This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.
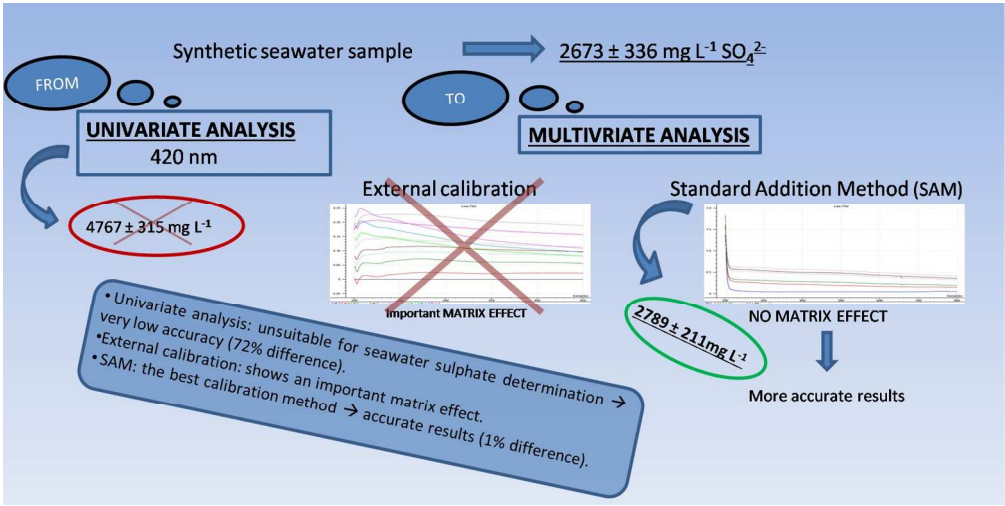
500x250mm (96 x 96 DPI)

TECHNICAL NOTE

Application of multivariate analysis to the turbidimetric determination of sulphate in seawater

Leire Kortazar *[a], Janire Sáez [a], Josu Agirre [a], Jon Kepa Izaguirre [a], Luis Angel Fernández [a]

Abstract. Spectroscopic techniques are widely used in the field of analytical chemistry for the determination of a huge number of analytes. The use of diode array spectrophotometers involves the possibility of collecting the whole UV-Vis spectrum, which provides an opportunity to make a multivariate analysis of the samples. The effectiveness of the use of multivariate data analysis against the univariate counterpart (which is the most common for classic spectroscopic methods) is demonstrated through application to the sulphate determination in seawater samples using a modified turbidimetric method. The original method recommends to perform the analysis at 420 nm this being an univariate analysis. The modification presented uses multivariate methods for the calibration with the whole UV-Vis spectrum from 200 to 800 nm instead of the single measurement at 420 nm. The external calibration shows that there is a matrix effect which can compromise the accuracy of the measurements, so that an in-house Iterative Multivariate Standard Addition Method (IMSAM) was successfully tested in order to avoid this effect. As the spectrum interval was too wide, both a Genetic Algorithm (GA) and Interval Partial Least Square (IPLS) regression methods were used as well in order to select the best wavelengths for the regression and back prediction of the samples. The results obtained in this work show that the performance of the method at 420 nm, as the original method recommends, leads to a considerable error and also that the best segment of the spectra for the determination of sulphate concentration in seawater is from 600 to 800 nm.

Analytical Methods Accepted Manuscript

[a] Department of Analytical Chemistry, University of the Basque Country, P.O. Box 644, 48080 Bilbao, Spain
  E-mail: kortazar.leire@gmail.com; Tel.: +34 94 6012723

Journal Name

TECHNICAL NOTE

Introduction

Ultraviolet and Visible (UV-Vis) spectroscopy is routinely used in analytical chemistry for the quantitative determination of different analytes [1, 2].

The simplest method of obtaining concentration information from UV-Vis spectroscopic data is through the use of an univariate regression technique such linear least squares, in direct application of Beer's law [3].

