Talanta 116 (2013) 108–114

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Chemometrics for the classification and calibration of seawater using the H^+ affinity spectrum

L. Kortazar*, J. Sáez, E. Astigarraga, N. Goienaga, L. Fernández

Department of Analytical Chemistry, University of the Basque Country, P.O. Box 644, 48080 Bilbao, Spain

ARTICLE INFO

Article history:
Received 9 February 2013
Received in revised form
15 April 2013
Accepted 24 April 2013
Available online 10 May 2013

Keywords:
Seawater
Chemometrics
Pollution
Principal component analysis
Partial least square regression
Proton affinity spectrum

ABSTRACT

In 1819 Alexander Marcet proposed that seawater contains small amounts of all soluble substances and that the relative abundances of some of them were constant. This hypothesis is nowadays known as Marcet's Principle or the principle of constancy of the composition of seawater. Based on this principle, the present research tried to prove that it is possible to detect polluted seawater samples using the seawater H^+ affinity spectrum by the application of the possibilities provided by chemometric tools. Seawater samples were classified using the principal component analysis (PCA) of the H_{Bound} spectra of the samples. It was concluded that the sampling points location does not have any influence in the cluster formation, while the season in which they were collected is significant. On the other hand, the seawater composition was calibrated using estuary water samples of different salinities. Once the major constituents were measured, the data analysis concluded that it is possible to make a calibration of the H_{Bound} spectrum *vs.* any of these constituents by means of partial least square (PLS) regression. Thus, the experimental evidence collected in this work confirms that it is possible to detect polluted sea or estuary water samples using these chemometric tools and the H^+ affinity spectrum because with polluted samples these multivariate methods lead to incoherent results. So, suspect polluted zones may be monitored in a simple way with a low cost method and spending much less time.

1. Introduction

In 1817 Joseph Louis Gay-Lussac found that the salt content of seawater in the open ocean remained almost constant geographically; this discovery was ratified with the time by John Murray in 1818 [1] and Alexander Marcet in 1819–1822 [2].

The hypothesis derived thereof, known as Marcet's Principle, proposed not only that seawater contained small quantities of all soluble substances, but also that the relative abundances of some of them remained constant. In subsequent studies (i.e., Georg Forchhammer in 1865 [3] and William Dittmar in 1884 [4]), since the results obtained by other researchers found out slight variations in the ratios of the major constituents greater than the experimental error, there were initial disagreements. In 1959, however, Carpenter and Carritt took for granted Marcet's Principle with complete regard for any reported data on probable variations in the composition of the seawater. It is true that the variations found were small and often of the same magnitude of the experimental error in the analytical methods employed, but for many purposes they can be considered negligible [5].

As the ratios between the concentrations of the main ionic constituents in seawater remain reasonably constant, it could be expected that the variations in the water content may be characterized by a single quantity, the salinity. Based on this criterion and according to standard seawater, the practical salinity scale was established at (i) 15 °C, (ii) atmospheric pressure and (iii) an electrical conductivity equal to that of a standard solution of KCl which contains exactly 32.4357 g of KCl per kg of solution. This standard of North Atlantic origin, known as Copenhagen water, has a chlorinity of 19.374‰ and a practical salinity value fixed by definition at 35‰. The experimental composition of the eleven most important components of a 'normal' seawater with salinity of 35‰ is given in Table 1 [6,7].

Speciation models of seawater can be nowadays constructed with the help of chemical equilibrium tools such as computer programs that can not only integrate equilibrium constant databases, but also make complex calculations and plots of the results. For instance, applying the MEDUSA computer program [8,9], it is possible to become aware of the peculiar behavior of the H^+ Affinity Spectrum in pH values close to those common in natural seawater (pH~8). This spectrum, known as ßeta, is defined as the plot of the first derivative of the H_{Bound} ($H_{Bound} = H_{Tot} - [H^+] - [OH^-]$) *vs.* for example, pH [7] and is numerically calculated by the program.

Considering, on the one hand, the great potential of chemometric tools to deal with this kind of 'spectroscopic' information some questions may arise. The first is related to the possibility of using

---

* Corresponding author. Tel.: +34 946012723.
E-mail addresses: kortazar.leire@gmail.com,
leire_kortazar@hotmail.com (L. Kortazar).

0039-9140/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2013.04.067

the H^+ Affinity Spectrum to classify seawater samples. Assuming that Marcet's Principle is exact, this won't be plausible in principle. However, if there would be even a slight spatial variation in the composition of the seawater, a geographical classification of the samples might be carried out. Furthermore, given the constancy in the proportions of the major components of seawater, the potential use of the H^+ Affinity Spectra (obtained at different salinities) to calibrate such composition as well as eventual exceptions to it can be also explored. According to Marcet's Principle, the dilution produced by the river freshwater has to be also proportional for each component if the mixture between seawater and freshwater in the estuary is complete. Due to the fact that the estuarine water samples taken at different times of the tide are more realistic than synthetic ones, these can be used for testing the possibility of calibrating the composition of the main constituents.

