European Journal of Chemistry
Journal homepage: www.eurjchem.com

First derivative-supervised pattern recognition for the flow-injection spectrophotometric discrimination of *s*-triazines in water

Judith Amador-Hernández[a,b,*], Miguel Velázquez-Manzanares[a,b] and Magaly Peral-Torres[a]

[a] *Ingeniería Ambiental, Universidad del Mar, Ciudad Universitaria Sin Numero, Puerto Ángel Oaxaca, 70902, México*
[b] *Instituto de Biotecnología, Universidad del Papaloapan, Circuito Central 200, Tuxtepec Oaxaca, 68301, México*

**Corresponding author at: Ingeniería Ambiental, Universidad del Mar, Ciudad Universitaria Sin Numero, Puerto Ángel Oaxaca, 70902, México. Tel.: +52.287.8759240; fax: +52.287.8759240. E-mail address: amadorjudith@live.com.mx (J. Amador).*

ARTICLE INFORMATION

Received: 02 February 2011
Received in revised form: 07 March 2011
Accepted: 08 March 2011
Online: 30 June 2011

KEYWORDS

Triazines
Pattern recognition
Water
Statistical data analysis
s-Triazines
Flow-injection

ABSTRACT

A new spectrophotometric method is presented for the automated differentiation of chloro- (0.5 to 5.0 μg/mL), thio- (0.5 to 5.0 μg/mL) and methoxy-triazines (1 to 10 μg/mL) in water samples. Classification models obtained by K-nearest neighbours, Soft Independent Modeling of Class Analogy, and Partial Least Squares-Discriminatory Analysis were constructed from zero order and first derivative absorption spectra as independent variables, in the spectral range from 210 to 270 nm. Binary responses were used as classifying variables (with/without certain group of triazines). With this dichotomous structure, parameters related to 2x2 contingency tables were used to evaluate the performance of the models. For tap and well water samples, sensitivity and selectivity values equal or higher than 50 % were obtained from autoscaled first derivative spectra, discriminated by Partial Least Squares-Discriminatory Analysis.

1. Introduction

Triazines represent more than 30% of the herbicides manufactured in the world. These compounds are used to control pre- and post-emerging weeds in the production of more than 50 crops in hundreds of countries, as well as in forestry conservation, and in home and garden care [1].

There are three groups of compounds that mainly comprise the s-triazines (s-TRZ), which share a six-member ring with atoms of carbon and nitrogen symmetrically distributed. Their main distinction is given by the substitution in position 2: a) a chlorine, in the case of chloro-triazines (CT, e.g. atrazine, simazine), b) a methoxyl group in methoxy-triazines (MT, e.g. prometron, atraton), and c) a methylthio group in thio-triazines (TT, e.g. terbutryn, ametryn). Substitutions in the fourth and sixth positions are generally amino alkyl groups, which are of less relevance in the biocide activity of these compounds [2-3].

The quantitative determination of these herbicides, as well as others pesticides, is usually carried out through separation-detection techniques, where a wide variety of analytical methods are available [4,5]. Techniques such as gas or liquid chromatography combined with mass spectrometry (GC-MS or LC) are well recognized as powerful analytical tools owing to their great selectivity and sensibility, in spite of the time-consuming, high-cost procedures involved [6]. In contrast, screening methods are less explored. From an analytical point of view, screening methods can be considered as simple and cost-effective, which generate less specific but short time responses, also useful [7-9]. Some authors have reported the development of screening methods based on chemometric strategies. Particularly, techniques such as Partial Least Squares Regression-Discriminatory Analysis (PLS-DA) and Soft Independent Modeling of Class Analogy (SIMCA) have been shown to be important tools in the resolution of qualitative problems [9-11]. Based on Principal Component Analysis (PCA), both are capable of finding intersample and intervariable relationships, and of reducing dimensionality of data.

In relation to the screening of triazines in water, Carabias-Martínez et al. reported the on-line simultaneous determination of both CT and TT in mixtures without chromatographic separation, by means of PLS [12]; however, MT were not included. Also, Beale et al. developed a fast screening method for the presence of atrazine and other triazines in water using flow injection with chemiluminescent detection, after the herbicides react with tris(2,2′-bipyridyl)ruthenium(III) [13].

Particularly, our group was interested to evaluate if chemometric tools enabled the differentiation between the three groups of triazines, in spite of the great overlapping between their absorption spectra, as a previous effort for the development of a quantitative exercise. Therefore, this work presents the discrimination of chloro-, thio, and methoxy-triazines in water through UV-Visible spectrophotometry. Sample manipulation is automated through a flow injection system (FI). Chemometric techniques, namely k-nearest neighbors (KNN), SIMCA, and PLS-DA are used to analyze the zero order and first derivative absorption spectra and results are discussed in each case.

