Regular Article

ANALYTICAL
AND BIOANALYTICAL
CHEMISTRY
RESEARCH

Published by the
Iranian Chemical Society

***Anal. Bioanal. Chem. Res.**, Vol. 11, No. 3, 327-334, July 2024.*

Simultaneous Spectrophotometric Quantification of 2-Nitrophenol and 4-Nitrophenol in Binay Mixtures Based on Partial Least Squares Method: Comparison Analysis of Five Types of Data Sets

S. Maryam Sajjadi*

Department of Chemical Engineering, Lamerd Higher Education Center, Shiraz University of Technology, Fars, Lamerd, Iran

(Received 9 January 2024, Accepted 8 April 2024)

In this study, simultaneous quantification of 2-nitrophenol and 4-nitrophenol in their binary mixtures was investigated spectrophotometrically. Since the signals of analytes were highly overlapped, a multivariate partial least squares (PLS) technique was proposed to analyze the data. The PLS method makes the analysis possible without the need for the separation of analytes by tedious separation procedures or using expensive instrumentation techniques such as chromatographic methods. Both 2-nitrophenol and 4-nitrophenol possess acid-base properties and it was required to investigate the effect of pH on the FOM of the calibration. Therefore, at three pH conditions, the calibration processes were evaluated and the results showed the best FOM and the least root mean squares error of prediction (RMSEP) for both analytes were achieved for the augmented data at 3.45 and 8.95 of pHs where only neutral or anionic forms of analytes were present in the solution. The analytical sensitivity, limit of detection, $R^{2,}$ and RMSEP were 108.3 ppm^{-1}, 0.08ppm, 1.00, 0.04 ppm and; 163.2 ppm^{-1}, 0.06 ppm, 0.9999, 0.04 ppm for 2-nitrophenol and 4-nitrophenol, respectively.

Keywords: Nitrophenols, Acid-base properties, Figures of merit, Partial Least Squares

INTRODUCTION

Phenolic compounds are known as organic pollutants in drinking water and aquatic environments because of their abundant applications in various industries such as essential raw materials or intermediate components in the production of dyes, pharmaceuticals, and pesticides. Among different kinds of phenolic compounds, 2-nitrophenol (2-NP) and 4-nitrophenol (4-NP) are widely employed in industries; and they have been ranked as 126th organic pollutants based on the reports of the Clean Water Act (CWA) [1]. The maximum permissible range of concentration of these compounds is 1-20 mg l^{-1} in soluble, and stable, a fact which makes them water [2]. These pollutants are poorly biodegradable, highly persistent in the soil, air, and groundwater. Therefore, these compounds accumulate in organisms and cause long-term damage to the environment as well as harmful effects on human health [3]. Considering the aforementioned, the quantification of 2-NP and 4-NP has attracted high attention in environmental issues and the protection of human health. There are a number of analytical methods and procedures developed for the quantification of these analytes, which include fluorescence [4], high-performance liquid chromatography [5], capillary electrophoresis [6], and flow injection analysis [7], electrochemical techniques [8]. However, most of these techniques are time-consuming, expensive, and poorly reproducible. Moreover, they require sophisticated operation and sample preparation techniques.

One of the most common strategies to determine nitrophenols is the spectrophotometric technique. Besides lots of benefits of this approach such as simplicity, availability, and wide application scope, some analytes cannot be measured due to having signal overlap problems [9]. Simultaneous spectrophotometric detection or

*Corresponding author. E-mail: sajjadi@sutech.ac.ir

quantification of nitrophenol isomers is difficult due to their similar structure and properties. Indeed, the absorption spectra of these analytes are severely overlapped, and overcoming this limitation is a great challenge in environmental monitoring. Multivariate calibration is defined as a developed mathematical model that relates unselective multiple instrumental signals with analyte concentrations and it can be employed for the analysis of data suffering signal overlapping problems. Some common multivariate calibration methods include Classical Least squares [10], Inverse Least squares (ILS) [11], principal component regression (PCR) [12], and partial least-squares regression (PLS) techniques [13,14]. Among them, PLS is the most sophisticated algorithm because it maximizes the linear combination of dependencies of signals with concentration of analyte and it uses the scores both relating to the concentration of analyte and carrying the most variation in the data.