Based on these assumptions, the aim of this work is to check if these chemometric tools to may allow the detection of anomalous seawater samples in those cases in which the classification or calibration models give rise to inconsistent results. Such situations may arise due to polluted environments, accidental spills or even extraordinary weather conditions. Thus, these tools could be used as an additional and simple monitoring method for testing seawater quality. When a spill occurs, this fact may be noticed in different ways. Sometimes it is just noticed visually by the effects it produces (i.e., water color changes, presence of death fishes, increase of the eutrophication, etc.). The problem is that in this case, the spill is detected much too late and the effects can be irreparable. On the other hand, regular monitoring of specific compounds can be made, but as it is impractical to control all the possible substances, those responsible of the harmful effects could be undetected. Finally, one the best ways to detect spills or polluted environments is the use of biomarkers, because pollution affects them from the beginning. The problem with this monitoring method is that it usually takes a long time and effort. The proposal presented in this work, is the use of the H^+ affinity spectra in order to detect the pollution, expressed in the form of changes in the basic composition of sea or estuarine waters, in a very fast way before the effects in the biota appear. This can be a very interesting and useful alternative, for example, for the detection of urban wastewater spills because it usually involves a drastic increment in the ammonia/ammonium concentration and, thus, in the buffer capacity.

Table 1
Composition of the 11 main components of seawater in molar scale using the experimental data obtained from references [6,7].

Component	mmol L^{-1}
Na^+	474.20
K^+	10.45
Mg^{+2}	54.33
Ca^{+2}	10.54
Sr^{+2}	0.095
F^-	0.076
Cl^-	548.08
Br^-	0.86
SO_4^{-2}	28.81
Alcalinity$^{*}_{Tot}$	2.39
BO_3^-	0.42

2. Materials and methods

2.1. Sampling

The seawater sampling process was conducted both in summer and winter in the locations indicated in Fig. 1. As it can be seen, this process was conducted into tow main scenarios. On the one hand, in the marine-coastal one, 12 samples were collected along the coast of Biscay (north of Spain) and 2 more picked offshore. On the other hand, in the estuarine sampling along the estuaries of Butroi and Nerbioi–Ibaizabal, 3 sampling stations were selected for each one. The seasonal sampling of the estuaries was performed in summer during 26 h (two full tidal cycles with 9 points) and in winter over 6 h (half of a tidal cycle with 7 points). The samples were taken on the surface of the water column with plastic bottles after having measured their physico-chemical properties (temperature, dissolved O_2, pH, redox potential, specific conductivity, salinity and atmospheric pressure) using a multiparametric probe (YSI 566 multi-parameter probe, YSI Environmental, Yellow Springs, USA). Afterwards, the samples were carried to the laboratory and filtered using 0.45 μm pore size filter disks for further analysis.

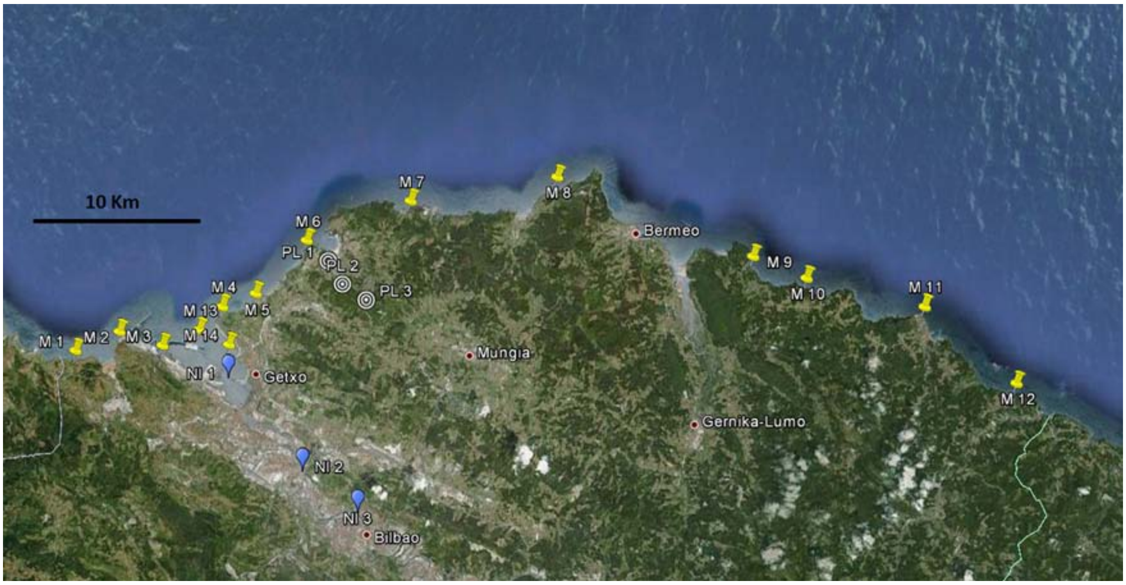

Fig. 1. Sampling point location. Coast samples: M1 (Covaron), M2 (Punta Lucero), M3 (Ciervana), M4 (La Galea), M5 (Sopelana), M6 (Barrica), M7 (Arminza), M8 (San Juan de Gaztelugatxe), M9 (Elantxobe), M10 (Ea), M11 (Lekeitio), M12 (Ondarroa), M13 (Puerto de Bilbao), M14 (Punta Piloto). Sampling points of the estuary of Nerbioi–Ibaizabal: NI 1 (Marina, Getxo), NI 2 (Erandio), NI 3 (Campo Volantín, Bilbao). Sampling points of the estuary of Butroi: PL 1, PL 2 and PL 3, all in Plentzia.