2. Experimental

2.1. Instrumentation

An UV-Visible spectrophotometer (Model Lambda EZ210, Perking Elmer) was used, controlled via a Pentium IV computer

ISSN 2153-2249 (Print) / ISSN 2153-2257 (Online)
DOI:10.5155/eurjchem.2.2.146-151.398

(Dell). The acquisition and storage of data were carried out with a PESSW v. 1.2.E. software (Perkin Elmer). The Pirouette v. 3.11 software (Infometrix Inc.) was also used for data treatment. The FI manifold is depicted in Figure 1. It consists of a) a peristaltic pump (Model Minipuls 3, Gilson), b) a low pressure injection valve (Model 504, Rheodyne), c) PTFE tubing of 0.5 mm i.d., and d) a flow cell with an optical path of 10 mm (Hellma), located in the detector.

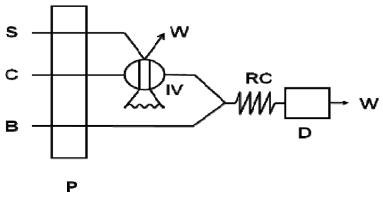

Figure 1. Dynamic manifold used for the discrimination of TRZ: S, sample solution; C, carrier solution; B, buffer solution; P, peristaltic pump; IV, injection valve; RC, reaction coil; D, detector; W, waste.

2.2. Reagents and solutions

All chemicals were of analytical-reagent grade. Atrazine, propazine, cyanazine, simazine, terbutryn, simetryn, ametryn, prometryn, terbumeton, atraton, and prometon were of pestanal grade (Riedel de Häen). An ultrapure water system (EasyPure Uv, Barnstead) was used throughout.

Stock solutions of the individual pesticides containing 100 μg/mL were prepared in methanol. All solutions were stored at 4 °C and kept stable for at least one month. Working standard solutions were prepared daily by adequate dilution. A buffer solution of KH_2PO_4/$NaKHPO_4$ 0.5 mol/L, pH = 6.8, was also used.

2.3. Procedure

The samples were prepared with one, two or three components, each one belonging to a different group. According to each group, the following compounds were selected as analytes: a) atrazine, propazine, simazine, and cyanazine as CT, b) ametryn, prometryn, simetryn, and terbutryn as TT, and c) atraton, prometon, and terbumeton as MT. Their concentrations varied in range from 0.5 to 5.0 μg/mL for CT and TT, and from 1 to 10 μg/mL for MT, where linearity was observed during one-compound calibration. Therefore, appropriate volumes of the stock solutions of interest were filled into a 10 mL volumetric flask; methanol content was added to complete to 20% (V/V), and the resultant mixture was diluted to volume with pure water.

For tap and well water analyses, the samples were fortified with the analytes of interest in different proportions and filtered by using a nylon membrane of 0.2 μm of pore size. Then the samples were managed in the same way as the samples described above. After that, the 500 μL loop of the injection valve was filled with the solution of each sample. This volume was injected into the methanol solution used as carrier and mixed with the buffer solution (KH_2PO_4/$NaKHPO_4$ 0.5 mol/L, pH = 6.8) in the reaction coil. After 180 seconds from injection, the analytes reached the flow cell located in the cell holder of the detector. The stream was then stopped for 70 seconds while the signal was stabilized and the absorption spectrum was recorded. Finally, the sample was driven away from the flow cell, through to the waste. In all cases, samples were injected in triplicate.

The absorption spectrum of each sample was recorded in the range of 200 to 300 nm against a reagent blank with a resolution of 0.2 nm and used as analytical data for sample classification with KNN, SIMCA, and PLS-DA techniques. Absorbance values lower than 1.2 were obtained in all cases.

3. Results and discussion

The chemical structures of the triazines of interest in this work are presented in Figure 2. The absorption spectra for atrazine (CT), ametryn (TT), and atraton (MT) are shown in Figure 3, as well as for atraton, terbumeton and prometon (all MT) in Figure 4. As can be seen, spectral profiles are more different between s-triazine groups than between compounds within a group. It seems that the functional group located in position 2 leads to more influence in the UV absorption spectra than substitution of amino alkyl groups in positions 4 and 6 (Figure 4), at least in absorption maximum and inflection points locations. From these observations, it was proposed the discrimination of CT, TT and MT in water by the use of supervised pattern recognition techniques.