To the best of my knowledge, there is no thorough investigation on the simultaneous determination of 2-NP and 4-NP based on cost-effectiveness and simple spectrophotometric method by taking advantage of their acid-base property to flourish the figures of merit of (FOM) of analytes' calibration. The main goal of the present study is the simultaneous quantification of 2-NP and 4-NP at different pH conditions to achieve the best FOM of analytes' calibration and prediction ability for the estimation of the analyte concentrations in unknown samples. It will be shown that augmentation of data at the pHs that consist of only acidic and basic forms of analytes is the best-collected data set for the purpose of analytes quantification.

EXPERIMENTAL

Reagents and Solutions

All the applied reagents were of analytical grade (Merck or Aldrich) and no further purification was conducted on them. 2-NP and 4-NP were obtained from Sigma-Aldrich (St Louis, MO, USA). Double-distilled water was used throughout the experiments. Stock standard solutions of 2-NP and 4-NP were prepared in double-distilled water with 300 ppm and 400 ppm of concentration, respectively. The solutions were stored for less than 1 month and nitrophenol solutions were obtained by diluting these stock solutions. The Britton-Robinson (B-R) universal buffer with pH values 3.45, 7.12, and 8.95 was prepared in distilled water and applied to adjust the pH of solutions.

Instrumentation and Software

The pH of the aqueous solution was measured by a digital pH meter (Metrohm 713, Germany) equipped with a glass combination electrode. The UV-Vis spectra of samples were recorded in the wavelength range of 250-520 nm with 2 nm of increment by a two-beam spectrophotometer (Cary 100 Bio, Varian, Australia) with 1cm quartz cell. The spectral data were exported in ASCII format from the instrument software to MATLAB 2013a version 8.1.0.604. The data was then analyzed by a first-order multivariate calibration user-friendly toolbox developed by Olivieri *et al.* which is available on the homepage: (http://www.iquir-conicet.gov.ar/descargas/mvc1.zip) [15].

PARTIAL LEAST SQUARES METHOD

The PLS technique is one of the famous first-order calibration methods relating variations in a signal **x** with variations of property of interest. In this study, **x** is the UV-Vis spectrum of the sample and the property is the concentration of 2-NP or 4-NP. This method is a two-step procedure; the first step is calibration, in which the relationship between spectra and analyte concentrations is created from a set of calibration samples, and the second step is prediction, where the calibration results are employed to calculate the component concentrations in unknown or prediction samples.

In the PLS method, the spectral information about the analyte concentration is ''extracted'' by the following equation:

$$\mathbf{c} = \mathbf{x}^{\mathrm{T}}\mathbf{b} \qquad (1)$$

The vector of regression coefficients **b** (N×1) is achieved by the spectra of M calibration samples measured at N wavelengths, **R**(M×N), and c is the reference concentrations to these calibration samples. The reference number 13 provides a more detailed explanation of the algorithm.

The most crucial task in executing the PLSR model is selecting the significant number of components; in this

regard, the leave-one-out-cross validation method (LOOCV) can be applied to the calibration data set [16]. In the LOOCV strategy, each sample is left out from the calibration set whose concentration is predicted using a model built with the data of the remaining samples under a different number of components. The predicted error sum of squares (PRESS) parameters can be calculated based on the sum of square errors for the prediction of the left-out samples as follows [17]:

$$PRESS = \sum_{m=1}^{M}(c_m - c_{m,pred})^2 \quad (2)$$

Where M is the total number of calibration samples, c_m and $c_{m,pred}$ are the real and predicted concentrations of analyte in the mth sample, respectively. The data set is analyzed based on a different number of latent variables from 1 to a number larger than the expected optimal components. Finally, the number of latent variables with minimum PRESS is selected as an optimal significant component used for further analysis.

