq.7}$$

Where e_j is the error, y_j is the output signal of node j , η is the learning rate, α is the momentum, and n is the iteration number. The ANN model training is to minimize the errors through repeatedly processing data by changing weight, and the iteration is finished once the error reaches a minimum. In the present work, a multi-layer feed-forward networks algorithm was used to determine parameters for the optimal ANN model in RStudio with “nnet” package. The analytical form of prediction of the target variable in the ANN is generally expressed as Eq.8:⁴¹

$$y_j(n) = \varphi_j \left(\sum_{i=0}^m \omega_{ji}(n) y_i(n) \right) \quad \text{Eq.8}$$

where m is the total number of inputs to node j , ω_{ji} is the weight connected to it and φ is the activation function (sigmoidal function by default in RStudio) at the neuron j at iteration n .

Each node in the hidden layer acts as a processing unit which locally performs transformations of the input data through the activation function, and the flexibility of freely adapting to the data is achieved by adjusting weight. Therefore, two significant parameters need to be identified in an ANN model: the number of nodes in the hidden layer (size, n) and the weight for each node (decay). To avoid overfitting, each calibration dataset (25 samples) was randomly divided into two sets of training and testing with 20 samples (80%) and 5 samples (20%), respectively. To determine the size of models, ANN models with size from 0 to 20 and default weight (0) were built for both the training dataset and the testing dataset. To avoid the local minima of error terms resulted from the gradient based training process employed in ANN, for each size, 10 networks with different random starting points were trained. A plot of mean square error (MSE) values (the average of 10 MSE at each size) as a function of the size of the model was used to determine the appropriate size. From MSE-size figure, an appropriate range of model size was chosen to optimize weight (0, 0.01, 0.001 and 0.0001) in the optimal ANN model.

Results and discussion

Regression by PLS

The PLS regression was applied to process the microwave induced temperature rises data for simultaneous determination of CNTs in the environment. Firstly, the number of components that yielded the smallest RMSECV need to be determined. For calibration dataset 1, as shown in Figure 1 (a), RMSECV dropped dramatically along with the number of components from 0 to 5 and then slightly decreased to the smallest RMSECV at component of 7. The variance in the predictors (temperature rises) and responses (mass of different types of CNTs) explained by the components for the calibration dataset 1 was presented in Table 2. The result shows that PLS model with 7 components can explain up to 93.73% and 85.35% of the variance in mass of MWCNT and SWCNT, respectively and 100% of the variance in predictors as expected. Therefore, a 7-component PLS model was identified for calibration dataset 1. The number of components was equal to the number of variables in the original calibration dataset 1, suggesting no dimensionality reduction occurs. As demonstrated in Figure 1 (b)-(d), the number of component was determined as 5 for MWCNT and 3 for SWCNT in sludge, 7 for MWCNT, 6 for SWCNT and 7 for MWCNT-COOH in soil and 3 for MWCNT, 7 for SWCNT and 1 for MWCNT-COOH in sludge, respectively. However, from Table S3, we observed that the variance of responses explained by the PLS with the optimal number of component was less than 80% for SWCNT in the calibration dataset 2, 3 and 4 and less than 40% for MWCNT-COOH in calibration dataset 3 and 4, indicating that the PLS may not be an appropriate model for processing the microwave induced temperature data.