Some samples were collected along the coast in order to explore the applicability of this work in unknown samples. Five of those samples (U1–U5) were taken in unpolluted places while D10 sample was taken near a water treatment plant, being a polluted sample.

2.2. Analytical procedures

The ions analyzed in the seawater samples were calcium, magnesium, chloride, fluoride and sulphate. The total concentration of carbonate, the alkalinity and the H^+ affinity spectrum were also measured. The experimental procedures used for analysis [10,11] were selected based on their performance as well as their availability, ease of use and cost.

Thus, it was decided to make the determination of chloride by conductimetric titration using $AgNO_3$, adding 1 mL of sample in a beaker which was filled with enough water so as to immerse the conductimetric probe CM35 (Crison, Spain). While stirring the solution, the $AgNO_3$ solution was added drop-wise using a 725 Dosimat automatic burette (Metrohm, Switzerland) to form the AgCl precipitate. The two lines obtained, which correspond to the differences in conductivity before and after the equivalence point, were fitted to straight lines by linear regression and the volume of titration equivalence was calculated as the intersection of the these two lines [12].

Fluoride determination was carried out with a standard addition procedure by adding 0.5 mg L^{-1} of an standard solution of 1000 mg L^{-1}, using a fluoride selective electrode, a Ag(s)/AgCl (s) reference electrode (both from Orion, Thermo Fisher Scientific, USA) [13] and a 744 pH Meter (Metrohm, Switzerland) for the potentiometric determination.

Calcium and magnesium concentrations were determined by means of complex formation titrations by using the complexing agent EDTA. Firstly, the sum of the Ca and Mg concentrations was measured using 1 mL of sample, 25 mL of a NH_4^+/NH_3 buffer solution of pH = 10 and Eriochrome Black T (EBT) as indicator. The titrant was a 0.1 M EDTA solution previously standardized [14]. Then, the total Ca concentration was calculated using 1 mL of sample, which was adjusted to pH = 12 with a NaOH pellet and using Murexide as an indicator. The difference between these values provided the concentration of magnesium present in the samples.

The determination of sulphate was performed spectrophotometrically using an in-house modified turbidimetric method based on the formation of the $BaSO_4$ precipitate when a $BaCl_2$ solution is added to the sample. In this method, an external calibration was carried out using sulphate solutions that ranged from 5 mg L^{-1} to 50 mg L^{-1} and blank solutions. The procedure consisted on the addition of different volumes of a stock solution of 1000 mg L^{-1} of Na_2SO_4, 5 mL of a NaCl/HCl mixture, 10 mL of a glycerine:methanol (3:1) mixture and 10 mL of a 15,000 mg L^{-1} of $BaCl_2$ solution before finally bringing the flasks to volume with deionized water. These solutions were shaken during a minute using a vortex mixer. After exactly 5 min from the addition of $BaCl_2$ the spectrum from 200 nm to 450 nm was obtained by a MultiSpec-1501 Diode Array spectrophotometer (Shimadzu Corp., Japan).

The alkalinity and the total carbonate concentration of seawater samples were determined by means of potentiometric acid-base titrations performed with a previously standardized HCl solution. The titrations were done by adding 0.020 mL of 0.1 M HCl increments to 100 mL seawater samples and measuring the electromotive force (EMF) after each addition until a total of 5 mL were added [15]. The determination and subsequent calculations were performed according to Gran's method as stated by Dyrssen and Sillén in 1967 [7,16].

3. Results and discussion

3.1. Buffer capacity spectra

As appreciated in Fig. 2, H_{Bound} spectra have less signal-to-noise ratios than the corresponding buffer capacity (βeta) spectra. To minimize this effect, a Savitzky Golay smoothing performed with a four-point average was conducted in the case of H_{Bound} spectra and a seven-point one for βeta spectra. While in this type of processing some information is lost, a compromise between noise reduction and loss of information must be adopted. Therefore, it was decided