In the calibration step, the analysis is evaluated based on computing some statistical parameters including R^2; the root of the mean squared error of calibration samples (RMSEC), sensitivity (SEN), analytical sensitivity, limit of detection (LOD), and limit of quantitation (LOQ). These parameters are calculated according to the following equations:

$$R^2 = 1 - \left(\frac{\sum_{m=1}^{M}(c_m - \widehat{c_m})^2}{\sum_{i=m}^{M}(c_m - \overline{c_m})^2}\right) \quad (3)$$

$$RMSEC = \sqrt{\frac{\sum_{m=1}^{M}(c_m - \widehat{c_m})^2}{M-1}} \quad (4)$$

Where M, $\widehat{c_m}$ and $\overline{c_m}$ are the total number of samples in the calibration set, the real and predicted concentrations of the given analyte in the mth calibration sample.

$$SEN = \frac{1}{\|\boldsymbol{b}\|} \quad (5)$$

Where **b** is the same as in Eq. (1) and || || denotes the norm of the vector.

SEN unit is (signal × concentration^{-1}), *i.e.*, this value depends on the signal type of the calibration model and it is not applicable for comparing the sensitivities derived from two determination techniques such as spectrophotometric and spectrofluorimetric determinations of analytes. Consequently, analytical sensitivity γ is defined as the most beneficial indicator, which is computed based on dividing sensitivity to instrumental noise:

$$\gamma = \frac{SEN}{\sigma_r} \quad (6)$$

Where σ_r is the constant uncertainty in instrumental noise as reported in [18].

LOD and LOQ are estimated as the concentration level, which are 3.3, and 10 times of root square of prediction error, respectively; and the range of LOD and LOQ depends on minimum and maximum leverage for a blank sample respectively [17].

In the prediction step, the root of the mean squared error of prediction samples (RMSEP) is calculated to evaluate the prediction ability of the PLS model for the analysis of unknown samples.

$$RMSEP = \sqrt{\frac{\sum_{f=1}^{F}(c_f - \widehat{c_f})^2}{F-1}} \quad (7)$$

Where F, c_f and $\widehat{c_f}$ are the total number of samples in the prediction data set, the real and predicted concentrations of the analyte in the fth prediction sample.

RESULTS AND DISCUSSION

Quantification of 2-NP and 4-NP in Binary Mixtures Using PLS Method

2-NP and 4-NP in aqueous solutions are present as neutral and anionic forms in the pH range of 0-14 (Fig. 1); the reported pKas of 2-NP and 4-NP are 7.41 and 7.19, respectively [9].

Since the main goal of this study is finding the optimal pH for simultaneous quantification of 2-NP and 4-NP in their mixture, the calibration data were collected at three pH conditions including 3.45, 712, and 8.95. Their pHs were chosen based on the pK_a values of the analytes as the solutions at pH of 3.45 and 8.95 are mainly composed of neutral and anionic species of the analytes, respectively; and at pH of 7.12, all of the forms of analytes are present which