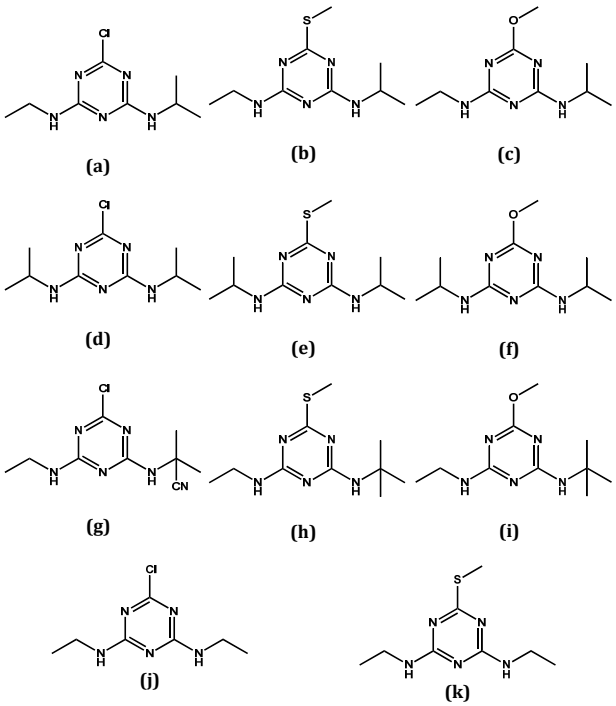

Figure 2. Molecular structures of the triazine herbicides: a) atrazine, b) ametryn, c) atraton, d) propazine, e) prometryn, f) prometon, g) cyanazine, h) terbutryn, i) terbumeton, j) simazine, k) simetryn.

3.1. Preliminary studies

Initially, the effect of pH on the absorption spectra of CT, TT and MT was studied. Two representative series of compounds were selected: a) atrazine (CT), ametryn (TT) and atraton (MT), and b) propazine (CT), prometryn (TT) and prometon (MT). A solution of 250 mL for each compound was prepared with a concentration between 2 and 4 μg/mL in 0.1 M KCl. The pH was varied from 1.0 to 11.0 with aqueous HCl or KOH solutions at different concentrations.

Major differences between the band shapes of the compounds were present at acidic conditions; however, a hypochromic effect was also observed while pH decreased. As an intermediate condition, a buffer solution of KH_2PO_4/$NaKHPO_4$ pH = 6.8 was selected as suitable.

3.2. Flow Injection system optimization

Variable optimizations were divided into chemical and hydrodynamic groups, and were performed by using the univariate method with atrazine as analyte. The absorbance at 222 nm was recorded as analytical signal. Buffer concentration was varied in the range from 0.05 to 0.50 mol/L. This did not influence the band shape of the analyte; 0.5 mol/L was chosen as an appropriate concentration. The methanol content between 0.5 and 5.0 % V/V was evaluated in the carrier solution, choosing 1% V/V as optimum.

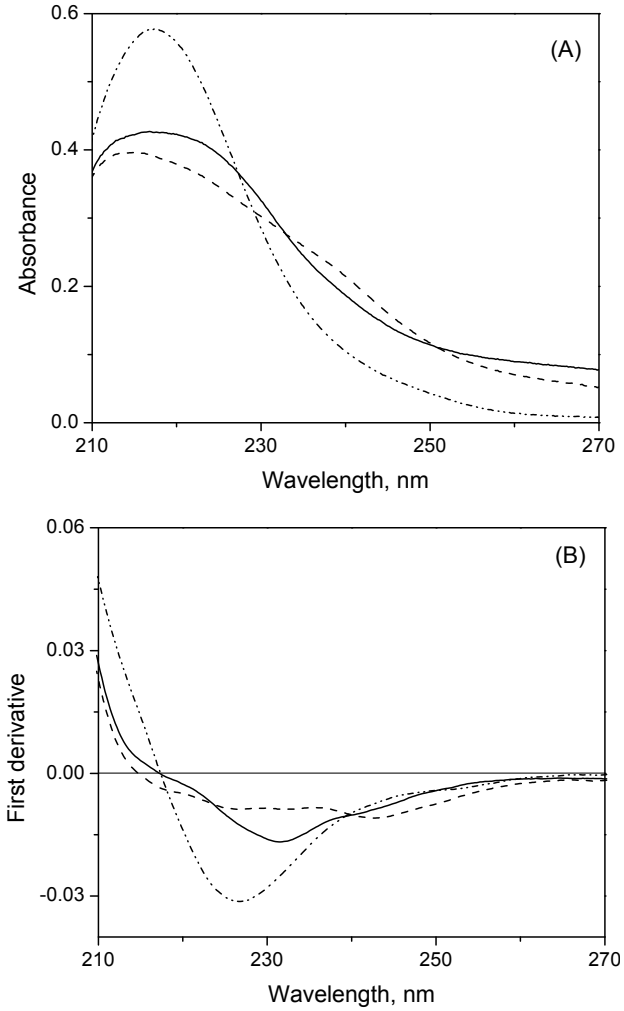

Figure 3. Absorption spectra of atrazine (3.0 μg/mL, solid line), ametryn (2.5 μg/mL, dash line) and atraton (7.0 μg/mL, dash dot dot line) under the proposed experimental conditions: (A) zero order, (B) first derivative.

In the case of the hydrodynamic variables, the flow rates of both the carrier and the buffer solutions varied in the range from 0.2 to 1.2 mL/min. A flow rate of 0.5 mL/min was selected as the most proper in both cases. The volumes of the injection loop and the reaction coil were also evaluated, both between 250 and 1000 μL. Volumes of 500 μL for the injection loop and 250 μL for the reaction coil were appropriate for next experiments. A sample throughput of 9 per hour was reached under the proposed conditions.

