A non linear approach to Non Symmetrical Data Analysis

Jean-François Durand , Yves Escoufier

ENSAM-INRA-UM II, Unité de Biométrie,
9 place Pierre Viala, Montpellier, France

Summary: The aim of this paper is to present a non linear additive extension for the linear non symmetrical analysis of two data arrays. The two groups of variables associated with each of these arrays play a non symmetrical role (the response variables and the explanatory variables). As in the linear approach the graphical issues provided by the additive approach are very easy to interpret. An application is given on spatial correlation models relating covariance to distance for different regions of the brain.

1. A spatial correlation problem in neurology

A set of $n = 15$ regions of the brain have been studied in Worsley et al. (1991) where two data arrays, $\mathbf{Y}$ and $\mathbf{X}$, are considered. The $n \times p$ matrix $\mathbf{Y}$, $p = 20$, contains the logarithms of the glucose metabolism measurements made on p subjects for the n regions, the $n \times n$ covariance matrix being denoted $\boldsymbol{\Sigma}$. The matrix $\mathbf{X}$ is the $n \times 3$ centered regions coordinates.

The aim of the Worsley et al.'s article is to model the covariance matrix $\boldsymbol{\Sigma}$ as a linear function of the squared distances between the regions

$$\boldsymbol{\Sigma} = \alpha_1 \mathbb{1}_n \mathbb{1}_n' - \alpha_2 \mathbf{D} + \alpha_3 \mathbf{I}_n \ , \tag{1}$$

where $\mathbf{D}$ is the $n \times n$ matrix of squared distances, $\mathbb{1}_n$ the vector of n repeated 1, and $\mathbf{I}_n$ the $n \times n$ identity matrix. The parameter α_1 may be interpreted as a subject effect, α_2 as a spatial effect and α_3 as an error or nugget effect.

In order to eliminate the subject effect, the $\mathbf{Y}$ data have been centered. Let

$$\mathbf{C} = \mathbf{I}_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n'$$

be the $n \times n$ centering matrix, then the normalised observation matrix is $\mathbf{CY}$, and its variance is

$$\mathbf{C}\boldsymbol{\Sigma}\mathbf{C} = 2\alpha_2 \mathbf{X}\mathbf{X}' + \alpha_3 \mathbf{C} \ . \tag{2}$$

It is clear that a sufficient condition for $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}$ to be positive definite is the nonnegativeness of α_2, α_3. Note that the covariance structure model is linear, plus a centered white noise.

The model (2) is extended in the following sections by using a more general framework: the non symmetrical data analysis of two data sets. Section 2 presents a general linear approach for solving problems that concern the structural analysis of multivariate data. The linear trend of the model (2) will then be extended in

$$\mathbf{C}\boldsymbol{\Sigma}\mathbf{C} = \mathbf{X}\mathbf{Q}\mathbf{X}' + \alpha_3 \mathbf{C} \ , \tag{3}$$

where the 3×3 symmetric positive semidefinite matrix $\mathbf{Q}$ replaces $2\alpha_2$ in (2).

A non linear additive extension of (3) is detailed in section 3. Each column of $\mathbf{X}$ is transformed by using a linear combination of normalized B-splines so that $\mathbf{X}$ is replaced by a matrix $\mathbf{X}(\mathbf{a})$ with the same dimensions and depending on $\mathbf{a}$, the vector of the spline parameters. Note that there exists a particular vector noted $\boldsymbol{\xi}$, very easy to compute from the spline knots, that gives $\mathbf{X}(\boldsymbol{\xi}) = \mathbf{X}$. The proposed model

$$\mathbf{C}\boldsymbol{\Sigma}\mathbf{C} = \mathbf{X}(\mathbf{a})\mathbf{Q}\mathbf{X}(\mathbf{a})' + \alpha_3\mathbf{C} , \tag{4}$$

may then be considered as a non linear extension for (3) and (2).

The last section compares the results of the three regression models applied on the brain data.

2. A general linear approach

2.1 Mathematical background

Let two statistical studies $\mathcal{E}_i$, $i = 1, 2$, be defined by the triples $(\mathbf{X}_i, \mathbf{Q}_i, \mathbf{W})$ where $\mathbf{X}_i$ are $n \times p_i$ data matrices whose elements are respectively the values of the p_i variables of $\mathcal{E}_i$ measured on the same n objects.