The comparison of goodness of fit of resulted models (PLS, LS-SVM and ANN) was summarized in Table 3. As observed by Table 3, the R^2 for the PLS models were low, especially for the samples containing 3 types of CNTs. The largest R^2 (0.87) was observed for MWCNT in mixture of 2 types of CNTs in soil, but the corresponding RMSECV was 0.0383 mg that was 43.98% higher than the RMSEP (0.0266 mg) for the calibration datasets. For other calibration datasets, the R^2 was in a range of 0.2–0.72, which was not acceptable to perform prediction for testing dataset. The PLS model based prediction of mass of CNTs were plotted against the actual mass in testing dataset in Figure 2. As shown in Figure 2 (a) and (b), the predicted mass of CNTs in the 2-component sample had a slightly increasing trend with the increasing actual mass, whereas the predicted mass in the 3-component samples randomly scattered with the increasing of actual mass (Figure 2 (c) and (d)). The prediction errors generated by different models for testing datasets were tabulated in Table 4. From Table 4, the largest RMSEP with PLS was 0.1021 mg for MWCNT-COOH in soil, which was 102% higher than the average mass of MWCNT-COOH in the testing dataset (0.1 mg). As a result, PLS model provided poor prediction performance in interpreting the microwave induced temperature data. The possible reason for bad performance of PLS was that PLS was not able to analyze heavy overlapping temperature rises spectra of different types of CNTs by establishing a linear relationship. Therefore, some nonlinear regression models were applied in subsequent sections.

3.2 Regression by LS-SVM

The optimal parameters of LS-SVM models for each calibration dataset were determined by searching for a minimal RMSECV with a 10-folder CV (listed in Table S4) and the fitness and prediction performances of the resulted models were summarized in Table 3 and Table 4. The R^2 was in a range of 0.74–0.93 for models trained with datasets containing two types of CNTs (calibration dataset 1 and 2), which was significantly improved when compared to R^2 of PLS models for the same datasets. With LS-SVM, the R^2 for MWCNT was greater than that for SWCNT. As discussed in our previous study, MWCNTs were more efficient in absorbing microwave energy than SWCNTs³⁶, so the stronger microwave response of MWCNTs made them easily distinguished from a mixture of several types of CNTs. The average RMSECV values (0.0172 mg and 0.0256 mg) were close to the RMSEP of the calibration dataset 1 and 2, no overfitting detected. Moreover, the average RMSECV of LS-SVM was 11.46% and 17.07% of average mass of each CNTs (0.15 mg) in calibration dataset 1 and 2, respectively, indicating the good fitness of LS-SVM model for the microwave data. For samples with three types of CNTs (calibration dataset 3 and 4), the R^2 varied from 0.64 to 0.95, which was also significantly superior to that of PLS models and the R^2 values for MWCNT-COOH were lower than R^2 for MWCNT and SWCNT. The average RMSECV of LS-SVM was 28.53% and 25.07% of average mass of CNTs (0.15 mg) for soil and sludge samples containing three types of CNTs, respectively, which was a little higher than that in 2-component samples.

The prediction ability of the LS-SVM models was assessed by the RMSEP of testing datasets (shown in Table 4). For the samples with two types of CNTs, the RMSEPs of the testing datasets were 0.0281 mg for SWCNT and 0.0251 mg MWCNTs in soil samples, and 0.0328 mg for SWCNTs and 0.0295 for MWCNTs in sludge samples. For the cases with

three types of CNTs in soil and sludge samples, the RMSEP of testing dataset varied from 0.0243 to 0.0410 mg. As illustrated in Figure 3, the LS-SVM-predicted mass of SWCNT and MWCNT was consistent with the actual mass of CNTs in testing datasets, especially in samples with two types of CNTs.

3.3 Regression by ANN

To build ANN models, a training dataset (80%) and a testing dataset (20%) were split from the original calibration dataset to determine the number of nodes in the hidden layer (size, n) and weight for each node (decay). For calibration dataset 1, from Figure 4 (a) and (b), the MSE values of the training dataset for both MWCNTs and SWCNTs dropped dramatically after the size of 1 but increased when the size was larger than 5. The MSE values of the testing datasets were low from size of 3. To reach a compromise between the training and testing datasets, the size ranging from 2 to 5 was chosen to determine the decay value of the model. As shown in Figure 4 (c) and (d), the model with size of 2 yielded the smallest MSE for both MWCNT and SWCNT when decay was 0.01, therefore size of 2 and decay of 0.01 were selected as optimal parameters to build the ANN model for calibration dataset 1. For other calibration datasets, the ANN models were built in the same way and the results were illustrated in Figure S1–S3. Having the optimal parameters of an ANN model, a 10-fold CV procedure was employed to check whether the ANN models were overfitted. As observed in the Table 3, the ratio of RMSECV to RMSEP of the calibration dataset was in a range of 0.81–1.16 for 2-component systems, demonstrating no overfitting. However, the ratio range was 1.27–1.60 for 3-component systems and such larger ratio indicated that ANN models captured not only relationship between the microwave induced temperature rises and mass of CNTs, but also the data noise in the datasets. ANN models are known for adaptability that can change its structure based on input and output data flowing through it, but the flexibility may be the reason for overfitting. The overfitting can lead to inaccurate results when using the model to predict quantity of CNTs. Moreover, the R^2 was 0.82 and 0.89 for SWCNTs and MWCNTs in calibration dataset 1 and 0.77 and 0.83 for SWCNTs and MWCNTs in calibration dataset 2, respectively, slightly lower than R^2 with LS-SVM. For mixtures of 3 types of CNTs, R^2 of ANN models were obviously greater than that of PLS but smaller than that of LS-SVM.

