Simultaneous spectrophotometric quantification of dinitrobenzene isomers in water samples using multivariate calibration methods

CrossMark

Tao Lu [a], Yuan Yuan [c], Yan Jiao [a], Zhining Wen [a], Lu Wang [a], Yihuan Zhao [a], Yuxiang Zhang [a], Menglong Li [a], Xuemei Pu [a,*], Tao Xu [b,*]

[a] College of Chemistry, Sichuan University, Chengdu 610064, People's Republic of China
[b] Institute of Chemical Materials, Chinese Academy of Engineering Physics, Mianyang 621900, People's Republic of China
[c] College of Management, Southwest University for Nationalities, Chengdu 610041, People's Republic of China

ARTICLE INFO

ABSTRACT

A facile yet efficient strategy was proposed by means of a combination of ultraviolet (UV) spectrophotometry with multivariate calibration methods, through which 1,2-dinitrobenzene, 1,3-dinitrobenzene and 1,4-dinitrobenzene in water samples could be simultaneously determined without any pre-separation process. The competitive adaptive reweighted sampling combined with successive projections algorithm (CARS–SPA) approach was used to diminish uninformative variables and select the important ones from spectral data measured. The multivariate calibration models were constructed by partial least squares (PLS-1) regression with high accuracy, in which the coefficients of determination of prediction (R^2_{pred}) were 0.9935, 0.9969, and 0.9971 and the root mean square error of prediction (RMSEP) were 0.7850, 0.5411 and 0.5414 for 1,2-dinitrobenzene, 1,3-dinitrobenzene and 1,4-dinitrobenzene, respectively. The optimized model was successfully applied to simultaneously determine the content of the three studied analytes in several real water samples with good recovery close to 100%. Finally, the elliptical joint confidence region (EJCR) tests further confirm that the proposed method has no proportional and constant error in the predicted concentrations, providing a statistic support for the accuracy of the model. These results indicate that it is promise for UV spectroscopy coupled to the multivariate calibration technique to establish a simple, quick, accurate and reliable analysis method for simultaneous determination of some nitroaromatic compounds in real environments. Also, the strategy proposed by the work will advance the analytical methods used in the other complicated sample systems.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Nitroaromatic compounds are not only environment pollutants but also explosively threatening to public security, which have serious negative impact on all living creatures and water used for drinking, household consumption, recreation, fishing, transportation, agriculture and commerce [1]. Among the nitroaromatic compounds, dinitrobenzenes (DNBs) are common contaminants of wastewater and the environment, and most of them have been included in the US Environmental Protection Agency (USEPA) list of priority pollutants [2,3]. The three isomers of DNB, 1,2-dinitrobenzene (1,2-DNB), 1,3-dinitrobenzene (1,3-DNB), and 1,4-dinitrobenzene (1,4-DNB) are the most biorefractory, persistent, and highly toxic dinitrobenzenes, which can induce methemoglobinemia and anemia on prolonged exposure in animals and in humans [4–6] and can also cause testicular toxicity [7,8] and brainstem damage [9]. As they have been used widely in industrial applications as

an intermediate in the chemical synthesis of some rubber chemicals, pesticides, dyes, and explosives or additive explosives [10,11], they may pour into environmental water with concentrations in range of ng L^{-1} to mg L^{-1} [12–14], leading to the contamination. Therefore, it is very important to detect and monitor the three isomers of DNB in aqueous environments.

Some analytical methods were already proposed for determining some DNBs in water, urine and oil, for example, high performance liquid chromatography (HPLC) for 1,3-DNB and nitrobenzene in water [1], organic light emitting diode (OLED) for 1,4-DNB [15], gas chromatography–mass spectrometry (GC–MS) for 1,2-DNB and 1,3-DNB [16], the dispersive liquid–liquid microextraction based on the solidification of floating organic droplet (DLLME-SRO) technique coupled to gas chromatography electron capture detection (GC-ECD) for the three isomers of DNB in different water samples [17], liquid chromatography–mass spectrometry (LC–MS) for measuring 1,2-DNB, 1,3-DNB and 1,4-DNB in water and oil [18,19]. In spite of these instrumental techniques provide good sensitivity and excellent selectivity, most of these techniques are time-consuming and expensive, require bulky

instrumentation and specialized operation [20–22]. Thus, it is highly desired to further develop some facile methods or improve the existing analytical techniques to realize faster and simpler determination on the three isomers of DNB in environments.

As accepted, UV spectrophotometry is a simple, fast, cost-effective and accurate analytical tool [23], which has been employed for quantitative determination of a specific component with a high degree of accuracy. It has been widely used and become one of the most important analytical methods in the modern laboratory [24]. However, the weak selectivity in the spectrophotometric method limits its application in complicated systems due to the strongly overlapped absorption bands from the multiple components.