The semimetrics $\mathbf{Q}_i$ that are $p_i \times p_i$ symmetric positive semidefinite matrices, are used to compute euclidian distances between objects of either $\mathcal{E}_i$. If the data arrays $\mathbf{X}_1$ and $\mathbf{X}_2$ are generally imposed, the statistician can choose different metrics along with the nature of the Data Analysis problems, see Sabatier et al. (1989), Escoufier and Holmes (1990).

The matrix $\mathbf{W}$, which is the same for the two studies, is called the matrix of weights associated with the objects. It is a $n \times n$ diagonal matrix with nonnegative diagonal elements that sum to 1.

A unifying tool for the linear Exploratory Data Analysis is the generalised Principal Component Analysis as defined in (Escoufier, 1987), see also Escoufier and Holmes (1990). This way of doing uses the eigenanalysis of $\mathbf{X}_i\mathbf{Q}_i\mathbf{X}_i'\mathbf{W}$ in order to have graphics for the representation of the objects of $\mathcal{E}_i$. This matrix is called the characteristic operator for the representation of the objects and the corresponding norm is $\|\mathbf{X}_i\mathbf{Q}_i\mathbf{X}_i'\mathbf{W}\| = \sqrt{tr((\mathbf{X}_i\mathbf{Q}_i\mathbf{X}_i'\mathbf{W})^2)}$.

The aim of the non symmetrical linear data analysis problem is to check the existence of linear relationships between the $\mathcal{E}_1$ response variables and the $\mathcal{E}_2$ explanatory variables by comparing the characteristic operators of those two studies

$$\min_{\mathbf{Q}_2} \|\mathbf{X}_1\mathbf{Q}_1\mathbf{X}_1'\mathbf{W} - \mathbf{X}_2\mathbf{Q}_2\mathbf{X}_2'\mathbf{W}\|^2 . \tag{5}$$

A solution is $\overline{\mathbf{Q}}_2 = (\mathbf{X}_2'\mathbf{W}\mathbf{X}_2)^-(\mathbf{X}_2'\mathbf{W}\mathbf{X}_1)\mathbf{Q}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_2)(\mathbf{X}_2'\mathbf{W}\mathbf{X}_2)^-$, see Escoufier (1987). If we note $\mathbf{P}_{\mathbf{X}_2}$ the W-orthogonal projector on the columns of $\mathbf{X}_2$, the eigenanalysis of the two operators respectively associated with the triples $(\mathbf{X}_2, \overline{\mathbf{Q}}_2, \mathbf{W})$ and $(\mathbf{P}_{\mathbf{X}_2}\mathbf{X}_1, \mathbf{Q}_1, \mathbf{W})$ provide the same representation of the objects. Choosing different $\mathbf{Q}_1$ and $\mathbf{W}$ than $\mathbf{I}_{p_1}$ and $n^{-1}\mathbf{I}_n$ can be a procedure worth considering, see Sabatier et al. (1989).

2.2 Application to the spatial correlation problem

Going back to model (3), we want to find the best $\mathbf{Q}$ and α_3 that are solution of the optimization problem

$$\min_{\alpha_3, \mathbf{Q}} \|\mathbf{C}\ \boldsymbol{\Sigma}\mathbf{C} - \mathbf{X}\mathbf{Q}\mathbf{X}' - \alpha_3\mathbf{C}\|^2 .$$

The objective function being rewritten, we have

$$\min_{\alpha_3, \mathbf{Q}} \|\mathbf{C}(\boldsymbol{\Sigma} - \alpha_3 \mathbf{I}_n)\mathbf{C} - \mathbf{X}\mathbf{Q}\mathbf{X}'\|^2 . \tag{6}$$

The comparison with (5) leads to the corresponding choices

$$\mathbf{X}_1 = \mathbf{C}, \quad \mathbf{Q}_1 = \boldsymbol{\Sigma} - \alpha_3 \mathbf{I}_n, \quad \mathbf{X}_2 = \mathbf{X}, \quad \mathbf{W} = n^{-1}\mathbf{I}_n .$$

Since $\mathbf{X}$ is centered, $\mathbf{CX} = \mathbf{X}$ and an explicit solution of the normal equations is

$$\begin{aligned} \overline{\alpha}_3 &= \frac{tr(\boldsymbol{\Sigma}(\mathbf{C} - \mathbf{P}_{\mathbf{X}}))}{n - 1 - r(\mathbf{X})} \\ \overline{\mathbf{Q}} &= (\mathbf{X}'\mathbf{X})^- \mathbf{X}' (\boldsymbol{\Sigma} - \overline{\alpha}_3 \mathbf{I}_n) \mathbf{X} (\mathbf{X}'\mathbf{X})^- , \end{aligned}$$

where $\mathbf{P}_{\mathbf{X}}$ is the the usual projector on the columns of $\mathbf{X}$. Note that $tr(\mathbf{C}) = n - 1$ and $tr(\mathbf{P}_{\mathbf{X}}) = r(\mathbf{X})$.