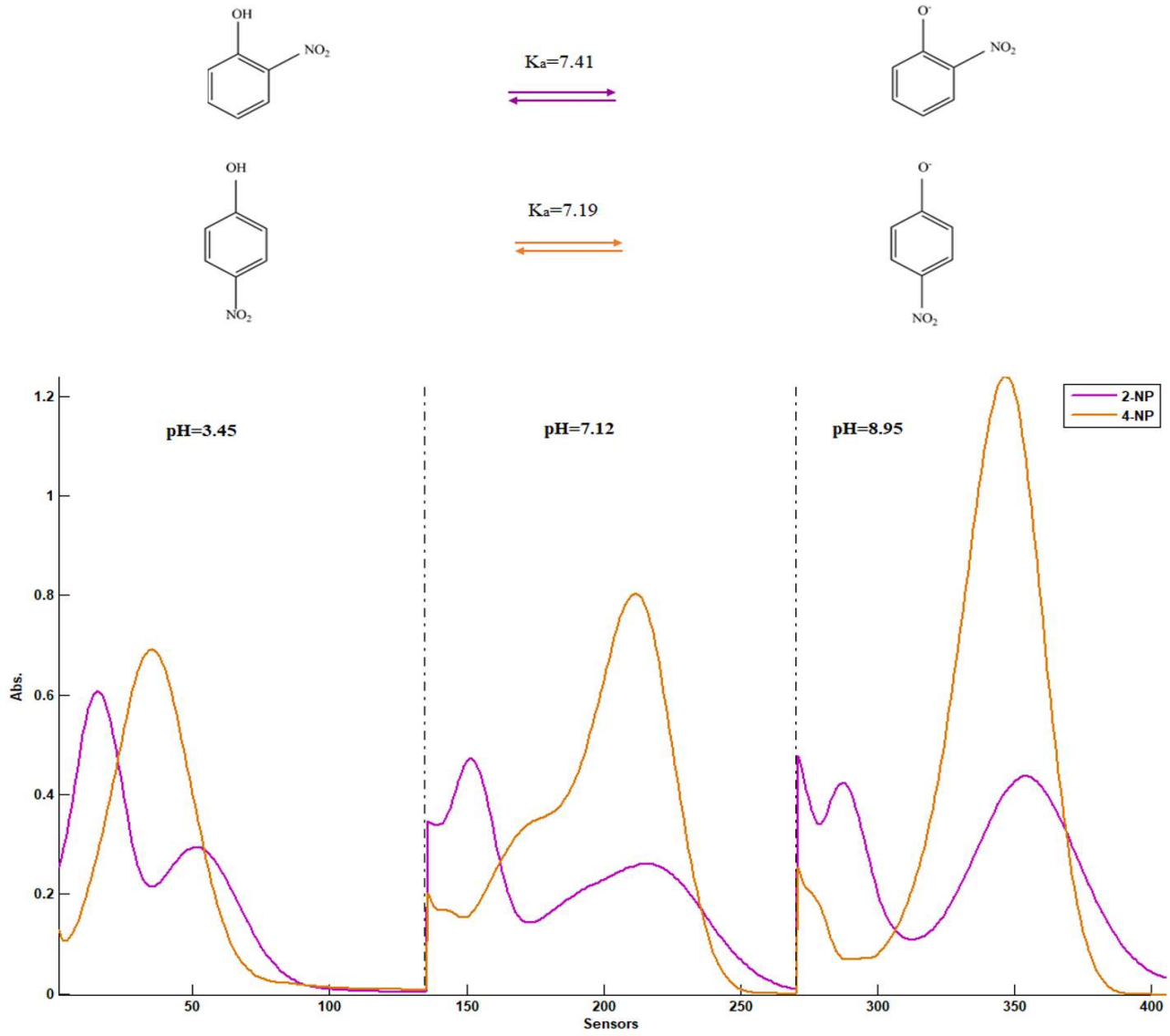

Fig. 1. Acid-based equilibrium for 2-NP and 4-NP in aqueous solution (upper part), pure spectral profiles of 2-NP (14.4 ppm) and 4-NP (9.6 ppm) at different pH*s* media.

makes the data to be more complex than those of two other pHs.

Figure 1 illustrates the pure spectra of individual 2-NP and 4-NP at different three pHs. As seen, the absorption spectra of two analytes, at all pH, are severely overlapped as analytes' quantification is not possible without previous chemical separation. Here, to overcome this signal overlapping problem, it is proposed to apply the PLS strategy to the multivariate spectrophotometric data of the analytes. In the PLS method, two data sets are needed: calibration and prediction. In this study, twenty-seven samples including different concentrations of 2-NP and 4-NP were synthesized (Table 1); twenty-five of which were designed based on a five-level full factorial design of two dye concentrations and two of which were the standard solution of the pollutants.

The same sample set including 27 samples was synthesized at three pH values including 3.45, 7.12, and 8.95. The spectrum of each sample was recorded in the wavelength range 250 to 520 with 2 nm intervals (Fig. 2). At each pH,