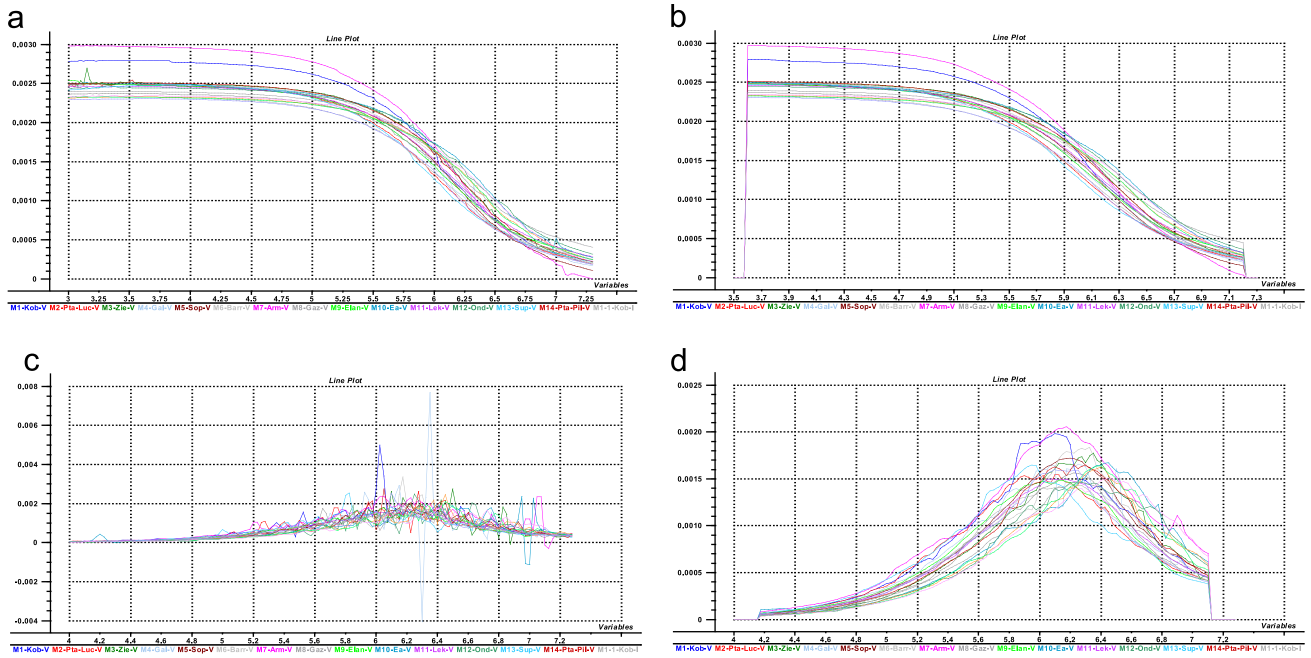

Fig. 2. Plot of the H_{Bound} spectra of the coastal samples (a) without any transformation and (b) with a Savitzky Golay smoothing of four-point average. Plot of the βeta spectra of the coastal samples (c) without any transformation and (d) with a Savitzky Golay smoothing of seven-point average.

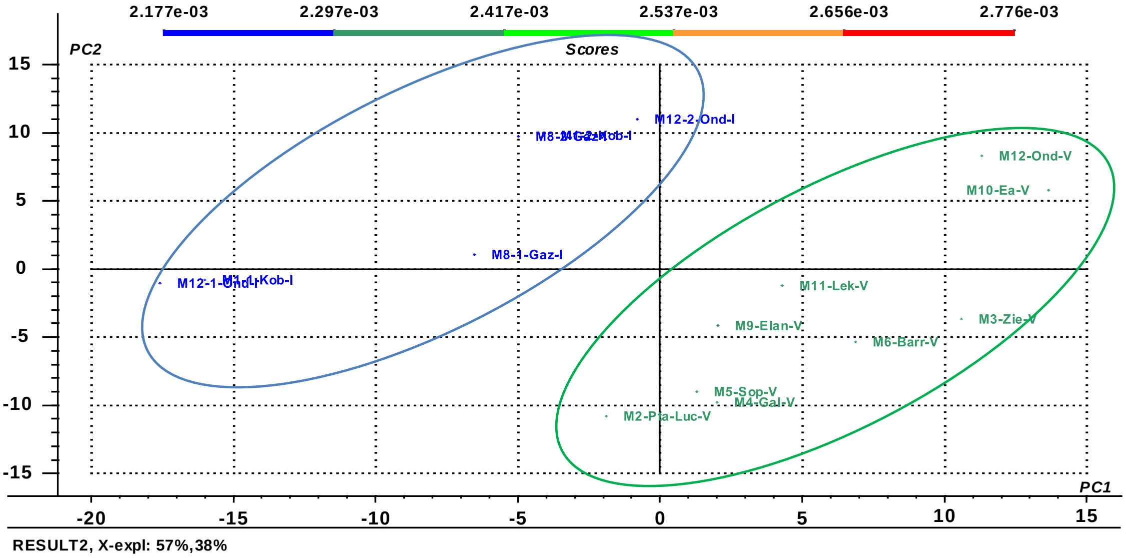

Fig. 3. PCA of the coastal samples of summer (right) and winter (left) using the H_{Bound} spectra.

