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

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*

*Corresponding author: Prof. X. M. Pu

Email address: xmpuscu@scu.edu.cn

Tel.: +86-028-85412290

Fax: +86-028-85412290

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 determination 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_{pred}^2) 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.

Keywords: Ultraviolet spectrophotometry; Partial least square regression; Successive projections algorithm; Competitive adaptive reweighted sampling; Dinitrobenzene isomers; Quantitative analysis

1. Introduction

Nitroaromatics compounds are not only environment pollutants but also explosive 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 droplets (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 method 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 contain 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 $\mathbf{X}$ of size $k \times j$, includes k samples in rows and j variables in columns. Vector $\mathbf{y}$ with order $m \times 1$ indicates the measured property of interest. Both $\mathbf{X}$ and $\mathbf{y}$ 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 coefficients 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 $(N_{\text{cal}} \times I)$ such that the i th variable x_i is associated to the i th column vector $x_i \in \mathbb{R}^{N_{\text{cal}}}$. Then, these column vectors are subjected to a series of projection operations that generate I chains of M variables, where $M = \min(N_{\text{cal}} - 1, I)$ 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_{pred}^2) for prediction set in the prediction process. These parameters can be calculated through Eqs.(1)-(3), respectively:

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

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

$$R^2 = 1 - \frac{\sum_{i=1}^m (C_{\text{act},i} - C_{\text{pred},i})^2}{\sum_{i=1}^m (C_{\text{act},i} - \bar{C})^2} \quad (3)$$

where $C_{\text{pred},i}$ is the predicted concentration of the interested component in i th mixture through the model, $C_{\text{act},i}$ is the real concentration, $\bar{C}$ is mean of real concentration in the calibration or prediction set and m is the number of samples in the calibration or prediction set.

3. Experimental

3.1. Chemical reagents and solutions

1,2-dinitrobenzene (1,2-DNB) was purchased from Fluka (Buchs, Switzerland), 1,3-dinitrobenzene (1,3-DNB) and 1,4-dinitrobenzene (1,4-DNB) were purchased from Aladdin (Shanghai, China). Analytical reagent grade acetonitrile was purchased from Chengdu Kelong Chemical Reagent Co. Ltd (Chengdu, China). Stock standard solutions of 1,2-DNB, 1,3-DNB and 1,4-DNB at a concentration of $500 \mu\text{g mL}^{-1}$ were prepared by exact weighing and dissolving of their solid powder in acetonitrile and then diluting to the mark with acetonitrile in 50 mL volumetric

flasks. Standard working mixtures with different concentrations were prepared by appropriate dilution of the stock standard solutions in acetonitrile.

3.2. Univariate calibration

Univariate calibration experiments were carried out in order to evaluate the linear concentration range for each analyte. Different volumes of standard stock solution of each compound was added to 10 mL volumetric flask and diluted with acetonitrile to the mark. The absorption spectra were recorded over the range of 200-400 nm at their maximum wavelength ($\lambda_{\max}$) 213, 235 and 260 nm for 1,2-DNB, 1,3-DNB and 1,4-DNB, respectively. The linearity range for each compound was determined by constructing the calibration curve using several points as absorbance values at its maximum absorption wavelength versus the sample concentrations. The linear concentration ranges were 0.5-35 $\mu\text{g mL}^{-1}$ for all three analytes, and the regression coefficients (R^2) are higher than 0.999 for all three analytes. The characteristic parameters for the regression equations of single calibration are summarized in Table 1.

3.3. Calibration and prediction sets as well as real samples

In order to optimize a calibration set for the simultaneous analysis of the three analytes in a mixture, a five-level orthogonal array design (OAD, $L_{25}(5^6)$) [56] was employed. A calibration set of 25 ternary-component mixture samples was prepared by mixing different volume of the stock standard solutions in 10 mL volumetric flask and diluting with acetonitrile to final concentrations in the range 2-32 $\mu\text{g mL}^{-1}$ for 1,2-DNB, 2-30 $\mu\text{g mL}^{-1}$ for 1,3-DNB and 2-32 $\mu\text{g mL}^{-1}$ for 1,4-DNB.

With the purpose of checking the predictive performance of the model after optimizing all calibration parameters, a prediction set of 16 ternary-component mixture samples was constructed with concentration within the corresponding calibration ranges, according to a four-level orthogonal

array design (OAD, $L_{16}(4^5)$) [57]. These prediction samples were prepared by adding appropriate volume of stock standard solutions in 10 mL volumetric flask and diluting to the mark with acetonitrile.