3. Additive spline non symmetrical analysis

The mathematical background of section 2.1 is now being modified while preserving the handy context of euclidian spaces and projectors. The use of regression spline functions for transforming the explanatory variables provides a natural nonlinear semi-parametric extension for the linear non symmetrical analysis as presented in section 2.1, see Durand (1993).

3.1 Transformation of the predictors

Let us choose, for simplicity, the same kind of B-splines for transforming the predictors: K interior knots and order m so that $r = m + K$ is the dimension of the spline space. The j^{th} column of $\mathbf{X}_2$, noted $\mathbf{X}_2{}^j$, is then replaced by

$$\mathbf{X}_2{}^j(\mathbf{a}^j) = \mathbf{B}^j \mathbf{a}^j ,$$

where $\mathbf{B}^j$ is the $n \times r$ coding matrix of $\mathbf{X}_2{}^j$, and $\mathbf{a}^j$ the vector of the r parameters. The matrix $\mathbf{X}_2$ is now being modified in

$$\mathbf{X}_2(\mathbf{a}) = (\mathbf{X}_2{}^1(\mathbf{a}^1)|\ldots|\mathbf{X}_2{}^{p_2}(\mathbf{a}^{p_2})) = \mathbf{BA} ,$$

with $\mathbf{B} = (\mathbf{B}^1|\ldots|\mathbf{B}^{p_2})$, $\mathbf{a}' = (\mathbf{a}^{1\prime}|\ldots|\mathbf{a}^{p_2\prime})$ and

$$\mathbf{A} = \begin{bmatrix} a_1^1 & 0 & . & . & 0 \\ . & . & . & . & . \\ a_r^1 & 0 & . & . & . \\ 0 & a_1^2 & . & . & . \\ . & . & . & . & . \\ . & a_r^2 & . & . & 0 \\ . & 0 & . & . & a_1^{p_2} \\ . & . & . & . & . \\ 0 & 0 & . & . & a_r^{p_2} \end{bmatrix} .$$

The knots sequence used for transforming the j^{th} predictor being written

$$\min_{k=1,\ldots,n} X_2{}_k^j = t_1^j = \ldots = t_m^j < t_{m+1}^j \leq \ldots \leq t_{m+K}^j < t_{m+1+K}^j = \ldots = t_{2m+K}^j = \max_{k=1,\ldots,n} X_2{}_k^j ,$$

the vector $\boldsymbol{\xi}^j$ of the parameters whose components are, for $i = 1, \ldots, r$,

$$\xi_i^j = \frac{1}{m-1} \sum_{k=1,\ldots,m-1} t_{i+k}^j ,$$

keeps the j^{th} predictor invariant,

$$\mathbf{X_2}^j(\boldsymbol{\xi}^j) = \mathbf{X_2}^j .$$

The role of the *nodal coefficients* (Chenin et al. (1985)) stored in $\boldsymbol{\xi}$ is twofold. First, the additive model as defined in section 3.2 can effectively take into account linear relationships. The numerical relaxation method inititializes with $\mathbf{a} = \boldsymbol{\xi}$ so that the solution of the linear problem is obtained at the first step of the algorithm.

3.2 The additive spline approach

Denote $\overline{\mathbf{a}}$ and $\overline{\mathbf{Q}}_2$ a solution of the optimization problem extending (5)

$$\min_{\mathbf{a},\mathbf{Q}_2} \|\mathbf{X}_1\mathbf{Q}_1\mathbf{X}_1'\mathbf{W} - \mathbf{X}_2(\mathbf{a})\mathbf{Q}_2\mathbf{X}_2(\mathbf{a})'\mathbf{W}\|^2 . \tag{7}$$

For fixed $\mathbf{a}$, an optimal solution is

$$\overline{\mathbf{Q}}_2(\mathbf{a}) = (\mathbf{X}_2(\mathbf{a})'\mathbf{W}\mathbf{X}_2(\mathbf{a}))^-(\mathbf{X}_2(\mathbf{a})'\mathbf{W}\mathbf{X}_1)\mathbf{Q}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_2(\mathbf{a}))(\mathbf{X}_2(\mathbf{a})'\mathbf{W}\mathbf{X}_2(\mathbf{a}))^- ,$$

so that $\overline{\mathbf{Q}}_2 = \overline{\mathbf{Q}}_2(\overline{\mathbf{a}})$.