The prediction capacity of ANN models for mixture of two or three types of CNTs in different media were summarized in Table 4. The RMSEPs of the testing dataset were a little larger than those with LS-SVM models for 2-component system. For samples with 3 types of CNTs, the RMSEP was 0.0823 mg and 0.0792 mg for MWCNT-COOH in soil and sludge, respectively, 82.3% and 79.2% higher than the average of MWCNT-COOH mass (0.1 mg). Figure 5 presents the plots between actual mass and ANN-predicted mass for SWCNT, MWCNT and MWCNT-COOH in different media, showing good prediction ability of ANN models for 2-component samples and imprecise prediction for 3-component samples.

Overall, our results show that the ANN based estimates were comparable to LS-SVM in accuracy for the environmental samples containing two types of CNTs, but not as accurate as prediction with LS-SVM for the samples containing three types of CNTs due to

overfitting. The ANN model can be effectively trained according to calibration dataset, but it is poorly generalized beyond its calibration data.¹⁹ Thus, more data need to be gathered to better optimize the ANN model. Among three types of CNTs, the prediction for MWCNTs was most precise, originating from the more significant sensitivity of MWCNTs to the microwave energy than the other two types of CNTs. By comparison to PLS, the better performance of LS-SVM and ANN may be partly attributed from the nonlinear regression function, which is not a feature in PLS. In Figure 6, differences in ΔT represents the difference in temperature rises of soil samples containing MWCNTs and SWCNTs and the summation of temperature rises of equivalent mass of each CNTs. It was observed that the temperature rises of soil samples containing SWCNT and MWCNT were not equal to the addition of individual temperature rise of each CNTs. The differences became larger with the increasing total mass of distinct types of CNTs, which probably resulted from the competition of the microwave energy. The differences for other calibration datasets were illustrated in Figure S4. Although LS-SVM and ANN can process data for which the relationship between the input and output variables are not clearly defined and approximate almost all kinds of nonlinear functions for modeling, one limitation of such artificial intelligent based methods is that they actually use a “black box” approach, which does not offer complete insight into the internal workings of the model for evaluating the interaction of inputs.³⁴ Thus, they are usually used for prediction, not for interpretation of the interactions in the system.

This study shows that the such multivariate calibration methods as LS-SVM and ANN can be applied for determination of quantify of individual CNTs in 2-component environmental samples. The difficulty in quantifying CNTs in the sample containing three or more types of CNTs are most likely attributed to the similarity in temperature rises for samples with different amounts of CNTs. Although MWCNT is most sensitive to the microwave energy, generating highest temperature rises compared to the temperature rises emitted by the same amount of SWCNT or MWCNT-COOH, it is possible that the temperature rises of sample with less MWCNT and more SWCNT and MWCNT-COOH are very close to the temperature rise of one sample with more MWCNT and less SWCNT and MWCNT-COOH. This can lead to heavy overlapping of temperature spectra. To improve the accuracy of prediction, CNTs loaded environmental samples could be exposed to higher microwave energy as the microwave responses of CNTs are more distinctive at higher microwave power.