To test the analytical performance of the proposed method in a practical situation, three types of water samples were prepared. Tap water was collected from our laboratory, pond water was obtained from lotus pond in Sichuan University (Chengdu, China) and river water was sampled from Jinjiang river (Chengdu, China). Tap water was not treated, pond water was filtered by using 0.45 μm cellulose acetate syringe filter to remove suspended particulate matter, and river water was filtered twice prior to analysis: first through a paper filter to remove suspended sediments and solid materials and then through a 0.45 μm cellulose acetate syringe filter. An appropriate amount of the stock standard solutions of 1,2-DNB, 1,3-DNB and 1,4-DNB were added to 1.0 mL of the prepared water samples and diluting to the mark with acetonitrile in 10 mL volumetric flask to get the desired concentration. These water samples were stored in the dark at 4 $^{\circ}\text{C}$ until the analysis was carried out.

3.4. Apparatus and software

The UV-Vis absorption spectra were collected with respect to a blank of acetonitrile in a 10 mm quartz cell using a U-2900 UV-Vis spectrophotometer (Hitachi Ltd, Japan) with a scan rate of 400 nm min^{-1} and slit width of 1.5 nm. Each spectrum was recorded between 200 nm and 400 nm with a resolution of 1 nm.

All data obtained from the experiments were gathered in a data matrix by using Microsoft Office Excel (version 2010) and transferred to MATLAB software. All computations for variable selection and multivariate calibration were performed in the MATLAB (Version 2013a, MathWorks,

Inc) environment. All the computations were performed on a general-purpose computer with Intel(R) Core(TM) i3-2120 3.30 GHz, 4 GB of RAM and Windows 8 as its operating system.

4. Results and discussion

Fig. 1 presents the chemical structure and the absorption spectra of 1,2-DNB, 1,3-DNB, 1,4-DNB and a mixture of the three isomers. As can be observed from Fig. 1, the absorption spectra of the analytes show a strong spectral overlapping over the entire working range of the UV spectrum, implying that the quantification of any of these compounds will be complicated, even biased if univariate calibration is employed as analytical method in the presence of each other without prior separation. Therefore, it is necessary to introduce multivariate calibration methods (such as PLS-1 and CARS-SPA-PLS-1) to realize the simultaneous and accurate determination on the content of 1,2-DNB, 1,3-DNB and 1,4-DNB in the mixture samples.

4.1. Selection of the optimum number of factors

The crucial step in PLS-1 algorithm is to accurately select the optimum number of factors, which determinates the predictive ability of the calibration model. Therefore, five-fold cross validation was utilized to evaluate the obtained models. The calibration sample set was divided into 5 subsets for the calibration. One subset was left out from 5 subsets as a prediction set in each iteration and the remaining subsets were taken as a calibration set for modeling. The procedure was iteratively repeated five times until all of the subsets pass through the calibration and prediction set. Finally, the known concentrations of the analytes in each reference sample were compared with the prediction concentrations of the analytes in each sample and the root mean square error of cross-validation (RMSECV) **values with respect to the number of factors** was calculated in terms of

Eq.(1). Based on the calculation results, the number of factors selected as the optimum numbers for the calibration models were determined to be 11, 13 and 17 for 1,2-DNB, 1,3-DNB and 1,4-DNB, respectively.

4.2. CARS-SPA for variable selection

As described in previous works [40,42], the number of Monte Carlo sampling runs does not have significant influence on the performance of CARS. Therefore, the number of sampling run was set to 100 as default during the compute process in the work. In order to select the variables of 1,2-DNB, 1,3-DNB and 1,4-DNB, the CARS calculation process was repeatedly performed. Eventually the selected wavelengths of 1,2-DNB, 1,3-DNB and 1,4-DNB were determined to be 36, 13 and 20, respectively. However, the refinement of partial least squares was still required, because too many variables are still retained after CARS was performed [42]. Thus, SPA was applied to refine the variables with minimum redundant information after CARS discarded the unimportant by regression in this work. Ultimately, the key selected wavelengths by SPA are 212, 240, 261, 271 and 337 nm for 1,2-DNB, 207, 229, 240, 269, 284 and 289 nm for 1,3-DNB and 204, 206, 210, 213, 218, 223, 234, 248, 255, 266 and 276 nm for 1,4-DNB. Also, Fig. 2 shows the selected variables for each analyte. As can be seen, the green lines indicate the selected wavelengths by CASR-SPA. For the purpose of illustrating the selected wavelengths are the key variables, the differential spectrum analysis was performed.