However, modern instrumentation provides the opportunity to collect many variables (for example, a diode array spectrophotometer involves the possibility of collecting the whole UV-Vis spectrum). All these data can be used to make a multivariate calibration using different chemometric tools which are based on the application of certain mathematical and statistical methods to chemical problems to extract useful information. There are several types of multivariate regression techniques [4] but in the presented work, the effectiveness of the use of multivariate regression instead of the univariate one will be checked through the application to the sulphate determination in seawater samples using PLS 1 regression.

Sulphate is found almost universally in natural waters at concentrations ranging from a few tenths to several thousand milligrams per liter. The sulphate concentration in seawater is about 2700 mg kg^{-1}, being the second major anion in seawater [6, 7] after chloride. Sulphate concentration provides a direct mean of evaluating the oxidation state of seawater. It is also a proxy for atmospheric oxidation state as sulphate accumulation in the ocean is a result of oxidative weathering of land sulphides [8]. Recently, chemometric tools have been used in sulphate analysis for the optimization of a bioluminescent method using experimental design techniques [9]. Therefore, it is interesting to perform the determination in a most accurate way as possible.

For the analysis of sulphate a turbidimetric method based on the precipitation of barium sulphate was chosen [10, 11] although some variations from the original were introduced. This method was chosen because among all the methods for the sulphate determination in seawater it is one of the fastest methods available and also gives the possibility of measuring a large number of samples in a short time. In this method the absorbance is measured only at 420 nm [11], being this, thus, an univariate method of analysis. However, it could be interesting to perform the analysis measuring the full UV-Vis spectra and transform the method to a multivariate one in order to check if its characteristics (ie., sensitivity, precision, etc.) improve. The performance of this modified method should be tested against proven standard methods. For this, both a classic volumetric method [9], due to its rapid and accurate performance and the use of synthetic samples [7] of known concentration can be used.

Environmental samples are usually subjected to a lot of interferences and matrix effects are also very usual in this kind of samples because of their inherent complexity. The sample's matrix may influence the analytical signal. If the matrices of the calibration standards and those of the samples differ, the calibration may be inaccurate [12].

Standard addition is considered to be the most suitable method for compensating matrix effects [13, 14]. There is not much information in the literature about the use of the traditional SAM for multivariate data analysis. Recently, Melucci et al. [14] have used a Multivariate Standard Addition Method (MSAM) for the improvement of the analytical performance in the differential pulse stripping voltammetry analysis of drugs. Concerning the use of multivariate data analysis methods, a final step in order to reduce model complexity is to resort to variable selection methods (ie, wavelengths in spectrophotometry). Besides, not all the collected data are always useful and sometimes they could even interfere the information extraction. Among the different strategies that can be used for this purpose, this work will concentrate on the usefulness of two of them as they are readily available in commercial software packages: (1) Genetic Algorithms (GA), and (2) Interval Partial Least Square (iPLS) regression, both available in the PLS Toolbox (Eigenvector Research, Wenatchee (WA), USA) developed for use with the Matlab (Mathworks, Natick (MA), USA) computer program. Unlike the GA, iPLS does a sequential, exhaustive search for the best variable or combination of variables. Furthermore, it can be operated in "forward" mode (F), where intervals are successively included in the analysis, or in "reverse" mode (R), where intervals are successively removed from the analysis.

Experimental methods

Sampling

Samples were collected on the surface water of the marine of Getxo in the Nerbioi-Ibaizabal estuary (43°20'24.40"N, 3°1'15.84"W) (used only for external calibration) and in the Azkorri (L1) (43°22'54.05"N, 3°0'53.16"W) and Arrigunaga (L2) (43°21'20.77"N, 3°1'15.51"W) beaches. These places are located in the Bizkaia coast (North of Spain).

The samples were taken on the water surface and using 1 L plastic bottles. After the sampling, the samples were carried to the laboratory and they were filtered using 0.45 μm pore size cellulose nitrate membrane filters (Whatman, Germany). The samples were kept at ≤ 6ºC in the fridge until the analysis.