There is not here an explicit solution for the normal equations. The expression of the gradient of the objective function with respect to $\mathbf{a}$ being computed, see Durand (1993), a relaxation method is processed which alternates a step of computing $\overline{\mathbf{Q}}_2(\mathbf{a})$ for fixed $\mathbf{a}$, with a step of a gradient descent method for fixed $\mathbf{Q}_2$.

The underlying model $\hat{\mathbf{X}}_1 = \mathbf{P}_{\mathbf{X}_2(\overline{\mathbf{a}})}\mathbf{X}_1$ is then additive since the characteristic operators associated with the triples $(\mathbf{X}_2(\overline{\mathbf{a}}), \overline{\mathbf{Q}}_2, \mathbf{W})$ and $(\mathbf{P}_{\mathbf{X}_2(\overline{\mathbf{a}})}\mathbf{X}_1, \mathbf{Q}_1, \mathbf{W})$ are identical. The projection matrix $\mathbf{P}_{\mathbf{X}_2(\mathbf{a})}$ may be considered as a multivariate linear smoother, see Hastie and Tibshirani (1990).

The eigenanalysis of the two preceding triples provides graphical displays as in the linear case (note that $\mathbf{X}_2(\mathbf{a})$ is centered for any $\mathbf{a}$ if $\mathbf{B}$ has been initially centered).

Different choices for $\mathbf{Q}_1$ and $\mathbf{W}$ allow the user to solve non linear data analysis problems. For example, Additive Spline Discriminant Analysis is the non symmetrical analysis of the indicator matrix $\mathbf{X}_1$ of p_1 classes associated with the Mahalanobis metric $\mathbf{Q}_1 = (\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}$ and the training sample matrix $\mathbf{X}_2$, see Durand (1992).

3.3 The non linear spatial correlation model

Now returning to the notation of section 1, the model (4) is associated with the optimization problem

$$\min_{\alpha_3,\mathbf{Q},\mathbf{a}} \|\mathbf{C}(\boldsymbol{\Sigma} - \alpha_3\mathbf{I}_n)\mathbf{C} - \mathbf{X}(\mathbf{a})\mathbf{Q}\mathbf{X}(\mathbf{a})'\|^2 . \tag{8}$$

Unlike the linear case the solution of the normal equations is not explicit. An iterative relaxation method is implemented that processes alternatively a step of computing $\overline{\mathbf{Q}}(\mathbf{a})$ as $\overline{\mathbf{Q}}_2(\mathbf{a})$ defined in section 3.2, a step of a gradient descent method with respect to $\mathbf{a}$, and a step of computing α_3 as in the linear case.

4. Results of the three correlation models

The present application does not use all the capabilities of the non symmetrical data analysis methods that are developped in sections 2 and 3. The only point we are interested in is the comparison of the three correlation models of the previous sections, see Vie (1993).

Let us note $\widehat{\mathbf{C\Sigma C}}$ the estimate of $\mathbf{C\Sigma C}$, $\widehat{\mathbf{R}}$ and $\mathbf{R}$ the corresponding correlation matrices. Let $\hat{\mathbf{r}}$ and $\mathbf{r}$ be the vectors of the 15×7 significant values derived from $\widehat{\mathbf{R}}$ and $\mathbf{R}$. The subscripts l, L and s point out the respective models (2), (3) and (4).

The efficiency of the three models can be gauged in Figure 1 where the optimal values of the corresponding loss functions are compared.

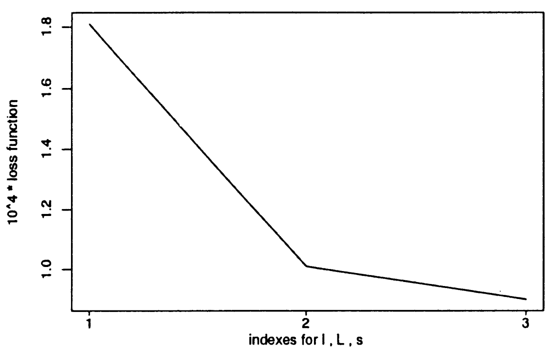

Figure 1: Gain in fit of the non symmetrical methods L and s compared with the simple linear approach l.