The differential spectra (the red curves in Fig. 2a) was obtained by deducing the spectrum of 1,3-DNB from the spectrum of 1,2-DNB. One can obviously observe that the selected wavelengths in the purple box are near the peak labelled as 1 in the differential spectrum between 1,2-DNB and 1,3-DNB, which stand for the biggest difference between the two spectra. The selected wavelengths

within the purple boxes are close to the peaks labelled as 2 and 3 in the differential spectrum between 1,2-DNB and 1,4-DNB in Fig. 2b, which also represent the biggest difference between the two spectra. Similarly, the selected wavelengths for 1,3-DNB and 1,4-DNB can be explained in the same way. Thus, it can be said that the selected wavelengths indeed could be served as the key variables.

Then, the multivariate calibration models (PLS-1 and CARS-SPA-PLS-1) were constructed by both the original spectra data and the selected wavelengths. The obtained results were summarized in Table 2. As can be observed from Table 2, the prediction ability of CARS-SPA-PLS-1 is better than PLS-1. Compared to the calibration models derived from the original spectrum data, the values of RMSECV obtained from the models with the selected wavelengths were reduced by 78.99%, 65.70% and 42.54% for 1,2-DNB, 1,3-DNB and 1,4-DNB, respectively. What's more, the number of both the selected variables and the factors were significantly reduced. The results demonstrated that CARS combined with SPA as a variable selection method is able to effectively extract informative variables from UV spectra.

4.3. Verification of the proposed method by the prediction set

For the purpose of evaluating the predictive ability of the optimized models, an independent prediction set was prepared, including sixteen samples. The established CARS-SPA-PLS-1 and full spectrum PLS-1 models were utilized to determine the content of 1,2-DNB, 1,3-DNB and 1,4-DNB in the prediction set. The results of the both models are listed in Table 3. A comparison of CARS-SPA-PLS-1 results with PLS-1 ones clearly shows that the unimportant variables can be excluded effectively and the predictive precision can be improved significantly by CARS-SPA approach. For instance, the values of REMSP of the three analytes in the CARS-SPA-PLS-1 model

are lower than PLS-1 model while the values of R_{pred}^2 in the CARS-SPA-PLS-1 model are higher than PLS-1 one. Also, the plots of the predicted values versus the actual ones in the prediction set for the three analytes are presented in Fig. 3, which clearly indicates that CARS-SPA can be a good approach for variables selection in spectra data sets.

4.4. Application to real samples

In order to evaluate the applicability and reliability of the proposed method in practice, the recovery study was carried out for the three real water samples (such as tap, river and pond waters) spiked with appropriate known amounts of 1,2-DNB, 1,3-DNB and 1,4-DNB at three different concentration levels. The results are shown in Table 4. Compared the two approaches in terms of the predicted values and the recovery values, it is can be found that the results derived from CARS-SPA-PLS-1 are still better than the PLS-1 ones, in which the average values of recoveries are 99.04% for 1,2-DNB, 100.25% for 1,3-DNB and 100.39% for 1,4-DNB. The results further demonstrate that the proposed method will be highly promising in practice.

4.5. Comparison between PLS-1 and CARS-SPA-PLS-1 models

To gain insight into the accuracy of the two mentioned methods in this study, a linear regression analysis of actual concentration values versus predicted ones by PLS-1 and CARS-SPA-PLS-1 for 1,2-DNB, 1,3-DNB and 1,4-DNB in the prediction set 16 samples and 9 real samples was implemented. The estimated slope and intercept were compared with their theoretical values of 1 and 0, respectively, based on the elliptical joint confidence region (EJCR) test [58]. Fig. 4 shows the EJCR plots for the two applied multivariate calibration methods for 1,2-DNB, 1,3-DNB and 1,4-DNB, respectively. As can be seen, all ellipses contain the ideal values (1) for the slope and (0) for the intercept at a confidence level of 95% for the two models, demonstrating the accuracy of

the proposed method. Besides, the elliptic size obtained with CARS-SPA-PLS-1 associated with 1,2-DNB, 1,3-DNB and 1,4-DNB is smaller, indicating that this method shows higher precision for the determination of the three studied compounds. In a whole, these results demonstrate that the proposed method have no proportional and constant errors in the predicted concentrations, and further confirming the excellent predictive ability of CARS-SPA-PLS-1.