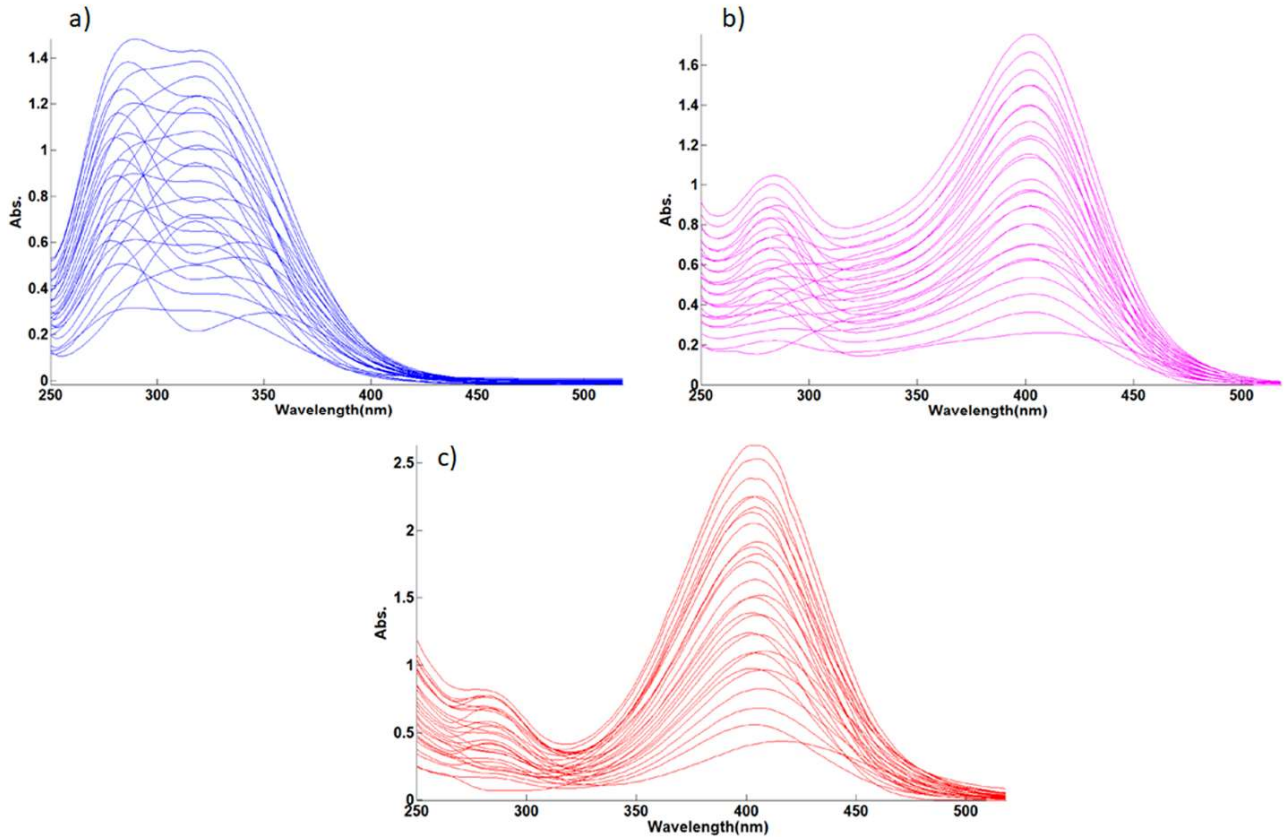

Fig. 2. Spectral data of the samples containing 2-NP and 4-NP at a) 3.45 of pH, b) 7.12 of pH and c) 8.95 of pH.

four out of 27 samples were selected for the prediction set, and the others were used as a calibration set. The collected calibration and prediction data were matrices with dimensions 135×23 and 135×4, respectively.

The PLS analysis was conducted on five types of data, three of which were collected data at mentioned pHs and the others were two augmented data sets. First, it was proposed to augment the total data (denoted as $\mathbf{D}_{total}$) to increase the information of data and in turn, improve the FOM for both analytes. It should be noted only neutral and anionic forms of analytes exist in the solution at 3.45 and 8.95 pHs, respectively; and the data at 7.12 pH is the most complex because both forms of analytes are present in the solution and as expected it could not add new information of data when it is augmented on data. Therefore, the second augmented data was constructed using the data at 3.45 and 8.95 of pHs, denoted as $\mathbf{D}_{1,3}$.

For each data set, PLS analysis was run by MVC1 software, and the LOOCV method was applied to the calibration data to find the optimal number of latent variables (LV) in the analysis, in obtaining the calibration model. The upper panel in Fig. 3 shows the cross-validation graph for 4-NP in data set $\mathbf{D}_{1,3}$; as seen the minimum value of PRESS was obtained at 5 number of LV. Although it seems that two samples out of 23 are diagnosed as outliers (lower panel, on the left of Fig. 3), these samples were not excluded from the calibration set because the calibration graph (lower panel, on the right) confirms a high correlation between the predicted and real concentration of 4-NP.