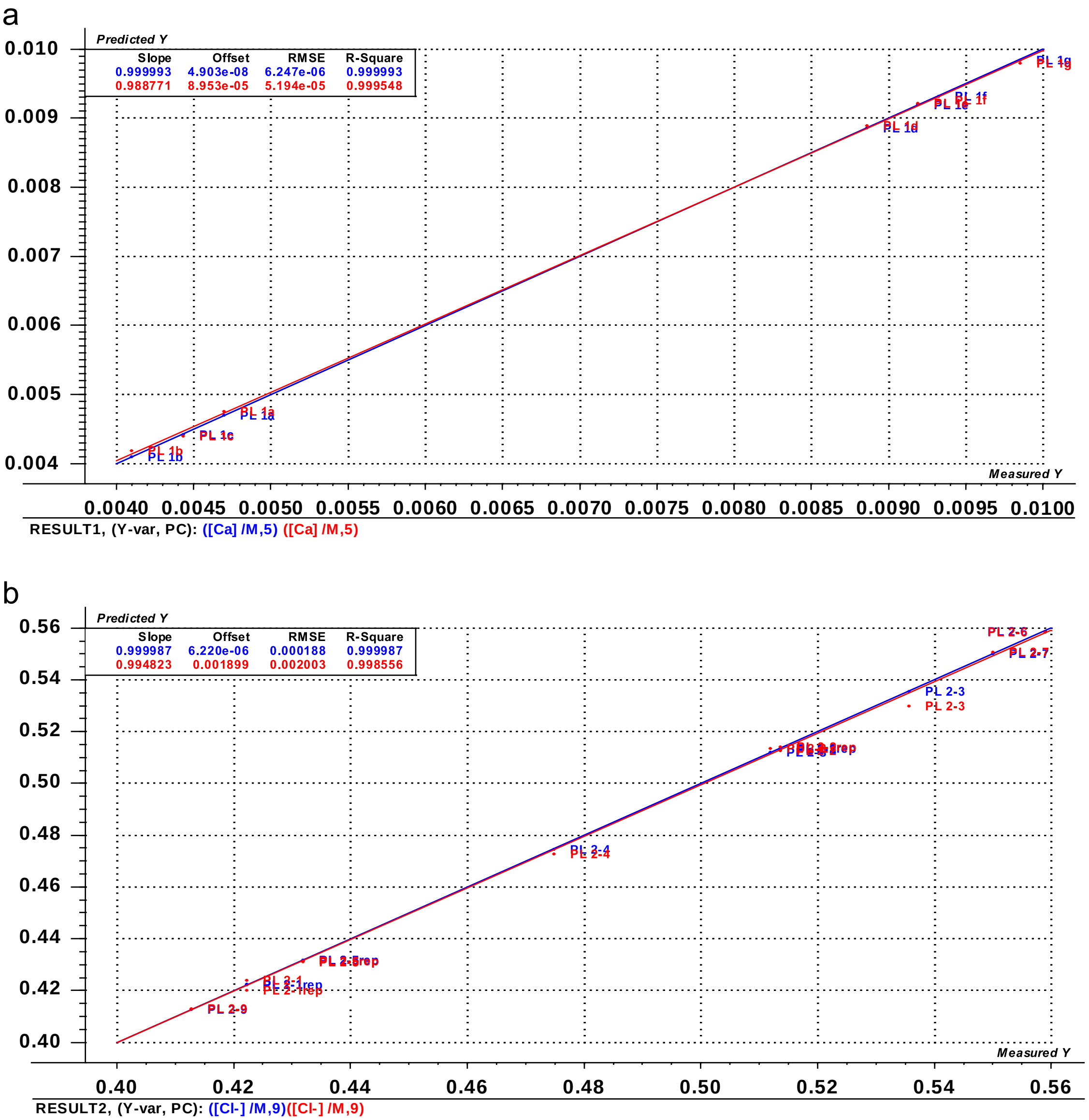

Fig. 4. Results of the PLS1 regression of the buffer capacity spectrum *vs.* (a) calcium concentration in the winter sampling and (b) chloride concentration in the summer sampling, for one of the positions located in the Butroi's estuary where blue line indicates the calibration curve while the red line indicates the validation curve. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

to use the H_{Bound} spectra as X variables fixed for subsequent chemometric analysis. All the spectra plots in Fig. 2 as well as the other chemometric calculations described in this work were made using The Unscrambler® (CAMO, Oslo, Norway) computer program.

3.2. Principal component analysis (PCA)

A PCA of the coastal samples was done using the H_{Bound} spectra in order to see if it was possible to observe any cluster formation which could help to classify the samples somehow.

The results in Fig. 3 show that it is not possible to classify the samples in space, or at least not in the relatively short distances (spatial distribution) in which the samples were picked. However, such was not the case of the seasonal differences since the samples taken were grouped into two clusters, one for the summer samples and another for the winter ones. It can also be appreciated that the cluster formation, and therefore the seasonal classification of the samples, is related to the first principal component (PC1).

3.3. Regression analysis

Partial Least Squares (PLS1) regression models were constructed using the estuarine water samples (of both estuaries) to see if it was possible to conduct a calibration via PLS1 regression of the H_{Bound} spectra *vs.* the concentration of any of the major constituents of the seawater. Fig. 4 summarizes the result of such regression for the cases of calcium and chloride.

This figure exhibits the potentiality of performing a calibration of the composition of the seawater using estuarine waters at different salinities. Similar results were obtained for the other measured components.

However, there are some situations in which it is not possible make this calibration: i.e., when the seawater and freshwater are not completely mixed or when the tide influence is too small (see Fig. 5).