5. Conclusions

In the work, we combined UV spectrometry and one multivariable calibration method (PLS-1) to successfully realize simultaneous determination on the content of 1,2-DNB, 1,3-DNB and 1,4-DNB in water samples without prior separation. Compared with PLS-1 model, the better results were obtained when using the variable selection method of CARS-SPA, in which RMSEP values are lower than 0.7850 and R_{pred}^2 values higher than 0.99 for all the three analytes in the prediction set, exhibiting high predictive ability and robustness. Moreover, good recovery values close to 100% were obtained by the application of CARS-SPA-PLS-1 model for the real water samples, further displaying high promise in practice. In addition, the statistical analysis results also confirm the accuracy and reliability of the PLS-1 method coupled with the variable selection by CARS-SPA. In a whole, our work demonstrates that the UV spectrophotometry coupled with chemometrics could be efficient method for simple, rapid, cost-effective and reliable determination for the three isomers of DNB in real environments. More importantly, the strategy proposed by us will advance the analytical methods for simultaneous determination on multiple components of the other complicated sample systems.

Acknowledgements

The authors gratefully acknowledge the National Science Foundation of China (Grant No. U1230121, 21273154 and 21573151) and Sichuan Province Science and Technology Support Program (Grant No. 2015GZ0193).

References

- 1 H. Parham, S. Saeed, Pre-concentration and determination of traces of nitrobenzene and 1,3-dinitrobenzene in water samples using anthracite adsorbent, *J. Ind. Eng. Chem.* 20 (2014) 1003-1009.
- 2 M.J. Lang, S.E. Burns, Improvement of EPA method 8330: complete separation using a two-phase approach, *J. Chromatogr. A*, 849 (1999) 381-388.
- 3 A. Preiss, A. Bauer, H.M. Berstermann, S. Gerling, R. Haas, A. Joos, A. Lehmann, L. Schmalz, K. Steinbach, Advanced high-performance liquid chromatography method for highly polar nitroaromatic compounds in ground water samples from ammunition waste sites, *J. Chromatogr. A*, 1216 (2009) 4968-4975.
- 4 T. Watanabe, N. Ishihara, M. Ikeda, Toxicity of and biological monitoring for 1,3-diamino-2,4,6-trinitrobenzene and other nitroamino derivatives of benzene and chlorobenzene, *Int. Arch. Occup. Environ. Health*. 37 (1976) 157-168.
- 5 M.F. Obasaju, D.F. Katz, M.G. Miller, Species differences in susceptibility to 1,3-dinitrobenzene-induced testicular toxicity and methemoglobinemia, *Fundam. Appl. Toxicol.* 16 (1991) 257-266.
- 6 P.A. Cossum, D.E. Rickert, Metabolism and toxicity of dinitrobenzene isomers in erythrocytes from Fischer-344 rats, rhesus monkeys and humans, *Toxicol. Lett.* 37 (1987) 157-163.

- 7 R.E. Linder, R.A. Hess, L.F. Strader, Testicular toxicity and infertility in male rats treated with 1,3-dinitrobenzene, *J. Toxicol. Environ. Health*. 19 (1986) 477-489.
- 8 T.E. Cody, S. Witherup, L. Hastings, K. Stemmer, R.T. Christian, 1,3-dinitrobenzene: toxic effects in vivo and in vitro, *J. Toxicol. Environ. Health*. 7 (1981) 829-847.
- 9 I. Romero, A.W. Brown, J.B. Cavanagh, C.C. Nolan, D.E. Ray, M.P. Seville, Vascular factors in the neurotoxic damage caused by 1,3-dinitrobenzene in the rat, *Neuropathol. Appl. Neurobiol.* 17 (1991) 495-508 .
- 10 T.B. Brill, K.J. James, Kinetics and mechanisms of thermal decomposition of nitroaromatic explosives, *Chem. Rev.* 93 (1993) 2667-2692.
- 11 N. Arshad, N.K. Janjua, S. Ahmed, A.Y. Khan, L.H. Skibsted, Electrochemical investigations of antioxidant interactions with radical anion and dianion of 1,3-dinitrobenzene, *Electrochim. Acta*, 54 (2009) 6184-6189.
- 12 F. Monteil-Rivera, C. Beaulieu, J. Hawari, Use of solid-phase microextraction/gas chromatography–electron capture detection for the determination of energetic chemicals in marine samples, *J. Chromatogr. A*, 1066 (2005) 177-187.
- 13 J. Yinon, S. Zitrin, *Modern Methods and Applications in Analysis of Explosives*, John Wiley & Sons, New York, 1993.
- 14 J. Yinon, *Forensic and Environmental Detection of Explosives*, John Wiley & Sons, New York, 1999.
- 15 E.C. Nallon, M.P. Polcha, V.P. Schnee, Electrically excited polymers for the detection of dinitrobenzene, *Sensor. Actuat. B: Chem.* 190 (2014) 578-584.
- 16 M. Bader, T. Goen, J. Muller, J. Angerer, Analysis of nitroaromatic compounds in urine by gas