4. Conclusions

This study is an initial exploration of the potential of the microwave induced heating method to simultaneously quantify individual CNTs in environmental matrices with a mixture of various types of CNTs using PLS, LS-SVM and ANN models. Among all models, despite overlapping spectra in our microwave method, the LS-SVM method may provide better precision in analyzing microwave induced temperature rises data, especially for samples containing less type of CNTs. ANN may generate larger prediction error for samples with more than 2 types of CNTs due to overfitting. Therefore, the application of the microwave measurements with a well-trained LS-SVM model can be used as an efficient, low cost and

rapid approach for quantitatively determination of CNTs in environmental media, which can be further employed in studies of environmental monitoring or toxicity of the nanotubes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded and conducted by the National Risk Management Research Laboratory of U.S. Environmental Protection Agency (EPA), Cincinnati, Ohio. This project was supported, in part, by an appointment in the Research Participation Program at the Office of the Research and Development (ORD). The research results presented in this paper do not necessarily reflect the views of the Agency or its policy. Mention of trade names or products does not constitute endorsement or recommendation for use. The authors would like to thank Dr. Raghuraman Venkatapathy for valuable comments on the manuscript and Mr. Phillip Cluxton for technical and laboratory support.

References

1. De Volder MFL, Tawfick SH, Baughman RH and Hart AJ, Science, 2013, 339, 535–539. [PubMed: 23372006]
2. Meike RJ der Zande van, Frank Walboomers X, and Jansen John A., Tissue Engineering Part B: Reviews, 2010, 17, 57–69. [PubMed: 21062222]
3. Leah AM, Jun G, Randy Vander W, Andrew G, Scott WB and Jacob DM, Toxicological Sciences, 2007, 100, 203–203. [PubMed: 17660506]
4. Cui D, Tian F, Ozkan CS, Wang M and Gao H, Toxicology Letters, 2005, 155, 73–85. [PubMed: 15585362]
5. Jeong SH, Kim KK, Jeong SJ, An KH, Lee SH and Lee YH, Synthetic Metals, 2007, 157, 570–574.
6. Li ZF, Luo GH, Zhou WP, Wei F, Xiang R and Liu YP, Nanotechnology, 2006, 17, 3692–3698.
7. Yang M, Kwon S, Kostov Y, Rasooly A, Rao G and Ghosh U, Environmental Chemistry Letters, 2010, 9, 235–241.
8. Schierz A, Parks AN, Washburn KM, Chandler GT and Ferguson PL, Environmental Science & Technology, 2012, 46, 12262–12271. [PubMed: 22970987]
9. Doudrick K, Corson N, Oberdörster G, Eder AC, Herckes P, Halden RU and Westerhoff P, ACS Nano, 2013, 7, 8849–8856. [PubMed: 23992048]
10. Doudrick K, Herckes P and Westerhoff P, Environmental Science & Technology, 2012, 46, 12246–12253.
11. Plata D. e. L., Reddy CM and Gschwend PM, Environmental Science & Technology, 2012, 46, 12254–12261. [PubMed: 22283840]
12. Sobek A and Bucheli TD, Environmental Pollution, 2009, 157, 1065–1071. [PubMed: 18952329]
13. Abdalla S, Al-Marzouki F, Al-Ghamdi AA and Abdel-Daiem A, Nanoscale Research Letters, 2015, 10, 358. [PubMed: 26377211]
14. Gemperline P, Practical Guide to Chemometrics, CRC Press, Boca Raton, 2006.
15. Brereton RG, Applied chemometrics for scientists, John Wiley & Sons, Hoboken, NJ, Chichester, England, 2007.
16. Qiu P, Ni Y and Kokot S, Analytical Letters, 2006, 39, 1967–1977.
17. Sørensen K, Westley M, Goodacre C, R., & Engelsen S, Anal Bioanal Chem, 2015, 47, 7787–7795.
18. Balabin RM and Lomakina EI, The Analyst, 2011, 136, 1703–1712. [PubMed: 21350755]
19. Kumar N, Bansal A, Sarma GS and Rawal RK, Talanta, 2014, 123, 186–199. [PubMed: 24725882]
20. Mohammadi S and Parastar H, Analyst, 2018, DOI: 10.1039/C7AN02059G.
21. Caramês ETS, Alamar PD, Poppi RJ and Pallone JAL, Food Analytical Methods, 2017, 10, 1609–1615.