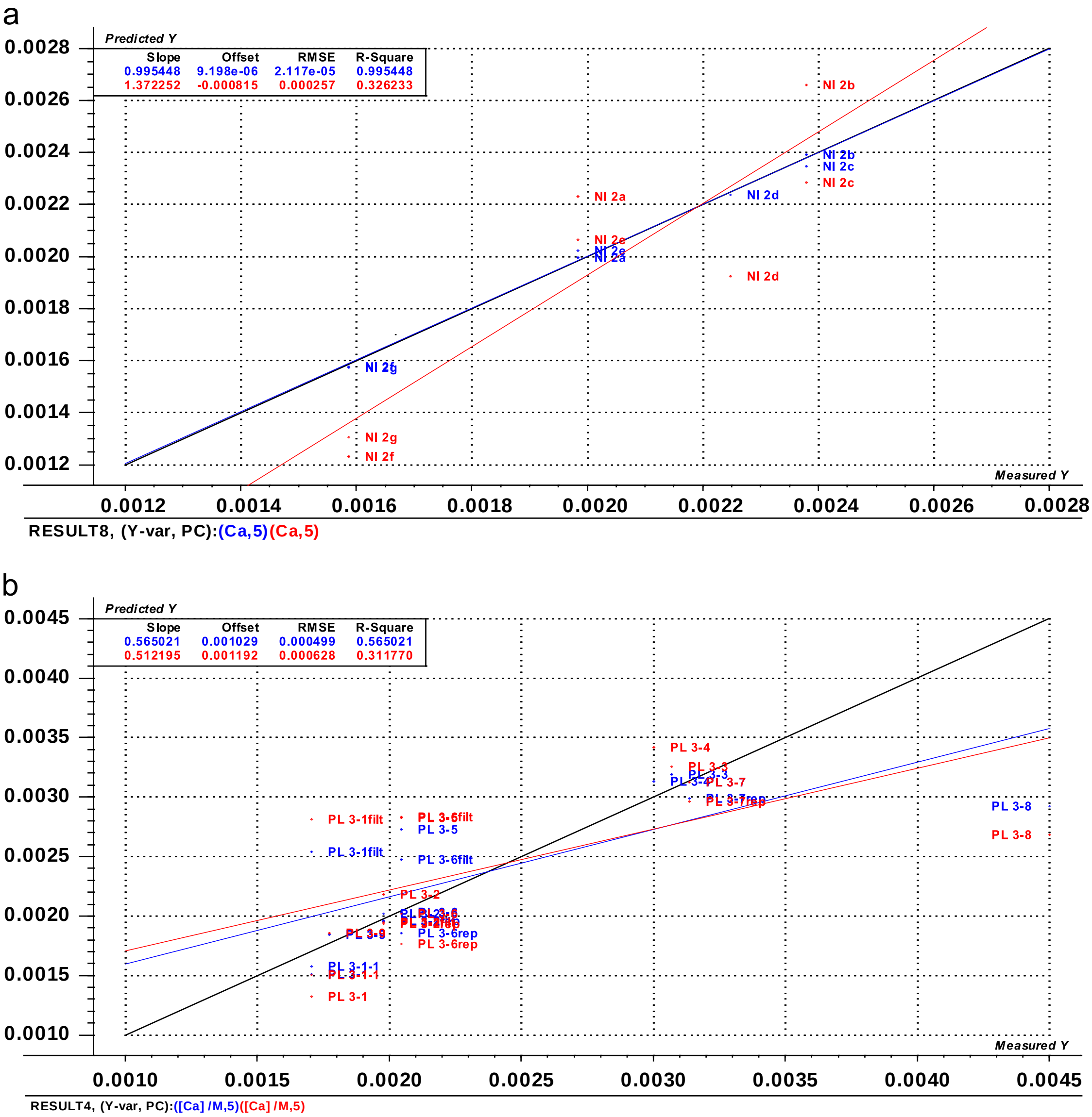

Fig. 5. Result of the PLS1 regression of the buffer capacity spectrum *vs.* calcium concentration (a) for the winter samples of the second position of the Nerbioi–Ibaizabal estuary, where the seawater and freshwater are not completely mixed and (b) for the summer samples collected in the third position of the Butroi estuary, where the influence of the tide is not big enough. Blue line indicates the calibration curve while the red line indicates the validation curve. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