- chromatography–mass spectrometry for the biological monitoring of explosives, *J. Chromatogr. B*, 710 (1998) 91-99.
- 17 Y.L. Wu, L.P. Dai, J. Cheng, F. Guo, J.K. Li, Application of DLLME based on the solidification of floating organic droplets for the determination of dinitrobenzenes in aqueous samples, *Chromatographia*, 72 (2010) 695-699.
- 18 C. Crescenzi, J. Albinana, H. Carlsson, E. Holmgren, R. Batlle, On-line strategies for determining trace levels of nitroaromatic explosives and related compounds in water, *J. Chromatogr. A*, 1153 (2007) 186-193.
- 19 R. Tachon, V. Pichon, M.B.L. Borgne, J.J. Minet, Use of porous graphitic carbon for the analysis of nitrate ester, nitramine and nitroaromatic explosives and by-products by liquid chromatography–atmospheric pressure chemical ionisation-mass spectrometry, *J. Chromatogr. A*, 1154 (2007) 174-181.
- 20 L. Senesac and T. G. Thundat, Nanosensors for trace explosive detection, *Mater. Today*, 11 (2008) 28-36.
- 21 K.L. Diehl, E.V. Anslyn, Array sensing using optical methods for detection of chemical and biological hazards, *Chem. Soc. Rev.* 42 (2013) 8596-8611.
- 22 J. Cho, R. Anandakathir, A. Kumar, J. Kumar, P.U. Kurup, Sensitive and fast recognition of explosives using fluorescent polymer sensors and pattern recognition analysis, *Sens. Actuators B*, 160 (2011) 1237-1243.
- 23 H.H. Perkampus, H.C. Grinter, T. Threlfall, *UV-VIS spectroscopy and its applications*, Springer-Verlag, New York, 1992.
- 24 D. Xu, W. Fan, H.Y. Lv, Y.Z. Liang, Y. Shan, G.Y. Li, Z.Y. Yang, L. Yu, Simultaneous

- determination of traces amounts of cadmium, zinc, and cobalt based on UV-Vis spectrometry combined with wavelength selection and partial least squares regression, *Spectrochim. Acta, Part A*, 123 (2014) 430-435.
- 25 R. Ergon, Reduced PCR/PLSR models by subspace projections, *Chemom. Intell. Lab. Syst.* 81 (2006) 68-73.
- 26 B Hemmateenejad, R. Ghavami, R. Miri, M. Shamsipur, Net analyte signal-based simultaneous determination of antazoline and naphazoline using wavelength region selection by experimental design-neural networks, *Talanta*, 68 (2006) 1222-1229.
- 27 S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.* 2 (1987) 37-52.
- 28 M.J. Arrabal, P.V. González, C.C. Gámez, A.S. Misiego, Spectrophotometric determination of mixtures of iron(III) and manganese(II) by complexation with 3-indolylacetohydroxamic acid and principal component regression multivariate calibration, *Analyst*, 119 (1994) 1537-1540.
- 29 T. Lu, Y. Yuan, X. He, M.L. Li, X.M. Pu, T. Xu, Z.N. Wen, Simultaneous determination of multiple components in explosives using ultraviolet spectrophotometry and a partial least squares method, *RSC Adv.* 5 (2015) 13021-13027.
- 30 M. de Luca, G. Ioele, A. Risoli, G. Ragno, Improvement of multivariate calibration techniques applied to 1-to-*N* component mixtures through an optimized experimental design, *Microchem. J.* 83 (2006) 24-34.
- 31 C.H. Spiegelman, M.J. Mcshane, M. Motamedi, Q.L. Yue, G.L. Coté, Theoretical justification of wavelength selection in PLS calibration: development of a new algorithm, *Anal. Chem.* 70 (1998) 35-44.

