UNIVERSITY OF COPENHAGEN

Comparation of bi and tridimensional multivariate methods in environmental monitoring

Gredilla, A.; Amigo Rubio, Jose Manuel; Fdez-Ortiz de Vallejuelo, S.; Bro, Rasmus; Diego, A. de; Madariaga, J.M.

Publication date:
2010

Document version
Early version, also known as pre-print

Citation for published version (APA):
Gredilla, A., Amigo Rubio, J. M., Fdez-Ortiz de Vallejuelo, S., Bro, R., Diego, A. D., & Madariaga, J. M. (2010). *Comparation of bi and tridimensional multivariate methods in environmental monitoring*. Paper presented at 7º Colloquium Chemiometricum Mediterraneun (CCM VII 2010 - Granada), Granada, Spain.

COMPARATION OF BI AND TRIDIMENSIONAL MULTIVARIATE METHODS IN ENVIRONMENTAL MONITORING

A. Gredilla[a]*, J.M. Amigo[b], S. Fdez-Ortiz de Vallejuelo[a], R. Bro[b], A. de Diego[a], J. M. Madariaga[a]

[a]Department of Analytical Chemistry, University of the Basque Country, P. O. Box 644; 48080 Bilbao, Basque Country.
[b]Department of Food Science, Quality and Technology, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
*e-mail: ainara.gredi@ehu.es

Abstract

Normally the use of univariate chemometric methods is not enough to obtain significant and relevant information about the studied area, thus, the use of multidisciplinary methods could be considered a powerful tool to discover contamination sources in the environmental studies. Therefore, the use of multivariate analysis may be considered as a suitable alternative for the correct understanding and interpretation of environmental dataset.

The most common multivariate used are classified in two main groups: those that decompose data into bilinear components, as Principal Component Analysis (PCA) and Multivariate Curve resolution alternating least squares (MCR-ALS); and those that work on the tridimensional arrangement of the dataset, as Parallel Factor Analysis (PARAFAC) and Tucker 3 models.

Since the arrangement of the dataset in each model is different, interpretation of the results obtained (usually different) may lead to wrong conclusions. That is why a critical comparison and discussion of bi and tridimensional chemometric methods may be of interest.

In this work, bi and tri-dimensional methods were used to investigate an environmental dataset consisting on the extractable concentration of 14 trace elements (Al, As, Cd, Co, Cu, Cr, Fe, Mg, Mn, Ni, Pb, Sn, V, Zn) in sediments collected at eight different sites of the Nerbioi-Ibaizabal estuary (Bilbao, Bay of Biscay, Basque Country) in different sampling campaigns from January 2005 to April of 2009.

Keywords: monitoring , trace elements, sediment, PCA, MCR-ALS, PARAFAC, Tucker 3.

INTRODUCTION

Trace elements, together with the organic pollutants, are the most monitored compounds in environmental research due to their toxicity and persistency. Their presence in sediments can be considered even more risky because they cannot be easily eliminated and they tend to accumulate [1].

In order to obtain more information, an easier interpretation and better visualization of the results of environmental monitoring studies, chemometric methods are appropriate strategies, since they could be considered multidisciplinary tools which are able to obtain analytical conclusions based on statistical analysis [2].

Depending on the chemometric method used the way of creating the studied matrix and consequently, the results obtained from it may differ. That is why the comparison between the two-way (PCA and MCR-ALS) and three-way (PARAFAC and Tucker 3) chemometric methods could be considered essential [3].

With this aim, our work was focused on the sediments of the Nerbioi-Ibaizabal estuary, which is an area that has been subjected to considerable industrial and mining activity throughout its history and it is located in one of the most important urban areas (Bilbao, Basque Country) of the Cantabrian coast [4].

EXPERIMENTAL

Sediments were collected in 8 different sites along the estuary. *Alde Zaharra* (AZ), *Zorrotza* (ZO) and *Arriluze* (AR) are located in the main flow of the river, and other sampling sites in four tributary rivers: *Kadagua* (KA), *Galindo* (GA), *Asua* (AS) and *Gobela* (GO). *Udondo* (UD) is a semiclosed dock close to the Gobela river mouth. The samples were analysed according to procedures previously optimised by our research group [5] in order to measure simultaneously the concentration of fourteen elements. The sampling campaigns were carried out from January of 2005 to April of 2009 every three months.