Another way of testing the efficiency of the methods is to plot $\hat{\mathbf{r}}$ against $\mathbf{r}$ for the three models, see Figures 2, 3 and 4.

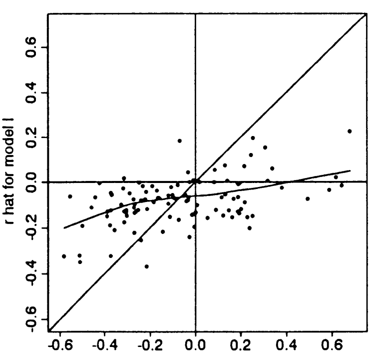

Figure 2: Validity of the linear model.

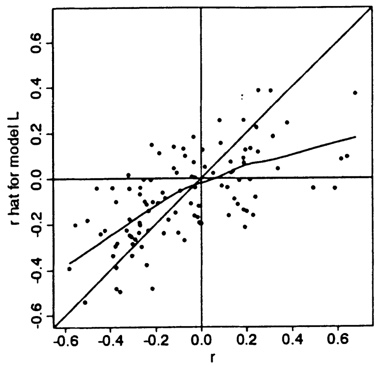

Figure 3: Validity of the generalized linear model.

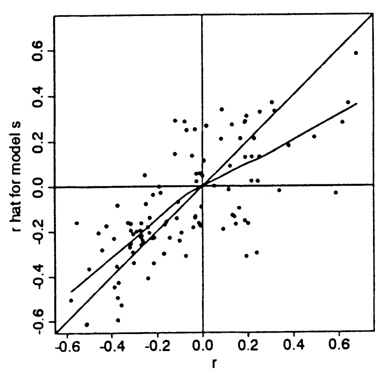

Figure 4: Validity of the additive spline model.

The trend of the response $\hat{\mathbf{r}}$ considered as a function of $\mathbf{r}$ is summarized by the locally-weighted smoother of Cleveland (1979). The gain in fit can easily be seen on Figures 2, 3 and 4 along with the three methods.

The geometry of the brain has been modified by the transformation of the predictors. Considering now the methods L and s only, let us write the Choleski decomposition of the optimal $\overline{\mathbf{Q}}$ in the (6) and (8) problems

$$\overline{\mathbf{Q}} = \overline{\mathbf{M}}\,\overline{\mathbf{M}}' .$$

The matrix of the region coordinates are linearily or additively transformed: $\mathbf{X}$ is replaced by $\mathbf{X}\overline{\mathbf{M}}$ in the L method, by $\mathbf{X}(\overline{\mathbf{a}})\overline{\mathbf{M}}$ in the additive spline s. Although three dimensions views are more significant, two dimensions scatterplots can bring information: for example, Figure 5 shows the 2 and 3 region coordinates for those three matrices. A scaling factor (10^{-2}) seems the only transformation for the linear method. Projected regions have been more revealingly modified (regions OC and MF) for the spline method.