- 32 Q. Ding, G.W. Small, M.A. Arnold, Genetic algorithm-based wavelength selection for the near-infrared determination of glucose in biological matrixes: initialization strategies and effects of spectral resolution, *Anal. Chem.* 70 (1998) 4472–4479 .
- 33 R. Leardi, Application of genetic algorithm–PLS for feature selection in spectral data sets, *J. Chemom.* 14 (2000) 643–655.
- 34 S.D. Osborne, R. Künnemeyer, R.B. Jordan, Method of wavelength selection for partial least squares, *Analyst*, 122 (1997) 1531-1537.
- 35 V. Centner, D.L. Massart, O.E. de Noord, S. de Jong, B.M. Vandeginste, C. Sterna, Elimination of uninformative variables for multivariate calibration, *Anal. Chem.* 68 (1996) 3851-3858.
- 36 W.S. Cai, Y.K. Li, X.G. Shao, A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra, *Chemom. Intell. Lab. Syst.* 90 (2008) 188-194.
- 37 B.C. Deng, Y.H. Yun, Y.Z. Liang, L.Z. Yi, A novel variable selection approach that iteratively optimizes variable space using weighted binary matrix sampling, *Analyst*, 139 (2014) 4836-4845.
- 38 Y.H. Yun, W.T. Wang, M.L. Tan, Y.Z. Liang, H.D. Li, D.S. Cao, H.M. Lu, Q.S. Xu, A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration, *Anal. Chim. Acta*, 807 (2014) 36-43.
- 39 H.D. Li, Y.Z. Liang, Q.S. Xu, D.S. Cao, Recipe for uncovering predictive genes using support vector machines based on model population analysis, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 8 (2011) 1633-1641.
- 40 H. Li, Y.Z. Liang, Q.S. Xu, D.S. Cao, Key wavelength screening using competitive adaptive reweighed sampling method for multivariate calibration, *Anal. Chim. Acta*, 648 (2009) 77-84.

- 41 M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame, V. Visani, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis, *Chemom. Intell. Lab. Syst.* 57 (2001) 65-73.
- 42 G. Tang, Y. Huang, K.D. Tian, X.Z. Song, H. Yan, J. Hu, Y.M. Xiong, S.G. Min, A new spectral variable selection pattern using competitive adaptive reweighted sampling combined with successive projections algorithm, *Analyst*, 139 (2014) 4894-4902.
- 43 H. Wold, in F. David (Ed.), *Research Papers in Statistics*, Wiley, New York, 1966.
- 44 S. Wold, in H. Martens, H. Runswurm (Eds.), *Food Research and Data Analysis*, Applied Science Publishers, London, 1983.
- 45 W. Henao-Escobar, O. Domínguez-Renedo, M.A. Alonso-Lomillo, M.J. Arcos-Martínez, Resolution of quaternary mixtures of cadaverine, histamine, putrescine and tyramine by the square wave voltammetry and partial least squares method, *Talanta*, 143 (2015) 97-100.
- 46 Q.Y. Luo, Y.H. Yun, W. Fan, J.H. Huang, L.X. Zhang, B.C. Deng, H.M. Lu, Application of near infrared spectroscopy for the rapid determination of epimedin A, B, C and icariin in Epimedium, *RSC Adv.* 5 (2015) 5046-5052.
- 47 J. El Haddad, F. de Miollis, J. Bou Sleiman, L. Canioni, P. Mounaix, B. Bousquet, Chemometrics applied to quantitative analysis of ternary mixtures by terahertz spectroscopy, *Anal. Chem.* 86 (2014) 4927-4933.
- 48 D. Chen, W.S. Cai, X.G. Shao, Removing uncertain variables based on ensemble partial least squares, *Anal. Chim. Acta*, 598 (2007) 19-26.
- 49 S. Mahramyari, E. Pourbasheer, A. Banaei, M. R. Ganjali, P. Norouzi, Simultaneous spectrophotometric determination of ceftazidime and sulbactam using multivariate calibration

methods, RSC Adv. 4 (2014) 41039-41044.