Figure 4 illustrates the plot of predicted vs. nominal concentrations of 4-NP and the elliptic joint confidence region (EJCR). The point (0, 1) lies inside the EJCR confirms bias is absent [15] and consequently, the prediction ability may be taken as 100% on a percentile scale. The prediction error vs. real concentration of 4-NP is shown in Fig. 4 (lower panel, on the left), which confirms the calibration model can estimate analytes in unknown samples efficiently.

Table 1. Sample Set for PLS Analysis: * are Prediction Samples and the Others are Calibration Set

Sample No.	2-NP (ppm)	4-NP (ppm)
1	4.8	3.2
2	9.6	3.2
3	14.4	3.2
4	19.2	3.2
5	24	3.2
6	4.8	6.4
7*	9.6	6.4
8	14.4	6.4
9*	19.2	6.4
10	24	6.4
11	4.8	9.6
12	9.6	9.6
13	14.4	9.6
14	19.2	9.6
15	24	9.6
16	4.8	12.8
17*	9.6	12.8
18	14.4	12.8
19*	19.2	12.8
20	24	12.8
21	4.8	16
22	9.6	16
23	14.4	16
24	19.2	16
25	24	16
26	14.4	0
27	0	9.6

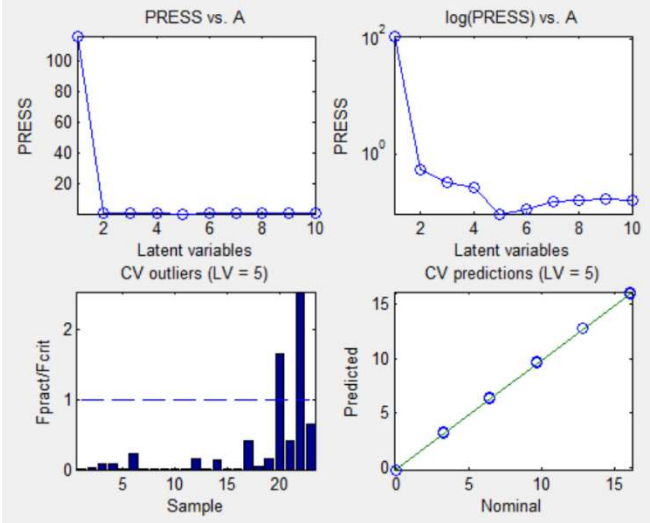

Fig. 3. Variation of the PRESS as a function of the number of latent variables; outlies diagnostics graph, Predicted *vs*. actual values for calibration sample set in PLS analysis of data set $\mathbf{D}_{1,3}$.

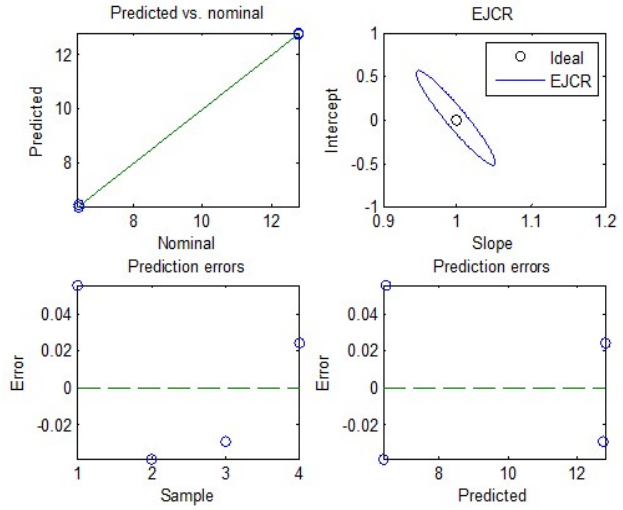

Fig. 4. Predicted *vs*. actual values for prediction sample set and EJCR; prediction error *vs*. sample number or real concentration of prediction sample set in PLS analysis of data set $\mathbf{D}_{1,3}$.