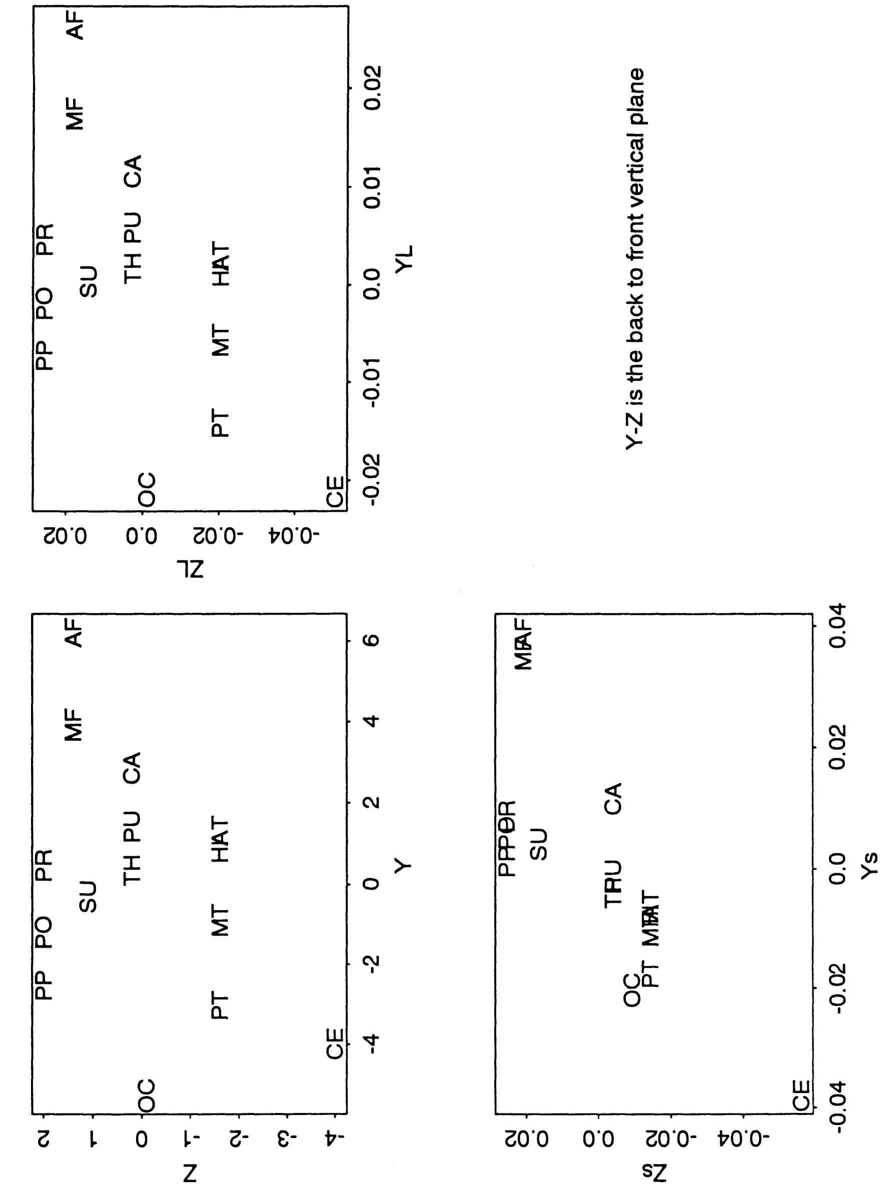

Figure 5: Transformation of the geometry of the brain for the vertical projection plane. L and s correspond to the generalized linear model and to the additive spline model.

ACKNOLEDGMENTS: We thank Frédérique Kazi Aoual for supplying us with the expression of the optimal value of α_3 in section 2.2.

References:

CHENIN, P., COSNARD, M., GARDAN, Y., ROBERT, F., ROBERT, Y. and WITOMSKI, P. (1985): Mathématiques et CAO, méthodes de base. Paris: Hermes publishing.

CLEVELAND, W.S. (1979): Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74, 368, 829-836.*

DURAND, J.F. (1992): Additive Spline Discriminant Analysis. In: Y. Dodge and J.C. Whittaker (Eds.): *Computational Statistics*, Heidelberg: Physica-Verlag, I, 145-150.

DURAND, J.F. (1993): Generalized Principal component analysis with respect to instrumental variables via univariate spline transformations. In press in *Computational Statistics & Data Analysis.*

ESCOUFIER, Y. (1987): Principal Components Analysis with respect to Instrumental Variables. *European Courses in Advanced Statistics*, University of Napoli, 285-299.

ESCOUFIER, Y., and HOLMES, S. (1990): Data analysis in France. *Biometric Bulletin, 7, 3, 27-28.*

HASTIE, T., and TIBSHIRANI, R. (1990): Generalized additive models. London: Chapman and Hall.

SABATIER, R., LEBRETON, J.D., CHESSEL, D. (1989): Principal component analysis with instrumental variables as a tool for modelling composition data. In: R. Coppi et S. Bolasco (Eds.), *Multiway Data Analysis.* Elsevier Sc. Pub., 341-352.

VIE, K. (1993): Ajustement non linéaire de corrélations spatiales. Application à la neuroscience. *Rapport de DEA de Biostatistique. UM II-ENSAM, Montpellier.*

WORSLEY, K.J., EVANS, A.C., STROTHER, S.C., and TYLER, J.L. (1991): A linear spatial correlation model, with applications to positron emission tomography. *Journal of the American Statistical Association, 86, 413, 55-67.*