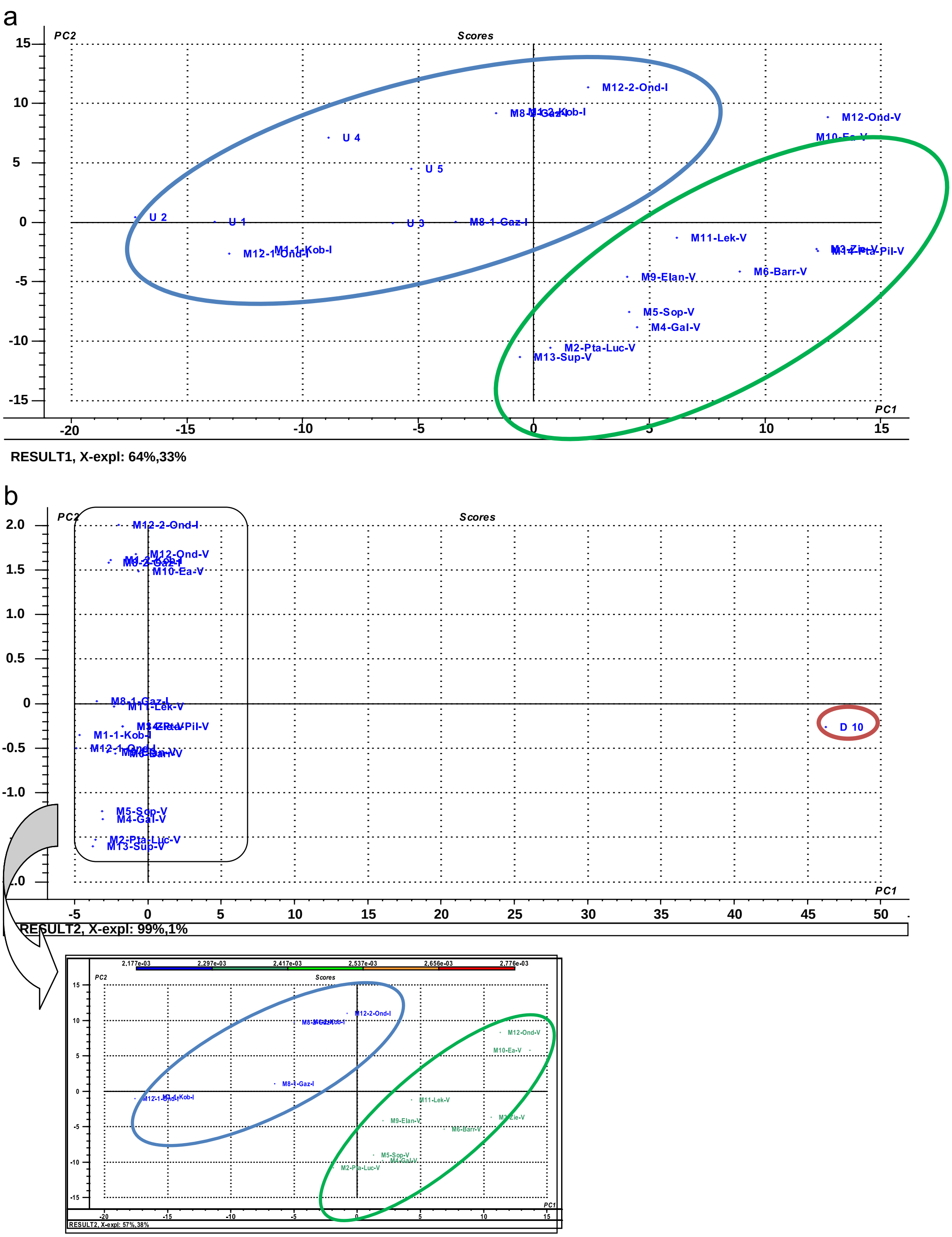

Fig. 6. PCA of the coastal samples of summer and winter using the H_{Bound} spectra with (a) some supposedly nonpolluted unknown seawater samples (U1–U5) and (b) a polluted seawater sample (D10) and a zoom in the zone of the coastal samples.

In the first case, the samples were taken after some very cold days when ice and snow began to melt in the mountains that provide water to the Nerbioi–Ibaizabal system. The temperature difference between seawater and freshwater was too high to achieve a complete water mixture. For that reason it is not possible to calibrate the seawater using solutions of estuarine water dilutions when this phenomenon takes place.

In the second case, the influence of the tide in this position of the estuary is very small, as it is too far from the river mouth. It is practically a river freshwater where it is not possible to perform the calibration because Marcet's principle does not hold any longer.

3.4. Pollution detection

As it was stated before, the aim of this work is the application of these chemometric tools, PCA and PLS, in order to detect polluted water samples. For that, some unknown coastal seawater samples

Table 2
The predicted values using both PL and NI regression models, the RSD values, the determined values by laboratory procedures and the errors of chloride concentration (in mol L^{-1}) and fluoride concentration (in ppm).

	Predicted chloride	RSD (%)	Determined chloride	Error (%)
Values calculated with PL model				
U 1	0.55	1.0	0.509	8
U 2	0.667	0.9	0.53	26
U 3	0.491	1	0.46	7
U 4	0.578	0.8	0.542	7
U 5	0.564	0.6	0.54	4
D 10	−3.727	3	< LOD	–
	Predicted fluoride	RSD (%)	Determined fluoride	Error (%)
U 1	1.446	0.5	1.396	4
U 2	1.715	0.4	1.592	8
U 3	1.24	0.5	1.286	4
U 4	1.626	0.3	1.516	7
U 5	1.587	0.2	2.439	35
D 10	−8.863	1	0.402	2305
Values calculated with NI model				
U 1	0.535	1	0.509	5
U 2	0.539	2	0.53	2
U 3	0.525	1	0.46	14
U 4	0.577	2	0.542	6
U 5	0.549	1	0.54	2
D 10	−0.434	37	< LOD	–
	Predicted Fluoride	RSD (%)	Determined Fluoride	Error (%)
U 1	1.262	0.01	1.396	10
U 2	1.498	0.01	1.592	6
U 3	1.027	0.004	1.286	20
U 4	1.491	0.01	1.516	2
U 5	1.463	0.01	2.439	40
D 10	−4.364	0.03	0.402	2305

(U1–U5 and D10), among which there was a polluted sample (D10) taken near a water treatment plant, were analyzed. The results of this sample were compared with other supposedly nonpolluted sample's results.