3.3. Chemometric strategies

Three supervised pattern recognition techniques were considered to develop classification rules with the capability to predict the category membership of new and unknown samples: KNN, SIMCA and PLS-DA. For supervised learning, a training set of 50 samples was prepared, with one, two or three components of the different groups and at distinct ratios (see Table 1), to incorporate as much variability as possible to the system [9,14]. Also, a validation set of 30 samples was used to rate the performance of these classification rules (Table 2), as well as a set of tap and well water samples fortified with the compounds of interest, which composition is described in Table 3.

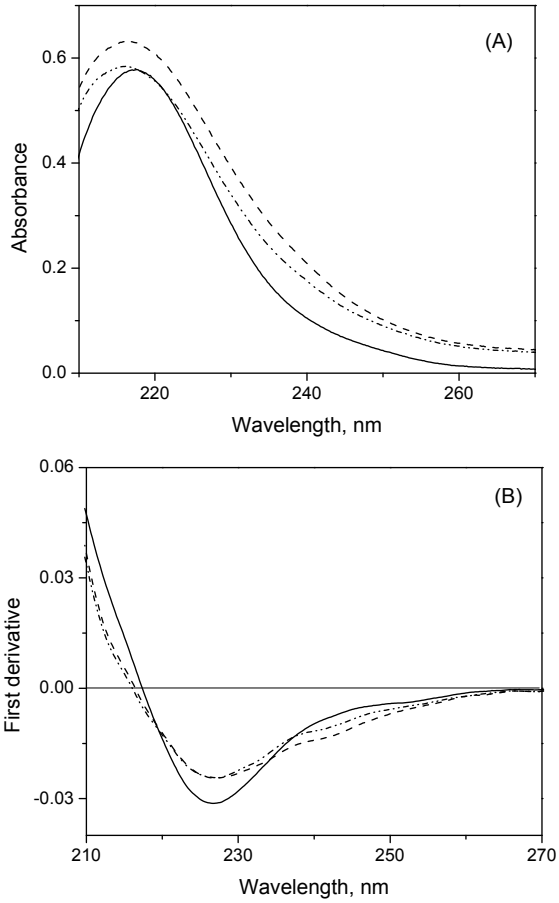

Figure 4. Absorption spectra of methoxy-triazines known as atraton (7.0 μg/mL, solid line), terbumeton (5 μg/mL, dash line) and prometon (8 μg/mL, dash dot dot line) under the proposed experimental conditions: (A) zero order, (B) first derivative.

Triazine groups (CT, TT, or MT) were analyzed individually. A binary classifier was applied to each group for simplicity reasons. In other words, samples were viewed in only two classes: positive (1, with a triazine of the group of interest), or negative (0, without a triazine of the group of interest). Thus, qualitative results could be tested through 2x2 contingency tables.

According to preliminary studies of original independent variables through PCA and SIMCA based on Modeling Power and Discriminant Power [15], valuable information was comprised between 210 and 270 nm. Therefore, this spectral region was selected in further analyses. Also, preprocessing (mean centering, autoscaling) and transformation of data (smoothing with 15 points and first derivative with five points) were considered. Other smoothing and first derivative conditions were studied, but the best results were obtained with the above conditions (see Figure 3B); therefore, they were selected in further calculations.

In order to estimate the performance of discriminant analyses, the following expressions were used:

Table 1. Composition of the training set of samples. All concentrations are in μg/mL.