50 H. Martens, T. Naes, Multivariate Calibration. Wiley, Chichester, 1989.

51 D.M. Haaland, E.V. Thomas, Partial least-squares methods for spectral analyses. 1. relation to other quantitative calibration methods and the extraction of qualitative information, Anal. Chem. 60 (1988) 1193-1202.

52 S.F. C. Soares, A.A. Comes, A.R. G. Filho, M.C.U. Araújo, R.K.H. Galvão, The successive projection algorithm, TrAC, Trends Anal. Chem. 42 (2013) 84-98.

53 S.F.C. Soares, R.K.H. Galvão, M.C.U. Araújo, E.C. Silva, C.F. Pereira, S.I.E. Andrade, F.C. Leite, A modification of the successive projection algorithm for spectral variable selection in the presence of unknown interferents, Anal. Chim. Acta, 689 (2011) 22-28.

54 H.M. Paiva, S.F.C. Soares, R.K.H. Galvão, M.C.U. Araújo, A graphical user interface for variable selection employing the successive projection algorithm, Chemom. Intell. Lab. Syst. 118 (2012) 260-266.

55 A.A. Gomes, R.K.H. Galvão, M.C.U. Araújo, G. Vêras, E.C. Silva, The successive projection algorithm for interval selection in PLS, Microchem. J. 110 (2013) 202-208.

56 R.G. Brereton, Introduction to multivariate calibration in analytical chemistry, Analyst, 125 (2000) 2125-2154.

57 W.G. Lan, M.K. Wong, N. Chen, Y.M. Sin, Orthogonal array design as a chemometric method for the optimization of analytical procedures. Part 2. Four-level design and its application in microwave dissolution of biological samples, Analyst, 119 (1994) 1669-1675.

58 A.V. González, M.A. Herrador, A.G. Asuero, Intra-laboratory testing of method accuracy from recovery assays, Talanta, 48 (1999) 729-736.

Figure Captions:

Fig. 1 Molecular structures and absorption spectra of 1,2-DNB ($12.50 \mu\text{g mL}^{-1}$), 1,3-DNB ($12.50 \mu\text{g mL}^{-1}$), 1,4-DNB ($12.50 \mu\text{g mL}^{-1}$) and a mixture containing 1,2-DNB ($12.50 \mu\text{g mL}^{-1}$), 1,3-DNB ($12.50 \mu\text{g mL}^{-1}$) and 1,4-DNB ($12.50 \mu\text{g mL}^{-1}$).

Fig. 2 Comparison of the selected wavelengths by CARS-SPA for 1,2-DNB, 1,3-DNB and 1,4-DNB. (a) and (b) Selected wavelengths of 1,2-DNB; (c) and (d) Selected wavelengths of 1,3-DNB; (e) and (f) Selected wavelengths of 1,4-DNB. All the black lines denote the spectra of mixture solution. The red lines denote the differential spectra of 1,2-DNB-1,3-DNB (a), 1,2-DNB-1,4-DNB (b), 1,3-DNB-1,2-DNB (c), 1,3-DNB-1,4-DNB (d), 1,4-DNB-1,2-DNB (e) and 1,4-DNB-1,3-DNB (f), respectively. The green lines denote the selected wavelengths by CARS-SPA.

Fig. 3 Predicted concentrations of 1,2-DNB (a), 1,3-DNB (b) and 1,4-DNB (c) versus the actual values for the prediction set by PLS-1 (red, left vertical axis) and CARS-SPA-PLS-1 (blue, right vertical axis).

Fig. 4 Elliptical joint confidence region (EJCR) plot at 95% confidence limit for the slope and intercept corresponding to regressions of actual concentrations *versus* PLS-1 and CARS-SPA-PLS-1 predicted concentrations of 1,2-DNB (a), 1,3-DNB (b) and 1,4-DNB (c), respectively. PLS-1 (red solid line) and CARS-SPA-PLS-1 (blue dashed line) and. The pentagram marks the ideal point (slope=1, intercept=0).

Fig.1

Fig.2

Fig.3

Fig.4

Tables:**Table 1.** Analytical characteristics for single component spectroscopic determination.

Analyte	$\lambda_{\max}$ (nm)	Linear range ($\mu\text{g mL}^{-1}$)	R^2	LOD ^a ($\mu\text{g mL}^{-1}$)	LOQ ^b ($\mu\text{g mL}^{-1}$)
1,2-DNB	213	0.50-35.00	0.9996	0.19	0.65
1,3-DNB	235	0.50-35.00	0.9992	0.09	0.30
1,4-DNB	260	0.50-35.00	0.9995	0.08	0.28

^a The LOD (limit of detection) was determined by a signal-to-noise ratio (S/N) = 3 for each analyte.