Volumetric sulphate determination

The volumetric determination of sulphate was performed based on the method described by the Wood Hole Oceanographic Institution and by the U.S. National Science Foundation [10].

Turbidimetric sulphate determination

The turbidimetric determination of sulphate was performed according to the method described by the United States Environmental Protection Agency (US EPA) [11] with the

following modifications: (1) the use of a 15000 mg L^{-1} BaCl$_2$ solution instead of in solid form and (2) the acquisition of the UV-Vis spectrum. Sulphate ion is converted to a barium sulphate suspension under controlled conditions. The resulting turbidity is determined by a spectrophotometer and compared to a curve prepared from a standard sulphate solution. Both, external calibration and standard addition methods were tested.

- The external calibration was carried out using sulphate standard solutions ranging from 5 mg L^{-1} to 30 mg L^{-1}. Blank solutions were also used for the determination. The spectrum from 200 to 450 nm was obtained using a MultiSpec-1501 Diode Array spectrophotometer (Shimadzu Corp., Kyoto, Japan). The same procedure was carried out for measuring the samples using 270 μL of the seawater samples instead of the sulphate stock solution.

- The standard addition method was also performed using the same procedure. In this case, 833 μL of the seawater sample were added into all the flasks. Then, additions of the sulphate standard solution were made ranging from 15 mg L^{-1} to 75 mg L^{-1}. The spectra were measured from 200 to 800 nm in this case.

Data analysis

As stated above, at first, an external multivariate calibration was tried using The Unscrambler ® (CAMO, Oslo, Norway) computer program. Different Partial Least Squares (PLS 1) regression models were obtained using the calibration samples and once the best model was chosen, the prediction of the unknown samples was made.

After that, different SAMs were tested. To begin, the method developed by Melucci et al. **[15]** was used. In analogy with the traditional univariate SAM, this method is based on a standard addition calibration followed by the prediction of the zero signal sample (blank).

On the other hand, to overcome the limitations found in the application of the previous MSAM method, an in-house IMSAM was tested. Its steps are the following:

Step 0: First, a PLS 1 regression is made using both the samples with additions (considering the added concentrations as the samples concentration values) and the blank sample. The offset of the blank sample of the validation curve is considered as the initial approximation to the sulphate concentration (see Figure 1) **[16]**. This is done so because, in contrast with univariate SAM methods in which the unknown sample concentration is read from the negative side of the concentration axis of the Signal vs. Added Concentration plot, in the multivariate case we just have the Predicted Y vs. Measured Y plots, where Y stands for concentration (Figure 1), and the only possible way of reading the predicted blank solution concentration is on the offset of the Predicted Y axis validation line.

Step 1: This initial concentration value is used as the new blank concentration and is also added to the other calibration solutions concentrations. With these new values, a PLS 1 regression is made and with the obtained model the sulphate concentration of the sample without any addition is predicted.
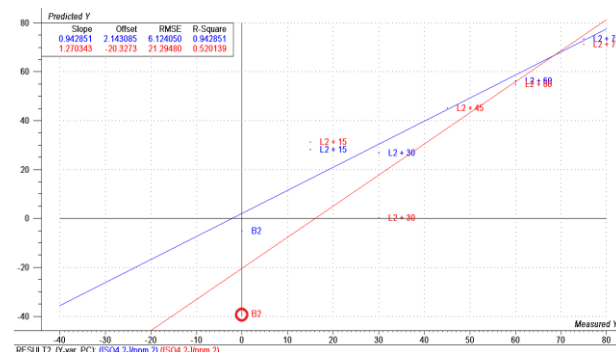

Fig. 1 Plot of the both calibration (blue) and validation (red) curves with a mark in the offset of the validation value for the blank which has to be taken into account as the initial approximation of the sulphate concentration in the iterative method proposed.

Steps 2 to n: This new concentration is used as the new blank concentration and a new PLS 1 and the corresponding prediction are made. The procedure is repeated until the deviation of the predicted concentration is the lowest possible.