Sample	CT				TT				MT		
	AZ	SZ	PZ	CZ	TY	PY	AY	SY	PN	AN	TN
1	0.5	-	-	-	0.5	-	-	-	5.0	-	-
2	-	2.5	-	-	-	2.5	-	-	5.0	-	-
3	-	-	5.0	-	3.0	-	-	-	-	10.0	-
4	-	1.0	-	-	-	5.0	-	-	-	-	6.0
5	1.0	-	-	-	-	3.0	-	-	-	-	10.0
6	3.0	-	-	-	-	-	1.0	-	-	6.0	-
7	-	-	-	3.0	-	-	5.0	-	-	2.0	-
8	5.0	-	-	-	-	-	-	1.0	-	-	10.0
9	-	-	5.0	-	-	3.0	-	-	2.0	-	-
10	-	-	-	2.0	-	-	-	2.0	-	9.0	-
11	-	-	4.5	-	-	-	2.0	-	4.0	-	-
12	2.0	-	-	-	4.5	-	-	-	-	4.0	-
13	-	-	-	-	0.5	-	-	3.5	-	-	7.0
14	-	3.5	-	-	-	-	0.5	-	-	7.0	-
15	-	-	3.5	-	-	3.5	-	-	1.0	-	-
16	-	2.5	-	-	-	-	-	1.5	-	-	8.0
17	2.5	-	-	-	-	-	-	4.0	-	3.0	-
18	-	-	4.0	-	1.5	-	-	-	-	-	5.0
19	-	-	-	1.5	-	2.5	-	-	8.0	-	-
20	-	1.5	-	-	4.0	-	-	-	5.0	-	-
21	-	-	-	4.0	-	-	2.5	-	-	3.0	-
22	-	-	1.0	-	-	-	-	2.0	-	-	-
23	2.0	-	-	-	-	-	-	-	-	-	6.0
24	-	-	-	-	-	-	3.0	-	-	4.0	-
25	-	-	-	4.5	-	0.5	-	-	-	-	-
26	0.5	-	-	-	-	-	-	-	-	-	9.0
27	-	-	-	0.5	-	-	-	-	-	-	-
28	-	2.5	-	-	-	-	-	-	-	-	-
29	-	-	-	2.5	-	-	-	-	-	-	-
30	3.0	-	-	-	-	-	-	-	-	-	-
31	-	-	3.5	-	-	-	-	-	-	-	-
32	4.0	-	-	-	-	-	-	-	-	-	-
33	-	4.5	-	-	-	-	-	-	-	-	-
34	-	-	5	-	-	-	-	-	-	-	-
35	-	-	-	-	0.5	-	-	-	-	-	-
36	-	-	-	-	-	0.5	-	-	-	-	-
37	-	-	-	-	1.0	-	-	-	-	-	-
38	-	-	-	-	-	-	-	2.0	-	-	-
39	-	-	-	-	-	-	2.5	-	-	-	-
40	-	-	-	-	3.0	-	-	-	-	-	-
41	-	-	-	-	-	3.5	-	-	-	-	-
42	-	-	-	-	-	-	-	4.0	-	-	-
43	-	-	-	-	-	-	4.5	-	-	-	-
44	-	-	-	-	-	-	-	-	-	-	1.5
45	-	-	-	-	-	-	-	-	1.0	-	-
46	-	-	-	-	-	-	-	-	-	-	2.0
47	-	-	-	-	-	-	-	-	-	3.0	-
48	-	-	-	-	-	-	-	-	-	-	10
49	-	-	-	-	-	-	-	-	5.0	-	-
50	-	-	-	-	-	-	-	-	-	7.0	-

S-triazine groups: CT, chloro-triazines; TT, thio-triazines, MT, methoxy-triazines. Individual compounds: AZ, atrazine; SZ, simazine; PZ, propazine; CN, cyanazine; TY, terbutryn; PY, prometryn; AY, ametryn; SY, simetryn; AN, atraton; PN, prometron; TN, terbumeton.

$$\text{Sensitivity} = \frac{\text{Number of resulting true positives}}{\text{Total number of actual positives}} = \frac{tp}{tp+fn} * 100 \quad (1)$$

$$\text{Specificity} = \frac{\text{Number of resulting true negatives}}{\text{Total number of actual negatives}} = \frac{tn}{tn+fp} * 100 \quad (2)$$

$$PPV = \frac{\text{Number of resulting true positives}}{\text{Total number of resulting positives}} = \frac{tp}{tp+fp} * 100 \quad (3)$$

$$NPV = \frac{\text{Number of resulting true negatives}}{\text{Total number of resulting negatives}} = \frac{tn}{tn+fn} * 100 \quad (4)$$

$$FPR = \frac{\text{Number of resulting false positives}}{\text{Total number of actual negatives}} = \frac{fp}{fp+tn} * 100 \quad (5)$$

$$FNR = \frac{\text{Number of false negatives}}{\text{Total number of actual positives}} = \frac{fn}{tp+fn} * 100 \quad (6)$$

where *tp* is true positive, *tn* is true negative, *fp* is false positive, *fn* is false negative, *PPV* is the Positive Predictive Value, *NPV* is the Negative Predictive Value, *FPR* is the False Negative Rate, and *FNR* is the False Positive Rate [16,17]. Some of the performance parameters obtained through the KNN, SIMCA, and PLS-DA techniques are summarized in Table 4.

Table 2. Composition of the validation set of samples. All concentrations are in μg/mL.

Sample	CT				TT				MT		
	AZ	SZ	PZ	CZ	TY	PY	AY	SY	PN	AN	TN
1	2.3	-	-	-	1.1	-	-	-	-	-	6.3
2	-	1.8	-	-	-	-	-	1.9	-	7.7	-
3	-	-	3.8	-	3.9	-	-	-	7.5	-	-
4	1.3	-	-	-	-	-	-	2.3	-	-	9.3
5	4.8	-	-	-	-	-	-	0.9	-	-	5.5
6	-	2.4	-	-	4.7	-	-	-	-	1.7	-
7	-	1.6	-	-	-	2.3	-	-	-	4.5	-
8	-	-	-	-	-	4.8	-	-	-	-	3.7
9	-	-	1.1	-	-	-	-	-	3.8	-	-
10	2.7	-	-	-	-	-	-	-	9.3	-	-
11	-	-	-	4.9	-	-	2.7	-	-	-	-
12	-	-	2.2	-	-	-	4.4	-	-	-	-
13	-	-	-	2.6	-	-	-	-	2.5	-	-
14	1.5	-	-	-	-	4.3	-	-	-	-	-
15	-	-	4.2	-	-	-	-	-	-	3.2	-
16	-	-	-	3.0	-	-	5.0	-	-	2.1	-
17	-	-	-	-	-	-	-	-	-	-	4.8
18	-	-	4.9	-	-	-	-	-	-	-	-
19	-	-	-	-	4.6	-	-	-	-	-	-
20	-	-	-	-	-	-	-	-	-	-	8.1
21	-	-	-	0.5	-	-	-	-	-	-	-
22	-	4.3	-	-	-	-	-	-	-	-	-
23	-	-	-	1.5	-	-	-	-	-	-	-
24	-	-	-	-	-	3.5	-	-	-	-	-
25	-	-	-	-	-	-	-	3.8	-	-	-
26	-	-	-	-	-	-	1.5	-	-	-	-
27	-	-	-	-	2.5	-	-	-	-	-	-
28	-	-	-	-	-	-	-	-	1.3	-	-
29	-	-	-	-	-	-	-	-	-	-	2.6
30	-	-	-	-	-	-	-	-	-	3.1	-