22. Pan W, Zhao J, Chen Q and Zhang D, Food Analytical Methods, 2015, 8, 749–757.
23. Borin A, Ferrão MF, Mello C, Maretto DA and Poppi RJ, Analytica Chimica Acta, 2006, 579, 25–32. [PubMed: 17723723]
24. Dinç E and Özdemir A, Il Farmaco, 2005, 60, 591–597. [PubMed: 15922340]
25. Tauler R, Viana M, Querol X, Alastuey A, Flight RM, Wentzell PD and Hopke PK, Atmospheric Environment, 2009, 43, 3989–3997.
26. Salau J. S. i., Tauler R, Bayona J and Tolosa I, Environmental Science & Technology, 1997, 31, 3482–3490.
27. Zacheis GA, Gray KA and Kamat PV, Environmental Science & Technology, 2000, 34, 3401–3407.
28. Ni Y, Xiao W and Kokot S, Journal of Hazardous Materials, 2009, 168, 1239–1245. [PubMed: 19349118]
29. Brunet D, Woignier T, Lesueur-Jannoyer M, Achard R, Rangon L and Barthès BG, Environmental Pollution, 2009, 157, 3120–3125. [PubMed: 19493598]
30. Suykens JAK and Vandewalle J, Neural Processing Letters, 1999, 9, 293–300.
31. Alves JCL and Poppi RJ, Talanta, 2013, 104, 155–161. [PubMed: 23597903]
32. Yu K-Q, Zhao Y-R, Liu F and He Y, 2016, 6, 27574.
33. Bishop CM, Neural Networks for Pattern Recognition, Oxford University Press, Cary, NC, 1996.
34. Hassaninejad-Darzi SK and Torkamanzadeh M, Water Science and Technology, 2016, 74, 2497. [PubMed: 27858806]
35. Cauchi M, Bianco L and Bessant C, Natural Computing, 2011, 10, 77–90.
36. He Y, Al-Abed SR and Dionysiou DD, Science of The Total Environment, 2017, 580, 509–517. [PubMed: 28040213]
37. Sch CBB,olkopf, A. Smola, The MIT Press, Cambridge, Massachusetts, 1999, pp. 25–43.
38. Suykens JAK, Least squares support vector machines, World Scientific, River Edge, NJ, 2002.
39. Xu S, An X, Qiao X, Zhu L and Li L, Pattern Recognition Letters, 2013, 34, 1078–1084.
40. Afkhami A, Abbasi-Tarighat M and Bahram M, Talanta, 2008, 75, 91–98. [PubMed: 18371852]
41. Haykin S, Neural networks and learning machines, Prentice Hall, New York, 2009.

Figure 1.

Plot of RMSECV vs. the number of components for (a) mixture of MWCNT and SWCNT in soil; (b) mixture of MWCNT and SWCNT in sludge; (c) mixture of MWCNT, SWCNT and MWCNT-COOH in soil; (d) mixture of MWCNT, SWCNT and MWCNT-COOH in sludge.

Figure 2.

Plot of PLS-predicted mass of CNTs vs. measured mass of (a) SWCNT and MWCNT in soil; (b) SWCNT and MWCNT in sludge; (c) SWCNT, MWCNT and MWCNT-COOH in soil; (d) SWCNT, MWCNT and MWCNT-COOH in sludge for testing datasets (the solid line with slope of 1).

Figure 3.

Plot of LS-SVM-predicted mass of CNTs vs. measured mass of (a) SWCNT and MWCNT in soil; (b) SWCNT and MWCNT in sludge; (c) SWCNT, MWCNT and MWCNT-COOH in soil; (d) SWCNT, MWCNT and MWCNT-COOH in sludge for testing datasets (the solid line with slope of 1).