Multivariate calibration methods in chemometrics can use mathematic separations to take place of chemical separations to solve the problems existed in the spectral data of complicated mixtures, such as collinearity, band overlaps and interactions [25,26]. Among these methods, partial least squares (PLS) and principal component regression (PCR) [27–30] have been frequently used to assist experiments with the purpose of directly determining the chemical components of interest in a complicated mixture without prior separation. The application of these methods to the spectrometric data usually requires the selection of spectral variables for the purpose of constructing well-fitted calibration models [31], because a full-spectrum calibration model probably contains a few uninformative variables and interfering ones, which will have negative impact on the prediction ability of the developed model. Therefore, some methods have been developed for the variable selection, including genetic algorithm (GA) [32,33], iterative PLS (iPLS) [34], uninformative variable elimination (UVE) [35], Monte Carlo based UVE (MC-UVE) [36], variable iterative space shrinkage approach (VISSA) [37], iteratively retains informative variables (IRIV) [38] and margin influence analysis (MIA) [39], competitive adaptive reweighted sampling (CARS) [40] and successive projections algorithm (SPA) [41]. Recently, one combination of CARS and SPA (CARS–SPA) was proposed for the spectral variable selection and exhibited better performance through selecting and optimizing important variables [42].

Based on the considerations above, we, herein, explore a novel method with simple, rapid and accurate advantages to simultaneously determine the 1,2-DNB, 1,3-DNB and 1,4-DNB in environment water samples through a combination of UV–Vis spectrophotometry and PLS coupled with CARS–SPA .

2. Methods

In this work, we suppose that the data matrix **X** of size $k \times j$, includes k samples in rows and j variables in columns. Vector **y** with order $m \times 1$ indicates the measured property of interest. Both **X** and **y** were mean-centered, when modeling.

2.1. Partial least squares regression (PLS)

Partial least squares (PLS) is a powerful multivariate statistical tool originally developed by Word [43,44] and has been successfully and widely applied to the multicomponent quantitative determination of mixtures up to now [45–49]. PLS involves a calibration step, in which the relationship between the spectra and the component concentrations is estimated from a series of standard samples, followed by a prediction step where the component concentrations of the unknown samples are estimated by using the calibration results [50]. The PLS-1 version is optimized for the determination of a single component of interest at a time. During the model training step, the PLS-1 performs the decomposition for the calibration data by an iterative algorithm, which correlates the data with the calibration concentrations employing a so-called 'inverse' model [51]. This offers a series of regression coefficients to be applied to a new sample. The PLS-1 method is well-known, along with a detailed explanation about it is easily available [50,51].

2.2. The competitive adaptive reweighted sampling (CARS)

The CARS, which is proposed recently by Li [40] is a novel variable selection method through selecting the variables with large absolute co-efficients in a multivariate linear regression like partial least squares. Its major idea is to apply the simple but effective principle 'survival of the fittest' on which Darwin's Evolution Theory is based. A detailed explanation of CARS can be found in ref. 40, and the principles of CARS are summarized briefly as follows:

(1) Monte Carlo for model sampling, CARS sequentially selects N subsets of wavelengths from N Monte Carlo sampling runs in an iterative and competitive manner based on the importance level of each variable. In each sampling run, a PLS-1 model is constructed by utilizing the randomly selected samples not all the samples in the calibration set.
(2) Exponentially decreasing function (EDF), EDF is employed to retain the wavelengths which are of relatively large absolute regression coefficients by force.
(3) Adaptive reweighted sampling (ARS), ARS is applied to further exclude wavelengths by means with a competitive way and select the key wavelengths.
(4) Finally, cross validation is used to choose the subset with the lowest root mean square error of cross validation (RMSECV).

2.3. The successive projections algorithm (SPA)

SPA is proposed as a variable selection method by Araújo [41] which exhibits the advantage of acquiring a small representative subset of full-spectrum variables with minimum collinearity. SPA can be divided into three phases: in the first phase, the spectral data of the calibration samples are arranged in a matrix X_{cal} of dimensions ($N_{cal} \times I$) such that the ith variable x_i is associated to the ith column vector $\boldsymbol{x}_i \in \mathfrak{R}^{Ncal}$. Then, these column vectors are subjected to a series of projection operations that generate I chains of M variables, where $M = \min(N_{cal} - 1, I)$ is the maximum number of variables that can be included in a MLR model. A new variable selected in each chain must have the maximum projection value on the orthogonal subspace of the previous ones with respect to all the remaining variables. The second phase consists of assessing the candidate subsets of variables which were extracted from the chains created in the first phase. For every one of those variable subsets, a MLR model is calibrated and the root mean square error of prediction (RMSEP) in the prediction set is also computed. The best subset of variables is selected on the basis of the smallest RMSEP for the prediction samples. The third phase consists of a backward elimination step in order to discard uninformative variables, so that the parsimony of the model could be improved. A detailed explanation of SPA is given elsewhere [41,52–55]. It should be emphasized that the maximum number of selected variables must be less than the number of calibration samples in the process of SPA calculated.

2.4. Statistical parameters

Several statistical parameters were selected to evaluate the performance of the calibration model for the simultaneous determination of 1,2-DNB, 1,3-DNB and 1,4-DNB in mixture samples. These parameters are the root mean square error of cross validation (RMSECV) for calibration set in the cross validation process, root mean square error of prediction (RMSEP) and determination coefficient (R^2_{pred}) for prediction set in the prediction process. These parameters can be calculated through Eqs. (1)–(3), respectively:

$$RMSECV = \left[\frac{1}{m-1} \sum_{i=1}^{m} \left(C_{act,i} - C_{pred,i} \right)^2 \right]^{1/2} \tag{1}$$