Thus, the overall idea of this iterative method is, so to say, to successively transform the MSAM method into a 'normal' multivariate calibration in which both the blank and the solutions with additions are iteratively transformed into a 'regular' set of calibration solutions by modifying their concentrations until the best possible model for the whole set is obtained.

Results and discussion

As stated before, the volumetric method was used as the reference method for the validation of this modified turbidimetric method and its good performance was confirmed using synthetic seawater prepared according to the concentrations of the major constituents of the standard seawater known as "Copenhagen" seawater **[6, 7]**. The sulphate concentrations, in mg L^{-1}, of the synthetic and the real samples are 2673 ± 336 with a Relative Standard Deviation (RSD) of 6%, 1217 ± 358 with a RSD of 14% and 2658 ± 324 with a RSD of 6%, respectively. The uncertainty was considered as the square root of the sum of the variance for repetition and the intrinsic variance. The last one was determined using the GUM (Guide to the expression of Uncertainty in Measurement) method **[17]**.

The real sulphate concentration of the synthetic sample was 2770 mg L^{-1}, and the error of the sulphate determination with respect to the volumetric method is 3%. So, the trueness of the method was considered appropriate in order to use it as the reference method. The precision of the method is acceptable too.

After the measurement of all the calibration and estuary water samples with the turbidimetric method, the data analysis was made using The Unscrambler® computer program. Given the simplicity of the shape and the characteristics of the spectra, there was no need of any kind of spectral pre-treatment. Only some spectral operations such as the substractions indicated in Figure 2 and the averaging of the repeated measurements of real samples were made.

Figure 2 shows the plot of the spectra of the calibration samples and the estuary water samples.
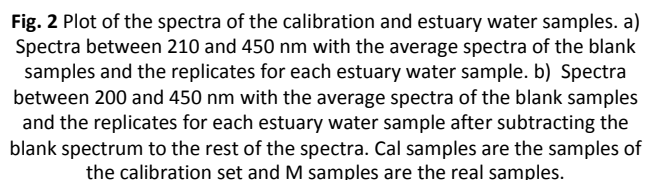
Fig. 2 Plot of the spectra of the calibration and estuary water samples. a) Spectra between 210 and 450 nm with the average spectra of the blank samples and the replicates for each estuary water sample. b) Spectra between 200 and 450 nm with the average spectra of the blank samples and the replicates for each estuary water sample after subtracting the blank spectrum to the rest of the spectra. Cal samples are the samples of the calibration set and M samples are the real samples.

As it can be seen, the spectra of the calibration samples are quite different compared with the spectra of the real estuary water samples. These spectra cross the calibration spectra. So, it seems that there is an important matrix effect which can affect the accuracy of the measurements. In order to try to avoid this effect, a standard addition procedure was tried.

Moreover, as it seems that the final parts of these spectra, those closer to 450 nm, are more constant, for the standard addition method the measurement of the spectrum was extended up to 800 nm. Figure 3 shows the spectra from 200 to 800 nm of both the artificial and the real seawater sample L1 using the standard addition method.

As it is possible to appreciate, in this case the spectra have the same shape, so the matrix effect was minimized to a great extent. However, at the beginning of the spectra the separation between them is bigger than at the end. For that reason, the sulphate concentration was determined using different segments of the spectra in order to find which segment gives the closest concentrations compared with the real value (in the case of the synthetic water sample) and with the reference method values.

As stated above, the multivariate standard addition method developed by Melucci et al. **[15]** was used. The full spectra and different segments of it were analyzed in order to see if there were differences in the results depending on the used wavelengths.

Using this data analysis method the precision is not acceptable: the uncertainty of the measurements is too high in most cases, with RSD values ranging from 8% to 73% for all the samples.