By applying the PCA it was found that the samples were correctly differentiated into summer and winter, except for the samples taken in places where the water was polluted, as shown in Fig. 6.

When the PCA was made including the polluted sample, it clearly appeared as an outlier.

A calibration model for Butroi estuary and another for Nerbioi–Ibaizabal estuary were used to try to predict the concentrations of the major constituents in unknown coastal samples. The regression models from both estuaries were used to check if similar results were obtained using water samples of different estuaries, that is, in order to explore the idea that it is possible to predict unknown samples using water from any estuary. In order to see the accuracy of the prediction, the chemical analysis of all the constituents was first made.

Thus, Table 2 shows the predicted values using both regression models, the Relative Standard Deviation (RSD) values, the determined values by laboratory procedures and the errors of alkalinity, chloride concentration and fluoride concentration.

In this case, the predicted values were rather close to those obtained after the chemical analysis of these constituents in the supposedly nonpolluted unknown coastal seawater samples (U1–U5) independently of which regression model was used. In these samples, the error comparing the predicted and the measured values is small in almost all cases. In the D10 polluted seawater sample, the concentration values obtained after the prediction based on calibration models were incoherent because in some cases the predicted value is even negative. Furthermore, the difference between the predicted and the measured values is generally enormous with both regression models, although the differences using NI estuary samples are bigger. That might be because the Nerbioi–Ibaizabal estuary is not quite clean and it is better to do the regression using a nonpolluted estuary water model.

Therefore, considering these results, some conclusions are reached: using chemometric tools such as PCA with the H_{Bound} spectra, it is possible to classify seawater samples according to the season in which they were taken, or at least if the samples were collected in summer or in winter.

Moreover, it is possible to make a natural calibration, using the Partial Linear Squares (PLS1) regression method, of the components of seawater *vs.* the H_{Bound} spectra using estuarine waters of different salinities collected at different times of the tide. Unfortunately there are situations where it is not possible to make a calibration: i.e., when the seawater and freshwater are not thoroughly mixed or when the tidal influence is too small.

Thus, assuming the constancy of the composition of seawater, chemometric tools were used to investigate their potential in detecting polluted waters. On the one hand, it has been possible to detect polluted samples using these chemometric tools and the H_{Bound} spectrum, which is very simple and inexpensive to obtain, as it is the case of sample D10. The generality of this conclusion was probed using the model obtained with the cleanest available water samples. On the other hand, it has been demonstrated that the unknown samples U1–U5 are unpolluted samples. Thanks to that, a very simple and fast monitoring method for sea or estuarine water pollution can be set up with a substantial reduction in its cost and time.

References

[1] J. Murray, Trans. R. Soc. Edinburgh 8 (1818) 205–244.
[2] A. Marcet, Philos. Trans. R. Soc. London 109 (1819) 161–208.
[3] G. Forchhammer, Philos. Trans. R. Soc. London 155 (1865) 203–262.
[4] W. Dittmar, Report on researches into the composition of ocean–water, collected by H.M.S. challenger, during the years 1873–1876, in: C.W. Thomson, J. Murray (Eds.), Physics and Chemistry, vol. 1 (1884).
[5] D.E. Carritt, J.H. Carpenter, The composition of sea water and the salinity–chlorinity–density problems, in: Proc. of the Conference on Physical and Chemical Properties of Sea Water (p. 67), Easton, Md. (USA), 1958.
[6] F. Culkin, Chemical oceanography, The Major Constituents of Sea Water, Academic Press, London, New York121–161.
[7] D. Dyrssen, L.G. Sillen, Alkalinity and Total Carbonate in Sea Water—A Plea for p-t-Independent Data, Tellus XIX, vol. 19, 1967, pp. 113–121.
[8] I. Puigdomenech, Windows software for the graphical presentation of chemical speciation, in: 219th ACS National Meeting, Abstracts of Papers, vol. 1, pages Abstract I&EC–248, San Francisco, CA, American Chemical Society, 26–30 March 2000.
[9] I. Puigdomenech, Input, Sed and Predom: Computer Programs Drawing Equilibrium Diagrams, Technical Report TRITA-OOK-3010 (ISSN 0348-825x), Royal Institute of Technology (KTH), Department of Inorganic Chemistry, SE-100 44 Stockholm, (Sweden) 1983.
[10] L.S. Clesceri, M.A.H. Franson, American Water Works Association, and Water Environment Federation. Standard Methods for the Examination of Water and Wastewater, American Public Health Association, 1985.
[11] K. Grasshoff, K. Kremlingl, M. Ehrhardt, Methods for Seawater Analysis, John Wiley-VCH, 1999.
[12] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, Pearson Education Limited, 2010.
[13] T. Anfalt, D. Jagner, Anal. Chim. Acta 53 (1971) 13–22.
[14] A.J. Vogel, Textbook of Quantitative Inorganic Analysis, 4th edition, Longmans, 1964.
[15] R. Cazallas, L.A. Fernandez, N. Etxebarria, J.M. Madariaga, Lab. Robot. Autom. 5 (1993) 161–169.
[16] G. Gran, Analyst 77 (1952) 661–671.