Figure 5 (upper panel, on the left) shows there is no outlier sample in the prediction set.

Figure 5 (upper panel, on the left) portrays the score plot of the whole data which reveals that 25 out of 27 total samples have been synthesized based on a 5-level full factorial design for two analytes, interestingly enough, this interpretation coincides with the above-mentioned strategy for samples' preparation. Moreover, Fig. 5 (lower panel, left) provides complementary graphs to analyze regression coefficients. Compared to Fig. 1, it can be concluded the high absolute coefficient values in this panel correspond to the sensors assigning to the maximum (or nearly maximum) of responses in pure spectra of the analytes (Fig. 1).

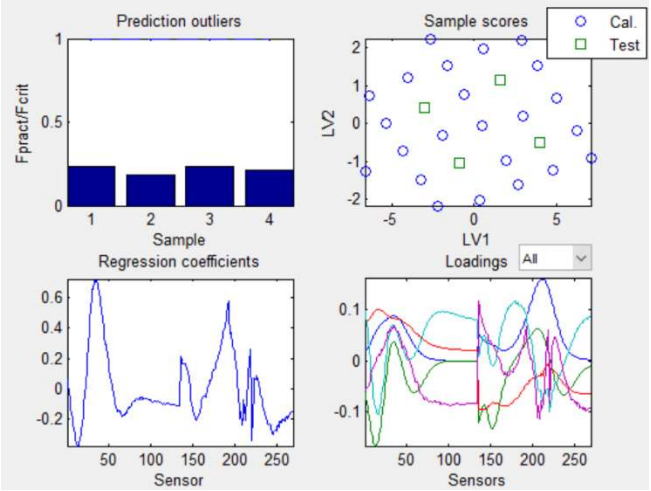

Fig. 5. Outlier diagnostics graph, score plot, regression coefficient *vs*. sensors, and loafing plot for PLS analysis of data set $\mathbf{D_{1,3}}$.

The generated graphs for the other data sets analysis were not illustrated here, due to the limitation of space. The obtained FOM including R^2, RMSEC, RMSEP, sensitivity, analytical sensitivity, LOD, and LOQ for each analyte calibration were calculated and reported in Tables 2 and 3 for2-NP and 4-NP, respectively.

It can be concluded from Tables 2 and 3 that the FOM of calibration for both analytes is the best one at 3.45 pH, where both analytes are in their neutral forms. Furthermore, the FOM for 4-NP was better than those of 2-NP at this optimal pH. The results revealed that the best FOM of calibration and the least RMSEP for the prediction set for both analytes were achieved when PLS was applied on $\mathbf{D_{1,3}}$. Although $\mathbf{D_{total}}$ was composed of three data sets, it was not the optimal data for the quantification of both analytes because the data at 7.12 pH could not add new information to $\mathbf{D_{1,3}}$. Indeed, in this pH all forms of analytes were present and it increases the complexity of data without adding any new information.

It should be mentioned that nitrophenols have three isomers: 2-NP, 3-NP, and 4-NP, and based on a literature search, especially 2-NP and 4-NP are widely used in the manufacturing of versatile compounds such as petrochemicals, pharmaceuticals, and pesticides. Therefore, these two isomers are mainly present in industrial wastewater. For the 3-NP compound, the pKa is 8.36 and the spectra of both neutral and anionic forms severely overlap with the signals of 2-NP and 4-NP. However, in an industrial process that employed all three types of isomers, the applied technique in this work can be extended to quantify three isomers simultaneously or quantify 2-NP and 4-NP in the presence of 3-NP as interference.