S-triazine groups: CT, chloro-triazines; TT, thio-triazines, MT, methoxy-triazines. Individual compounds: AZ, atrazine; SZ, simazine; PZ, propazine; CN, cyanazine; TY, terbutryn; PY, prometryn; AY, ametryn; SY, simetryn; AN, atraton; PN, prometron; TN, terbumeton.

Table 3. Composition of the test series consisting in tap and well water samples fortified with chloro-, thio- and/or methoxy-triazines; all concentrations are in μg/mL. Well water samples were made up the same way as tap water samples (ten samples in total).

Sample	CT				TT				MT		
	AZ	SZ	PZ	CZ	TY	PY	AY	SY	PN	AN	TN
1	2.0	-	-	-	-	2.0	-	-	-	-	2.0
2	-	-	-	3.0	-	-	-	-	-	-	-
3	-	-	-	-	-	4.0	-	-	-	-	-
4	-	-	-	-	-	-	-	-		2.8	-
5	-	-	1.0	-	-	-	-	-	1.0	-	-

S-triazine groups: CT, chloro-triazines; TT, thio-triazines, MT, methoxy-triazines. Individual compounds: AZ, atrazine; SZ, simazine; PZ, propazine; CN, cyanazine; TY, terbutryn; PY, prometryn; AY, ametryn; SY, simetryn; AN, atraton; PN, prometron; TN, terbumeton.

3.3.1. K-nearest neighbors

This is a simple but powerful classification technique, commonly used as a standard when comparing pattern recognition procedures [18]. In this work, the technique was applied for the classification of samples associated with zero order and first derivative absorption spectra, each one under mean center or autoscale preprocessing. Also, the discriminant capability of PCA-KNN was studied, taking into account that absorption spectra are collinear in an extensive way [19]. Therefore, KNN was applied to scores of relevant principal components (PC) obtained both from preprocessed, zero order and first derivative absorption spectra. In each case, one was selected as the number of maximum neighbors.

For PCA, the full cross-validation by leaving out one sample was carried out in all cases, and the corresponding PRESS

Table 4. Performance parameters obtained for training and validation series with the optimum models for KNN, SIMCA and PLS-DA. Percentage values equal or higher than 90 are highlighted.

Parameter[1]	Calibration Set								
	KNN			SIMCA			PLS-DA		
	CT	TT	MT	CT	TT	MT	CT	TT	MT
Sensitivity	**94**	87	**94**	**100**	**94**	**100**	84	80	**94**
Specificity	61	84	79	**100**	**100**	89	**100**	**100**	84
PPV[2]	81	**90**	88	**100**	**100**	**94**	**100**	**100**	**91**
NPV[3]	85	80	88	**100**	**90**	**100**	78	76	89
FPR[4]	39	16	21	0	0	11	0	0	16
FNR[5]	6	13	6	0	6	0	16	20	6
UnS[6]	-	-	-	0	0	0	0	0	0
	Validation Set								
	KNN			SIMCA			PLS-DA		
	CT	TT	MT	CT	TT	MT	CT	TT	MT
Sensitivity	83	**100**	**100**	**94**	**100**	92	78	81	86
Specificity	83	50	38	**91**	73	50	**100**	**100**	75
PPV[2]	88	75	58	**94**	84	63	**100**	**100**	75
NPV[3]	77	**100**	**100**	**91**	**100**	88	75	80	86
FPR[4]	17	50	63	9	27	50	0	0	25
FNR[5]	17	0	0	6	0	8	22	19	14
UnS[6]	-	-	-	1	3	4	0	2	0

[1] S-triazine groups: CT, chloro-triazines; TT, thio-triazines, MT, methoxy-triazines.
[2] Positive Predictive Values.
[3] Negative Predictive Values.
[4] False Negative Rate.
[5] False Positive Rate.
[6] Unclassified samples (by definition, this type of samples does not exist in KNN).

Table 5. Performance parameters estimated in the classification of real samples (tap and well water), by means of the three supervised pattern recognition techniques. Percentage values equal or higher than 90 are highlighted.