For the application of the two-way methods (PCA and MCR-ALS) bidimensional data augmentation was obtained placing the different sampling campaigns one on top of the other, so the dimensions of the augmented data matrix were 136 x 14 (8 sites x 17 campaigns and 14 variables).

For the creation of the three-way augmented data, the 17 sampling campaigns were placed one behind the other. Three loading matrices were created after the dataset augmentation: loadings of the first mode (also called scores for two-way data) contained the sampling sites, loadings of the second mode contained the metal composition (loadings for two-way data), and the loadings in the third mode contained the different sampling campaigns.

RESULTS AND DISCUSSION

After developing two-way (PCA and MCR-ALS) and three-way (PARAFAC and Tucker 3) models, the first component always divided the sampling sites according to the pollution profile described by all the metal content (Fig.1). This profile could classifies the sampling sites in two principal groups: the sites with high metal content and the sites with a lower concentration of these trace elements. Thus, AZ, KA and ZO were the sites with the less metal content and AR, GO and UD the sampling sites which presented the highest concentration of the pollutants. AS always appeared between these two groups.

Taking into account the geographical location of the different sites accordingly to the distribution shown by the first principal component, it could be appreciated that the sites placed at the upper part of the estuary were the cleanest ones, whereas the general pollution was increasing down stream (Fig. 2).

Nevertheless, results obtained with PARAFAC model were slightly different, because by this model only GO, AS and UD presented high positive loading values for almost all the trace elements; KA, AZ, GA and ZO showed negative values and AR was between these two groups. It presented an intermediate rank, with positive values for some of the pollutants and negative values for others.

Despite the fact that the results obtained by the second component of the models brought different classifications of the sites, they also gave the chance to know more about the studied areas, supplying more information on which metals had more influence in each zone. GA, for example, was mostly influenced by Ni, Mn Cr and Co, AR was directly related to the Al and Mg content and GO and UD both presented high loading values for V, As, Cu and Cd.