Table 2. FOM of Different Data Sets for 2-NP Obtained from PLS Analysis

Data	R^2	SEN	Anal. Sen	LOD min	LOD max	LOQ min	LOQ max	RMSEP
Data1	1	0.0304	93.2292	0.0949	0.1676	0.2848	0.5029	0.051
Data2	0.9991	0.1195	104.6667	0.1793	0.2389	0.538	0.7166	0.3018
Data3	1	0.0701	36.5631	0.146	0.2503	0.4379	0.7508	0.0998
Data 1,3	1	0.1016	108.3393	0.0765	0.1483	0.2295	0.4449	0.0411
Total	0.9999	0.1773	98.6392	0.0956	0.1612	0.2868	0.4836	0.1306

Table 3. FOM of Different Data Sets for 4-NP Obtained from PLS Analysis

Data	R^2	SEN	Anal. Sen	LOD min	LOD max	LOQ min	LOQ max	RMSEP
Data1	1	0.2236	256.7244	0.0791	0.1257	0.2372	0.3772	0.0383
Data2	0.9999	0.2943	257.7866	0.155	0.2072	0.4651	0.6216	0.099
Data3	1	0.2059	120.8002	0.109	0.2039	0.327	0.6116	0.049
Data 1,3	0.9999	0.2552	163.1907	0.0627	0.1122	0.1882	0.3366	0.0388
Total	1	0.2778	189.0058	0.0574	0.1082	0.1722	0.3247	0.0454

CONCLUSION

In this study, the PLS method as a genuine multivariate calibration method was applied to the collected spectrophotometric data containing 2-NP and 4-NP analytes to quantify the analytes simultaneously. The absorption signals of the analytes were too overlapped to determine the analytes in the presence of each other without chemical separation, and the proposed method makes analytes' quantification possible without any effort to conduct tedious pre-separation procedures and use expensive instrumentation techniques such as chromatographic methods. In addition, the FOM of the calibrations of analytes was investigated thoroughly at different pH conditions since both analytes possessed acid-base properties. The results showed the best FOM is obtained when the data is augmented at the pHs where only neutral or anionic forms of analytes are present in the solution.

REFERENCES

[1] R. Abazari, A.R. Mahjoub, G. Salehi, J. Hazard. Mater. 365 (2019) 921.

[2] N.Wang, G. Lv, L. He and X. Sun, J. Hazard. Mater. 403 (2021) 123805.

[3] V.V. Kadam, S.D. Shanmugam, J.P. Ettiyappan, R.M. Balakrishnan, Environ. Sci. Pollut. Res. 28 (2021) 12119.

[4] Z.Y. Lu, X.T. Lu, Y.H. Zhong, Y.F. Hu, G.K. Li, R.K. Zhang, Anal. Chim. Acta 1050 (2019) 146.

[5] X.F. Gu, Y. Zhao, K. Li, M.X. Su, F. Yan, B. Li, Y.X. Du, B. Di, J. Chromatogr. A 1474 (2016) 130.

[6] J. Fischer, J. Barek, J. Wang, Electroanal 18 (2006) 195.

[7] S. Clavijo, J. Avivar, R. Su´arez, V. Cerd`a, Trends Analyt. Chem. 67 (2015) 26.

[8] Y. Fang, D. Wang, X. Lv, X. Xu, H. Zhou, P. Liu, B. Cui, L. Wang, Sens Actuators B Chem 333 (2021) 129568.

[9] S.M. Sajjadi, Z. Asadollah-pour, S.H. Sajjadi, S.N. Nabavi, Z. Abed, F. Farzin, A. Emadi, B. Abdous, New J. Chem. 45 (2021) 12974.

[10] H.F. Grahn, P. Geladi, Techniques and applications of hyperspectral image analysis, Wiley, 2007.

[11] A.C. Olivieri, The Inverse Least-Squares Model. In: Introduction to Multivariate Calibration. Springer, Cham., 2018.

[12] B.R. Kowalski, M.B. Seasholtz, J. Chemom. 5 (1991) 129.

[13] J.P.M. Andries, G.H. Tinnevelt, Y.V. Heyden, Talanta 239 (2022) 123140.

[14] R. Khani, R. Rahmanian, N.V. Motlagh, Food Anal. Methods 9 (2016) 1112.

[15] A.C. Olivieri, H.C. Goicoechea, F.A. In˜o´n, Chemom. Intell. Lab. Syst. 73 (2004) 189.

[16] J. Ferre´, N.K.M. Faber, Chemom. Intell. Lab. Syst. 69 (2003) 123.

[17] G.M. Escandar, A.C. Olivieri, Practical Three Way Calibration, Elsevier Inc. 2014.

[18] A.C. Olivieri, Chem. Rev. 114 (2004) 5358.