Figure 4.

Effect of the number of node (n) on the MSE of resulting ANN models for (a) MWCNT and (b) SWCNT; Effect of decay on the MSE of resulting ANN models for (c) MWCNT and (d) SWCNT.

Figure 5.

Plot of ANN-predicted mass of CNTs vs. measured mass of (a) SWCNT and MWCNT in soil; (b) SWCNT and MWCNT in sludge; (c) SWCNT, MWCNT and MWCNT-COOH in soil; (d) SWCNT, MWCNT and MWCNT-COOH in sludge for testing datasets (the solid line with slope of 1).

Figure 6. Temperature differences in temperature rises of the mixed samples (MWCNT and SWCNT) and summation temperature rises of corresponding CNTs in soil.

Table 1.

Description of original calibration datasets.

Calibration Dataset	Mixture	Environmental Medium
1	SWCNT and MWCNT	Soil
2	SWCNT and MWCNT	Sludge
3	SWCNT, MWCNT and MWCNT-COOH	Soil
4	SWCNT, MWCNT and MWCNT-COOH	Sludge

Table 2.

Variance of predictors and responses explained by the components of the PLS.

Calibration Dataset#	Component	Explained variance of predictor (cumulative %)	Explained variance of response (cumulative %)	
			MWCNT	SWCNT
1	1	99.51	77.71	20.54
	2	99.71	85.45	70.62
	3	99.89	90.42	76.86
	4	99.92	92.79	83.03
	5	99.98	93.56	84.83
	6	99.99	93.73	85.35
	7	100.00	93.73	85.35

Table 3.

Summary of model fitness.

Model	CNTs	Soil			Sludge		
		RMSECV (mg)	R ²	RMSEP (mg)	RMSECV (mg)	R ²	RMSEP (mg)
				Calibration			Calibration
PLS	SWCNT	0.0586	0.70	0.0406	0.0636	0.34	0.0711
	MWCNT	0.0383	0.87	0.0266	0.0837	0.64	0.0442
	SWCNT	0.0754	0.49	0.0530	0.0856	0.35	0.0611
	MWCNT	0.0561	0.72	0.0424	0.0775	0.47	0.0656
	MWCNT-COOH	0.1044	0.23	0.1269	0.1050	0.20	0.0957
LS-SVM	SWCNT	0.0172	0.89	0.0208	0.0256	0.74	0.0317
	MWCNT		0.93	0.0143		0.85	0.0248
	SWCNT	0.0428	0.78	0.0429	0.0376	0.83	0.0342
	MWCNT		0.88	0.0364		0.95	0.0219
	MWCNT-COOH		0.64	0.0572		0.82	0.0409
ANN	SWCNT	0.0376	0.82	0.0341	0.0365	0.77	0.0383
	MWCNT	0.0251	0.89	0.0216	0.0353	0.83	0.0436
	SWCNT	0.0561	0.68	0.0442	0.0558	0.71	0.0399
	MWCNT	0.0461	0.86	0.0349	0.0289	0.80	0.0181
	MWCNT-COOH	0.0613	0.52	0.0421	0.0376	0.69	0.0261

Table 4.

Prediction accuracies of resulted models for testing datasets.

	PLS		LS-SVM		ANN	
	Soil	Sludge	Soil	Sludge	Soil	Sludge
CNTs	RMSEP (mg)	RMSEP (mg)	RMSEP (mg)	RMSEP (mg)	RMSEP (mg)	RMSEP (mg)
SWCNT	0.0250	0.0456	0.0281	0.0328	0.0340	0.0503
MWCNT	0.0209	0.0378	0.0251	0.0295	0.0322	0.0466
SWCNT	0.0785	0.0542	0.0321	0.0243	0.0611	0.0740
MWCNT	0.0694	0.0992	0.0325	0.0313	0.0403	0.0358
MWCNT-COOH	0.1021	0.0665	0.0410	0.0312	0.0823	0.0792