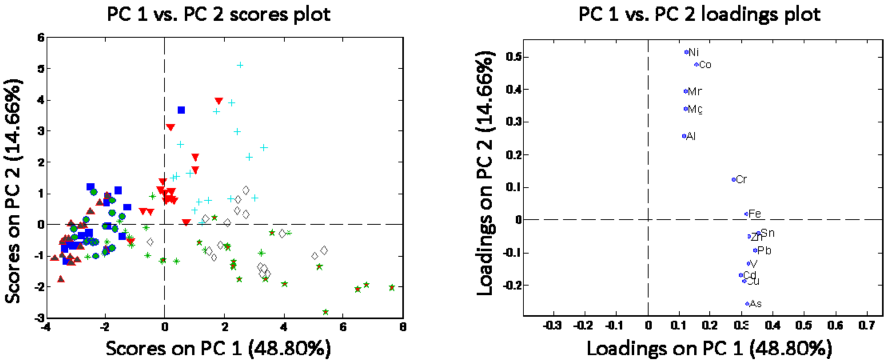

Fig. 1.Scores and Loadings plots from PCA for the first two principal components

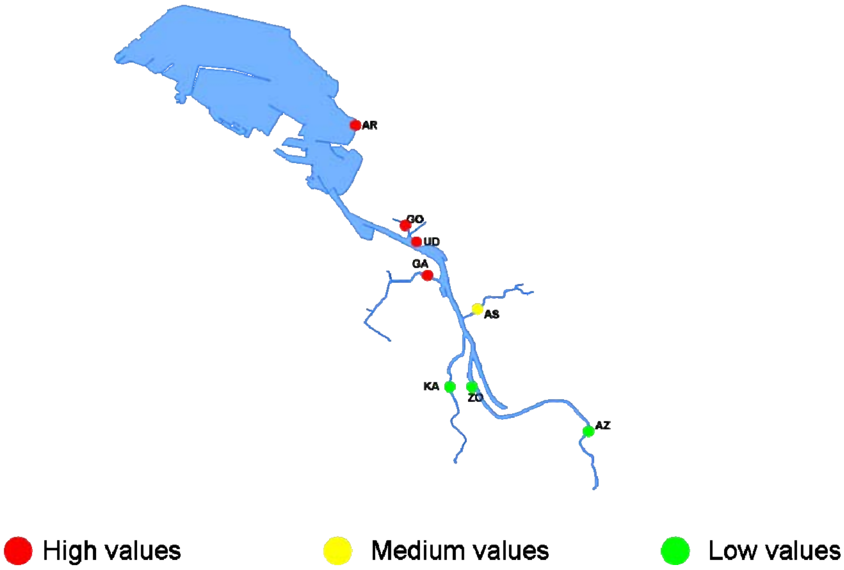

Fig. 2. Contamination level of the sampling sites according to the model created by the loadings with positive values for the first component (PCA, MCR and Tucker 3 models).

Although no significant trend was appreciated with time using bidimensional models (this may be due to the loss of the tridimensional structure of data), the three way chemometric methods allowed us to detect a decreasing trend in the metal content with time (Fig. 3).

It could be said that the decreasing pattern was clearer for the Tucker 3 model, because in the PARAFAC model this tendency was only described by the first component and in the Tucker model this decreasing behaviour of the last loading profile, was always the same for all the components.

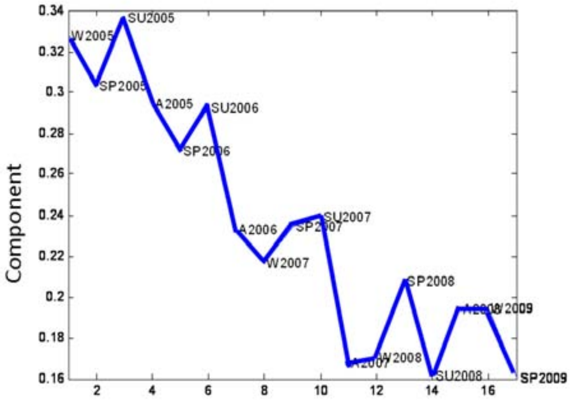

Fig. 3. Plot of the loadings of the third mode from the Tucker 3 model

CONCLUSIONS

In this particular case, the application of a two-way model, such as PCA or MCR, was enough to describe the current situation of the estuary, classifying the sampling sites according to the concentration of the contaminants, but no temporal information was obtained.

Nevertheless, the three-way models (PARAFAC and Tucker 3) presented a clear decreasing tendency with time on the contamination profiles. They gave the possibility to consider a cleaning process in the sites with high concentration of the trace elements, happening together a simultaneous deterioration of the sites with lower metal concentration.

Therefore, in order to obtain more information about the studied area and its near future, the three-way models could be considered more appropriate.

Acknowledgements: This work has been financially supported by the Basque Government through the project IT07-245. A. Gredilla is grateful to the University of the Basque Country (UPV/EHU) for her pre-doctoral fellowship.

REFERENCES

[1] Moros J, Fdez.Ortiz de Vallejuelo S, Gredilla A, de Diego A, Madariaga J M, Garrigues S, De la Guardia M (2009), Environmental Science & Technology 43: 9314-9320

[2] Peré-Trepat E, Ginebreda A, Tauler R (2007), Chemometrics and Intelligent Laboratory Systems 88: 69-83

[3] Evrim A, Rasmus B, Bonnie S (2008), Journal of Chemometrics 22:91-100

[4] Cearreta A, Irabien M J, Leorri E, Yusta I, Quintanilla A, Zabaleta A (2002), Marine Pollution Bulletin 44: 487-503

[5] Fdez.Ortiz de Vallejuelo S, Barrena A, Arana G, de Diego A, Madariaga J M (2009), Journal of Marine Systems 80: 434-439