Moreover, the differences between the values obtained with both methods and the real value in the synthetic sample are also too high, having the lowest one a 39% error. Therefore, this method for the data analysis is considered unacceptable for the sulphate

determination in seawater. As there are important differences depending on the segment of the spectra which was analyzed, it may be interesting to delve into the variable selection in order to find the most accurate results as possible.
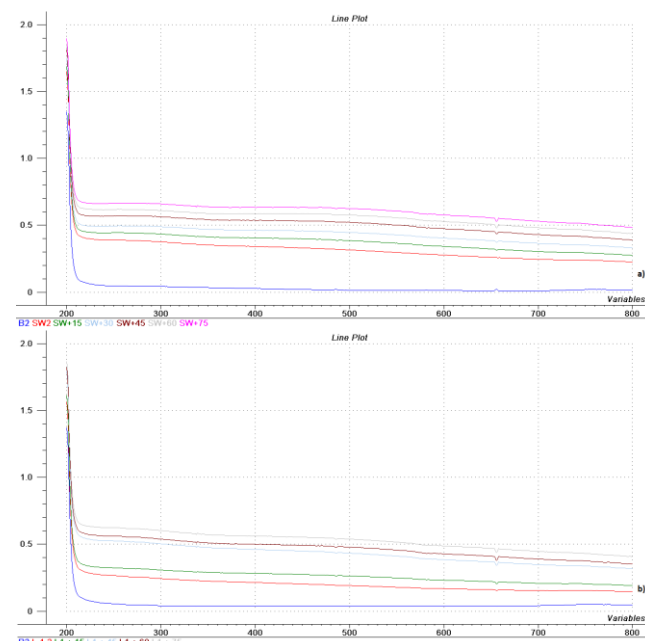

Fig. 3 Plot of the complete spectra of the standard addition procedure of a) synthetic seawater sample and b) L1 real sample.

Variable selection

There are several ways to select which variables are the most important for the sulphate determination.

As the method developed by Melucci et al. **[15]** doe not work as expected, an in-house IMSAM was tried. In this case, several segments of the spectra were analyzed: 250-350 nm, 350-450 nm, 450-550 nm, 550-650 nm, 650-750 nm, 600-700 nm, 700-800 nm and 600-800 nm, in addition to the full spectra and the single 420 nm absorbance value which is the one used in the original US EPA method.

The best segments for the sulphate determination lay around 600 to 800 nm and also considering the full spectrum. The truest and most precise segments seem to be, on the one hand, 650-750 nm with a value of 2740 ± 226 mg L^{-1} and a difference of 1% with the real value. On the other hand, the 600 to 800 nm segment is also one of the most precise with a value of 2789 ± 211 mg L^{-1} and a difference of 1% with the real value (these data are shown in Table 1 in the ESI†).

However, the matrix of this synthetic seawater is not the same as that of real seawater. For that reason, real samples were tested using this method (as shown in Table 2 in the ESI†) in order to see if the results match between them.

In the L1 sample, there are different segments giving true results. The truest is the 600-800 nm with a 4% error, although the 350-450 nm and the 550-650 nm segments also give a low error. On the other hand, the RSD is quite high ($\geq 12\%$) which means that the precision is not very good.

However, in the L2 sample the sulphate concentration is closest

to the reference value at the end of the spectra, being the 700-800 nm segment the truest one with an error of 12%, which can be considered appropriate. In this case, the RSD of all the samples is very low (from 1% to 5%).

5 Considering both samples, the results given by the last segments of the spectra are generally the closest to the reference value.

The analysis was also made at 420 nm alone, as indicated the original US EPA method [11]. The error using this wavelength is 78% for the synthetic and the L1 samples and 60% for the L2 10 sample.

Another way to carry out the variable selection is the use of Genetic Algorithms (GA) and Interval Partial Least Square (iPLS) methods. With GA the errors obtained for the synthetic sample are 10% with the reference methods' value and 7% with 15 the real value. In the case of the L1 and L2 real samples the errors were 25% and 40%, respectively, considering the results given by the volumetric method. So, this variable selection technique seems not to be adequate enough for the sulphate determination in seawater. On the other hand, the Interval Partial Least Square 20 (iPLS) was used obtaining errors of 4% for L1 and 12% for L2. These errors are the same obtained by the in-house procedure. The problem is that the best segments obtained for each sample do not agree, so, this method cannot be considered acceptable because it is not possible to reach any solid conclusion.