^b The LOQ (Limit of quantitation) was determined by a signal-to-noise ratio (S/N) = 10 for each analyte.

Table 2 Statistical parameters obtained for PLS-1 or CARS-SPA-PLS-1 in the calibration set.

Models	Figures of merit	Analytes		
		1,2-DNB	1,3-DNB	1,4-DNB
PLS-1	Concentration range ($\mu\text{g mL}^{-1}$)	2-32	2-30	2-32
	$n\text{VAR}^a$	201	201	201
	$n\text{LVs}^b$	11	13	17
	RMSECV c ($\mu\text{g mL}^{-1}$)	1.6148	1.3343	0.9370
	Concentration range ($\mu\text{g mL}^{-1}$)	2-32	2-30	2-32
CARS-SPA-PLS-1	$n\text{VAR}$	5	6	11
	$n\text{LVs}$	4	6	11
	RMSECV ($\mu\text{g mL}^{-1}$)	0.3392	0.4576	0.5384

^a $n\text{VAR}$ denotes the number of variables selected.

^b $n\text{LVs}$ denotes the number of latent variables.

^c RMSECV is the root mean square error of the cross validation.

Table 3 Statistical parameters of the optimized models using the PLS-1 and CARS-SPA-PLS-1 in the prediction set.

Parameters	1,2-DNB		1,3-DNB		1,4-DNB	
	PLS-1	CARS-SPA-PLS-1	PLS-1	CARS-SPA-PLS-1	PLS-1	CARS-SPA-PLS-1
RMSEP ^a ($\mu\text{g mL}^{-1}$)	1.4497	0.7850	1.6442	0.5411	0.7778	0.5414
R_{pred}^2	0.9778	0.9935	0.9714	0.9969	0.9940	0.9971

^a The root mean square error of prediction.

Table 4 Simultaneous determination of 1,2-DNB, 1,3-DNB and 1,4-DNB in spiked water samples by PLS-1 and CARS-SPA-PLS-1 methods.

Sample	Actual value (μg mL ⁻¹)			Predicted value ^a (μg mL ⁻¹)					
	1,2-DNB	1,3-DNB	1,4-DNB	PLS-1			CARS-SPA-PLS-1		
				1,2-DNB	1,3-DNB	1,4-DNB	1,2-DNB	1,3-DNB	1,4-DNB
Tap water									
(I)	10	20	14.4	10.23(102.30)	18.27(91.35) ^b	15.50(107.64)	9.62 (96.20)	19.89 (99.45)	13.78 (95.69)
(II)	20	10.4	10.4	20.32(101.15)	8.37 (80.48)	11.33(108.94)	20.08(100.40)	10.37 (99.71)	9.69 (93.17)
(III)	15	14.4	20	16.33(108.87)	11.98 (83.19)	19.29 (96.45)	15.03(100.20)	14.65(101.74)	20.16(100.80)
River water									
(I)	10	20	14.4	10.22(102.20)	17.71 (88.55)	13.87 (96.32)	9.31 (93.10)	20.09 (99.31)	14.30 (99.31)
(II)	20	10.4	10.4	18.99 (94.95)	9.45 (90.87)	9.33 (89.71)	19.27 (96.35)	10.57(101.63)	11.46(110.19)
(III)	15	14.4	20	16.85(112.33)	11.72 (81.39)	22.65(113.25)	15.20(101.33)	14.51(100.76)	19.59 (97.95)
Pond water									
(I)	10	20	14.4	9.15 (91.50)	18.57 (92.58)	15.82(109.86)	9.78 (97.80)	19.77 (98.85)	14.96 (103.89)
(II)	20	10.4	10.4	20.77(103.85)	8.26 (79.42)	11.03(106.06)	20.67(103.35)	10.09 (97.02)	10.27 (98.75)
(III)	15	14.4	20	16.55(110.33)	11.68 (81.11)	21.07(105.35)	15.39(102.60)	14.27(103.75)	20.75 (103.75)

^a The average value of the three independent determinations.

^b The values in parentheses are the recoveries for each component. Recovery (%) = $100 \times (C_{\text{pred}}/C_{\text{act}})$, C_{pred} represents the prediction concentration, C_{act} represents the actual concentration.