	KNN			SIMCA			PLS-DA		
	CT	TT	MT	CT	TT	MT	CT	TT	MT
Sensitivity	83	75	50	83	**100**	**100**	67	**100**	50
Specificity	**100**	67	**100**	**100**	33	25	**100**	67	**100**
PPV[2]	**100**	60	**100**	**100**	50	67	**100**	67	**100**
NPV[3]	80	80	57	80	**100**	**100**	67	**100**	57
FPR[4]	0	33	0	0	67	75	0	33	0
FNR[5]	16	25	50	17	0	0	33	0	50
UnS[6]	0	0	0	0	0	0	0	0	0

[1] S-triazine groups: CT, chloro-triazines; TT, thio-triazines, MT, methoxy-triazines.
[2] Positive Predictive Values.
[3] Negative Predictive Values.
[4] False Negative Rate.
[5] False Positive Rate.
[6] Unclassified samples (by definition, this type of samples does not exist in KNN).

(Prediction Residual Error Sum of Squares) was calculated. Further, the F-test criterion recommended by Haaland and Thomas and the first local minimum value of PRESS were considered for this purpose [20,21]. As a result, three PC were identified as significant, except in the case of autoscaled data of zero order absorption spectra, where two PC were chosen. Thus, scores of two or three PC were considered as independent variables for the classification of samples through KNN.

Better discriminatory capabilities were obtained when zero order absorption spectra were autoscaled and first derivative was obtained for the transformation of data before KNN classification. Table 4 shows the parameters estimated for these KNN models. As can be observed in the performance of training and validation sets, better sensibility than specificity were obtained for the three groups, a favorable condition, considering that false positives instead of false negatives are preferred due to the nature of the assay. As can be established, an increase in sensitivity is usually gained at the expense of a decrease in specificity [16]. Also, all samples were assigned to one of the two classes; none of the samples remained unclassified, an imposition of the KNN model. On the other hand, false positives were more evident in three-component samples, affecting both specificity and FPR indicators in CT, TT and MT. As these types of samples are not included in tap and well water samples, both parameters show better values, as can be observed in Table 5.

3.3.2. Soft independent modeling of class analogy

SIMCA is a well-known pattern recognition technique, used to classify samples in complex systems [18,22]. As with the other techniques that were proposed, several mathematical models were obtained through SIMCA to find the best classification rule capable of identifying the presence of triazines belonging to a certain group (CT, TT or MT), while the presence of compounds of the other groups did not interfere.

Preprocessing and treatment of data were the same as with KNN. For the selection of the number of optimum factors, parameters such as Cumulative Variance, Interclass Residuals, and Interclass Distances were used. Two or three factors were chosen in all cases, depending on the pretreatment or transforming strategy, as well as on the group of interest. Next, the performance of the classification models was estimated, by comparing the predicted versus the actual category, in both training and validation sets.

The best results were obtained with the model based on autoscaled first derivative spectra with two factors as representatives in all cases, with the exception of the three factors required for the first category (negative, without) in the TT model. As can be appreciated in Table 4, satisfactory results in terms of sensitivities were obtained in all cases (values higher than 90 % in training and validation sets for CT, TT, and MT). False positives observed for TT and MT in validation sets lead to more unfavorable specificity than sensitivity values (with FPR of 27 and 50 %, respectively). Also, it is important to

mention that one (CT), three (TT) and four (MT) samples from the validation set remained unclassified through SIMCA, another unfavorable result when all of them contain as analytes the compounds considered in the training step. On the other hand, satisfactory results were observed for tap and well water analyses (see Table 5), although poor specificity results were obtained for TT and MT (FPR values higher than 65%).

3.3.3. Partial least squares regression-discriminatory analysis

PLS is a powerful multivariate calibration technique, useful in the quantification of macro- or micro-components in a wide variety of matrices [23]. However, it can be used for qualitative instead of quantitative purposes (PLS-DA) [18,22]. Classification of samples dealing with a dichotomous structure is done according to whether the prediction is closer to 0 or 1 (an arbitrary choice to identify both classes). In this case, the application of PLS-DA was carried out by using the same preprocessing and transformation strategies for data as with KNN and SIMCA techniques, and it was used on the same sets of samples.

The selection of the number of factors was decided through a cross validation of the training samples, by applying the local minimum and F-test criteria related to PRESS. The Cumulative Variance, Standard Error of Calibration, and the Correlation Coefficient parameters were also considered. The optimum numbers of factors were three for the CT, TT, and MT models under the different preprocessing and transformation conditions.

The models derived from the autoscaled first derivative spectra provided the best results (similar to SIMCA), which are listed in Table 4. In general, consistency between performance parameters obtained from training, validation and real samples sets could be observed for the three groups, as seen in Tables 4 and 5. Only two samples from the validation set remained unclassified (in the TT model), a more favorable condition than with SIMCA. Those results show the discriminatory capability of PLS, in spite of continuous (non-discrete) variables were obtained directly through this technique and whose values were later rounded to assign them to one of the two classes.

