Lecture at
CMA4CH 2014, Mediterraneum Meeting
Employ the Multivariate Analysis and Chemometrics in Cultural Heritage and Environment Fields
5th ed., Rome, Italy, Europe, 14-17 December 2014

Chemometric Methods for the Analysis of Data Coming From Designed Experiments

F. Marini
Dept. of Chemistry, University of Rome “La Sapienza”, P.le Aldo Moro 5, I-00185 Rome, Italy

Abstract

In the present communication, strategies for dealing with multivariate data coming from designed experiments will be presented. In particular, attention will be focused on the recently proposed ANOVA-Simultaneous Component Analysis (ASCA) and related techniques. ASCA works by decomposing the overall variation within a data set into the contribution of the controlled factors and their interactions, and it models the effect matrices by PCA. The potential of this approach will be illustrated by means of examples taken from environmental chemistry, food chemistry and nutritional metabolomics.

Introduction

Design of experiments, firstly introduced by Fisher [1], is a discipline aiming at finding the experimental conditions under which to carry out the study in order to obtain the maximum information, often about the effect of some process or intervention (the factors) on one or more properties (the responses) of the investigated systems. As a consequence, when analyzing the multivariate responses coming from designed exeriments, particular care should be taken, as the presence of multiple sources of variation can hinder the interpretation (exploratory approach) or introduce nonlinearities (predictive modeling). On the other hand, in the univariate case, analysis of variance [2] allows to partition the overall variation observed in a response into the contribution of the individual sources (main effects and interaction). Based on these considerations, in the last years an approach based on coupling a multivariate ANOVA-like decomposition to the modeling power of bilinear methods has been proposed in the literature, the main member of this family of methods being ANOVA-Simultaneous Component Analysis (ASCA) [3]. While this strategy was originally proposed in the framework of metabolomics, it can be easily generalized to other fields of applications, where experiments are carried out according to a designed scheme.

In the present communicaiton, successful application of this approach to the fields of environmental chemistry, food chemistry and nutritional metabolomics will be presented.

Materials & Methods

ANOVA-Simultaneous Component Analysis (ASCA) [3] is a chemometric method combining an ANOVA-like decomposition of the data matrix **X**, collecting the multivariate signals recorded for the designed experiments, and the bilinear modeling of the resulting effect matrices by means of PCA. For instance, if the experiments are carried out to evaluate the effect of two factors and of their binary interaction on a multivariate signal, ASCA proceeds by partitioning the matrix **X** (after mean centering) into:

$$\mathbf{X}\text{-}\mathbf{1m}^{T} = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_{12} + \mathbf{X}_{res} \qquad (1)$$

Where 1 is a unity vector, **m** is the row vector collecting the means of the columns of **X**, $\mathbf{X}_1$ and $\mathbf{X}_2$ are the matrices accounting for the main effects of factors 1 and 2, respectively, $\mathbf{X}_{12}$ is the effect matrix for the interaction and $\mathbf{X}_{res}$ contains the residual variation. Given this formulation, the entity of the multivariate effect is expressed by the sum of square of the corresponding matrix, and validation of its significance is carried out by means of permutation tests [4].

Successively, each of the effect matrices $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_{12}$ is described by an individual PCA model, so that inspection of the corresponding loadings can provide information about which regions of the multivariate signal are most affected by the controlled factors.

Results

As anticipated, the proposed approach has been applied by the author and coworkers to solve problems in different fields, from food chemistry and nutritional metabolomics to phytochemistry and environmental chemistry. In particular, in the field of environmental chemistry, this approach has been adopted to investigate the applicability of a "toolbox" designed for microbially assisted phytoremediation [4]. In particular, the effect of viromine, mychorriza and bacteria (together with field position) on soil and plant parameter was investigated. The use of ASCA allowed to assess that only the effect of viromine (also modulated by the interaction either with mychorriza or bacteria) was statistically significant.

PCA analysis on the effect matrices evidenced that the first principal component accounts for the main effect of viromine and for that part of the interaction between viromine and either bacteria or mychorriza which modulates the main effect only in magnitude but does not change its direction, while PC2 accounts for the effect of the interaction that results also in a change in the direction of the viromine-induced variation due to the presence or absence of the other constituent (Figure 1).

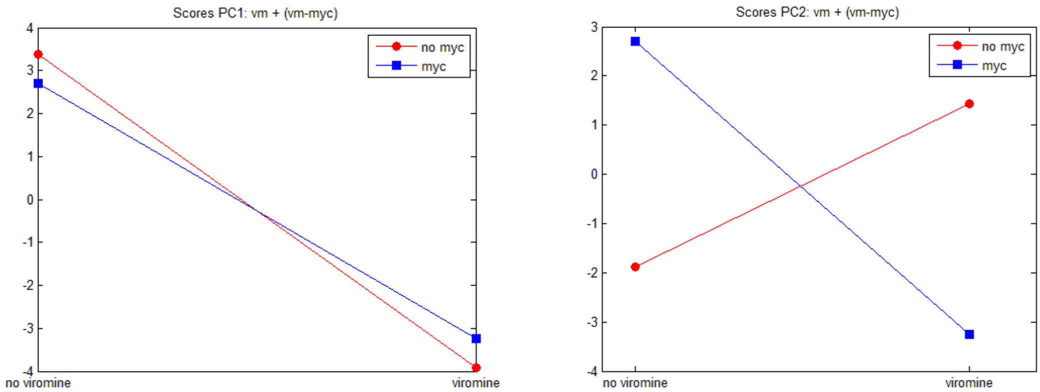

Figure 1 – PCA on the matrix corresponding to the main effect of viromine (vm) and its interaction with the mychorriza (myc): projection of the design point onto the first (PC1) and second (PC2) component.

On the other hand, inspection of the loadings allows interpretation of the effects in terms of the change of the measured soil and plant parameters.

Conclusions

ASCA and related methods constitute a valid and effective approach to analyze data from designed experiments not only in the field of metabolomics, for which they were originally designed, but also in other different ambits of applied chemometrics.

References

1) R.A. Fisher, *The design of experiments (9th ed.)*, McMillan Publisher Ltd, London, UK, (1971). ISBN 0-02-844690-9.
2) H. Scheffé, *The Analysis of Variance*, Wiley, New York, NY, (1959). ISBN 978-0471345053.
3) J.J. Jansen, H.C.J. Hoefsloot, J. van der Greef, M.E. Timmerman, J.A. Westerhuis, A.K. Smilde, ASCA: analysis of multivariate data obtained from an experimental design, *J. Chemometr.*, 19, (2005), 469–481.
4)A.R. Sprocati, C. Alisi, V. Pinto, M.R. Montereali, P. Marconi, F. Tasso, K. Turnau, G. De Giudici, K. Goralska, M. Bevilacqua, F. Marini, C. Cremisini, Assessment of the applicability of a "toolbox" designed for microbially assisted phytoremediation: the case study at Ingurtosu mining site (Italy), *Env. Sci. Pollut. Res.*, 21(11), 2014, 6939-6951.