25 Summarizing, the best data analysis method for MSAM in this case, has been proved to be the in-house IMSAM using the 600-800 nm segment.

Conclusions

30 The plot of the spectra using the original external calibration procedure shows that there is an important matrix effect which can lead to incorrect results.

To avoid this matrix effect a standard addition procedure was performed successfully using a sulphate standard solution. Taking 35 into account all the obtained information, both with the synthetic and the real samples, some conclusions can be drawn. On the one hand, the in-house IMSAM is truest at the end of the spectra, around the 600 to 800 nm region. The error in these final segments is generally lower than in the rest of the spectra. On the 40 other hand, the analysis at 420 nm, indicated by the US EPA method, gives very high errors for all the samples, so it was concluded that this wavelength is not appropriate for the seawater sulphate analysis, in contrast with the recommendations of the

EPA, because it leads to a large error. So, multivariate data 45 analysis has been shown to highly improve the accuracy of this determination.

Finally, it can be also concluded that the multivariate standard addition methods available in literature are not suitable for the turbidimetric sulphate determination in seawater and that the 50 strategies for the variable selection are not useful either despite or maybe because of the featureless characteristics of the involved spectra. Thus, the best way to determinate sulphate concentration in seawater is to use the in-house IMSAM developed and the 600-800 nm segment.

55 **References**

[1] D. A. Skoog, D. M. West, F. J. Holler, S. R. Crouch, Fundamentals of Analytical Chemistry, 9th ed., Books/Cole, Cengage Learning, Belmont, 2013.

[2] American Public Health Association, American Water Works 60 Association, Water Environment Federation. Standard Methods for the Examination of Water and Wastewater, 22nd ed., American Public Health Association, 2012.

[3] R. J. Hendershot, R. Vijay, B. J. Feist, C. M. Snively, J. Lauterbach, Meas. Sci. Technol., 16 (2005) 302–308.

65 [4] J. N. Miller, J. C. Miller, Statistics and Chemometrics for Analytical Chemistry, 5th ed., Prentice Hall, Harlow, 2000.

[5] K.H. Esbensen, Multivariate Data Analysis - In practice, Camo Software AS, 5th edition, 2001.

[6] F. Culkin, The major constituents of sea water, in: J. P. Riley, G. 70 Skirrow, Chemical Oceanography, Academic Press, London and New York, 1965, 121–161.

[7] D. Dyrssen, L. G. Sillen, Tellus 19 (1967) 113–121.

[8] J. Foriel, P. Philippot, P. Rey, A. Somogyi, D. Banks, B. Menez, Earth Planet. Sci. Lett., 228 (2004) 451– 463.

75 [9] s. M. Marques, J. C. G. Esteves da Silva, Anal. Methods, 5 (2013) 1317-1327.

[10] Woods Hole Oceanographic Institution, Limnol. Oceanogr. 23 (1978) 1069-1072.

[11] Environmental Monitoring and Support Laboratory of the 80 Environmental Research Laboratory Methods for chemical analysis of water and wastes, EPA-600/4-79-020, U. S. Environmental Protection Agency, Cincinnati, 1983.

[12] K. Grasshoff, K. Kremlingl, M. Ehrhardt, Methods for Seawater Analysis, third ed., Wiley-VCH, Weinheim, 1999.

85 [13] M. Stüber, T. Reemtsma. Anal. Bioanal. Chem. 378 (2004) 910–916.

[14] R. G. Brereton, Analyst 112 (1987) 1635-1657.

[15] D. Melucci, C. Locatelli, J. Electroanal. Chem. 675 (2012) 25-31.

[16] Camo Software AS, The Unscrambler: Methods, 2006.

[17] Joint Committee for Guides in Metrology. Evaluation of 90 measurement data - Guide to the expression of uncertainty in measurement. JCGM 100:2008, GUM 1995 with minor corrections. Joint Committee for Guides in Metrology, 2008.