Finally, it must be remarked that triazinones like metribuzine, metamitrone, and hexazinone did not interfere at the same magnitude of concentrations as s-triazines, in none of the three chemometric techniques.

4. Conclusion

The proposed method proved to be a useful tool to discriminate between chloro-, thio-, and methoxy-triazines in water, even in multi-component samples, despite the high overlapping observed in absorption spectra. The pretreatment of data by autoscaling, smoothing and first derivative obtaining before the application of the supervised pattern recognition techniques was crucial in getting satisfactory results. By assuming binary categories, i.e. presence/absence of compounds pertaining to certain s-triazine group, it was possible to use parameters related to 2x2 contingency tables, in order to evaluate the performance of classification rules. In general, PLS-DA showed the best sensitivity and specificity values (higher than 75 %) in training and validation sets, although lower values were obtained in real samples, as can be expected. Finally, the simplicity of the FI manifold allowed for an acceptable sample throughput of the qualitative analyses.

Acknowledgement

The authors gratefully acknowledge financial support from the Consejo Nacional de Ciencia y Tecnología (CONACyT, México), Project J37143-E.

References

[1]. LeBaron, H. M.; McFarland, J. E.; Burnside, O. C. The Triazine Herbicides, 50 years Revolutionizing Agriculture, Elsevier, 2008.
[2]. Kamrin, M. A. Pesticide Profiles: Toxicity, Environmental Impact and Fate, CRC Press, 1997.
[3]. Montgomery, J. H. Agrochemicals Desk Reference, 2nd edition, CRC Press, 2000.
[4]. Wang, Y.; You, J.; Ren, R.; Xiao, Y.; Gao, S.; Zhang, H.; Yu, A. *J. Chromatogr. A* **2010**, *1217(26)*, 4241-4246.
[5]. Gao, S.; You, J.; Zheng, X.; Wang, Y; Ren, R.; Zhang, R.; Bai, Y.; Zhang, H. *Talanta* **2010**, *82(4)*, 1371-1377.
[6]. Sanchez-Ortega, A.; Unceta, N.; Gómez-Caballero, A.; Sampedro, M. C.; Akesolo, U.; Goicolea, M. A.; Barrio, R. J. *Anal. Chim. Acta* **2009**, *641(1-2)*, 110-116.
[7]. Romberg, B.; Müller, H. *Anal. Chim. Acta* **1997**, *353(2-3)*, 165-172.
[8]. Jurado-Sánchez, B.; Ballesteros, E.; Gallego, M. *Talanta* **2007**, *73(3)*, 498-504.
[9]. Amador-Hernández, J.; Fernández-Romero, J. M.; Luque de Castro, M. D. *Anal. Chim. Acta* **2001**, *448*, 61-69.
[10]. López-de-Alba, P. L.; López-Martínez, L.; Cerdá, V.; Amador-Hernández, J. *J. Braz. Chem. Soc.* **2006**, *17*, 715-722.
[11]. Serrano, A.; Gallego, M. *J. Chromatogr. A* **2004**, *1045(1-2)*, 181-188.
[12]. Carabias Martínez, R.; Rodríguez Gonzalo, E.; Santiago Toribio, M. P.; Hernández Méndez, J. *Anal. Chim. Acta* **1996**, *321(2-3)*, 147-155.
[13]. Beale, D. J.; Porter, N. A.; Roddick, F. A. *Talanta* **2009**, *78(2)*, 342-347.
[14]. Berzas Nevado, J. J.; Murillo Pulgarín, J. A.; Amador-Hernández, J.; Gómez Laguna, M. A. *Analyst* **1998**, *123*, 483-488.
[15]. Infometrix. User's Manual, Pirouette Multivariate Data Analysis for Windows 95/98 and NT. V. 3. 1, 2002.
[16]. Pulido, A.; Ruisánchez, I.; Boqué, R.; Rius, F. X. *TRAC-Trend Anal. Chem.* **2003**, *22(9)*, 647-654.
[17]. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. Handbook of Chemometrics and Qualimetrics Part A, Elsevier, 1997.
[18]. Lavine, B. K.; Davidson, C. E. Classification and Pattern Recognition, In Practical Guide to Chemometrics, Gemperline P. editor, 2nd edition, CRC PRESS, 2006.
[19]. Naes, T.; Isaksson, T.; Fean, T.; Davies, T. A. User-friendly Guide to Multivariate Calibration and Classification, NIR Publications, 2002.
[20]. Thomas, E. V.; Haaland, D. M. *Anal. Chem.* **1990**, *62(10)*, 1091-1099.
[21]. López-de-Alba, P. L.; López-Martínez, L.; Wrobel-Kaczmarczyk, K.; Wrobel-Zasada, K.; Amador-Hernández, J. *J. Radioan. Nucl. Ch. Le.* **1997**, *220*, 167-171.
[22]. Brereton R. G. Chemometrics, Data Analysis for the Laboratory and Chemical Plant, Wiley, 2003.