

## Análisis de la evolución en el tiempo para datos con estructura de grupos: STATIS dual canónico y modelo de medidas repetidas doblemente multivariantes

Analysis of Time Evolution for Group Structured Data: Canonical  
Dual STATIS and Doubly Multivariate Repeated Measures Model

AMPARO VALLEJO<sup>1,a</sup>, JOSÉ LUIS VICENTE<sup>2,b</sup>, PURIFICACIÓN GALINDO<sup>2,c</sup>,  
MARGARITA FERNÁNDEZ<sup>3,d</sup>, CAMINO FERNÁNDEZ<sup>3,e</sup>, ELOY BÉCARES<sup>3,f</sup>

<sup>1</sup>DEPARTAMENTO DE MATEMÁTICAS, UNIVERSIDAD DE ANTIOQUIA, MEDELLÍN, COLOMBIA

<sup>2</sup>UNIVERSIDAD DE SALAMANCA, SALAMANCA, ESPAÑA

<sup>3</sup>UNIVERSIDAD DE LEÓN, LEÓN, ESPAÑA

---

### Resumen

En este trabajo proponemos dos soluciones al problema que se plantea cuando se pretende analizar datos multivariantes, para un conjunto de individuos con estructura de grupos, que además han sido replicados bien sea en ocasiones o en situaciones experimentales diferentes. La primera solución se obtiene aplicando la versión dual del STATIS canónico propuesto por Vallejo-Arboleda et al. (2007); la segunda, aplicando el modelo de medidas repetidas doblemente multivariantes. Usamos los datos del proyecto SWALE (Stephen et al. 2004) para una selección de 7 variables físico-químicas medidas para 4 tratamientos, cada uno con 9 réplicas, durante 4 semanas (elegidas entre las 10 iniciales del proyecto) para comparar las dos soluciones.

**Palabras clave:** Análisis canónico, medidas repetidas.

---

<sup>a</sup>Profesora. E-mail: [avallejo@matematicas.udea.edu.co](mailto:avallejo@matematicas.udea.edu.co)

<sup>b</sup>Profesor. E-mail: [villardon@usal.es](mailto:villardon@usal.es)

<sup>c</sup>Profesora. E-mail: [pgalindo@usal.es](mailto:pgalindo@usal.es)

<sup>d</sup>Profesora. E-mail: [margarita.alaez@unileon.es](mailto:margarita.alaez@unileon.es)

<sup>e</sup>Profesora. E-mail: [camino.alaez@unileon.es](mailto:camino.alaez@unileon.es)

<sup>f</sup>Profesor. E-mail: [ebcm@unileon.es](mailto:ebcm@unileon.es)

### Abstract

In this work we propose two solutions to the problem we consider when we have multivariate experimental data of individuals with structure of groups that have been repeated at different occasions or experimental situations. We obtain the first solution applying dual version of canonical STATIS proposed by Vallejo-Arboleda et al. (2007), and the second with doubly multivariate repeated measures model. We use the data of SWALE project (Stephen et al. 2004), with 7 physical-chemical variables, measured in 4 treatments, each one with 9 repetitions and during 4 weeks (selected between 10 original weeks in the project) to compare these solutions.

**Key words:** Canonical analysis, Repeated measurements.

## 1. Introducción

Frecuentemente se obtienen datos multivariantes experimentales de individuos, con estructura de grupos, que se han replicado en diferentes ocasiones, como los que se obtienen en diseños multivariantes de medidas repetidas. Uno de los objetivos que se proponen es buscar las diferencias entre los efectos de los tratamientos en las diferentes ocasiones y si las diferencias en los efectos tienen, o no, una estructura común.

Se propone una solución aplicando la versión dual del STATIS canónico, propuesto por Vallejo-Arboleda et al. (2007) y buscando un subespacio de referencia común para el análisis de variables canónicas de todas las ocasiones. La proyección de las matrices de medias en este espacio de referencia común es una representación biplot propuesta por Gabriel (1971), que permite caracterizar las trayectorias de las medias de los tratamientos en el espacio de las variables promedio. Esta técnica, llamada STATIS dual canónica, incluye además una representación biplot de las medias de los grupos en el espacio de las variables canónicas comunes. El método STATIS dual canónico tiene la misma estructura del STATIS-ACT dual introducida por Escoufier (1973) y L'Hermier des Plantes (1976), desarrollado por Lavit (1988). En este caso, se parte de las matrices de medias de los grupos en todas las variables para cada ocasión y se toman la inversa de la matriz de covarianza común y la matriz diagonal de tamaños de los grupos como métricas para los grupos y las variables, respectivamente. Por tanto, se supone que las matrices de covarianza son homogéneas en todas las ocasiones dentro de los grupos y es necesario obtener un estimador de la matriz de covarianza común dentro de los grupos. El objetivo es determinar las combinaciones de variables que maximizan la dispersión entre grupos, relativa a la dispersión dentro de los grupos. Otra solución es aplicar el diseño de medidas repetidas doblemente multivariantes. Si se cumplen los supuestos de multinormalidad, homogeneidad de las matrices de dispersión independencia de las puntuaciones e igualdad de las varianzas correspondientes a las diferencias entre las ocasiones de medida (esfericidad), tales diseños permiten contrastar dos tipos de hipótesis: sobre el efecto total de los tratamientos y sobre el efecto de los tratamientos a través de las diferentes ocasiones. Cuando el supuesto

de esfericidad no se cumple, el análisis puede abordarse con los grados de libertad corregidos.

## 2. El método STATIS dual canónico

### 2.1. Descripción de los datos y definiciones

Se denota por  $K$  el número de ocasiones o condiciones experimentales. Para  $k = 1, 2, \dots, K$ , las observaciones se disponen en matrices  $\mathbf{X}_k$  de tamaño  $I \times K$ , donde  $J$  es el número de variables e  $I$  el número de individuos clasificados en  $G$  grupos mutuamente excluyentes. El tamaño de cada grupo es  $I_g (g = 1, \dots, G)$ , con  $\sum_g I_g = I$ . Sin pérdida de generalidad, puede suponerse que cada matriz está centrada por columnas. Sean las matrices  $\bar{\mathbf{X}}_k$  de tamaño  $(G \times J)$ , que contienen las medias de cada grupo para todas las variables en la ocasión  $k$ ,  $\mathbf{D} = \text{diag}(I_1, \dots, I_G)$ , con los tamaños muestrales de los grupos en la diagonal;  $\mathbf{B}_k = \frac{1}{G-1} \bar{\mathbf{X}}_k^T \mathbf{D} \bar{\mathbf{X}}_k$ , la matriz de covarianzas entre los grupos; y  $\mathbf{S}_k = \frac{1}{I-G} \mathbf{X}_k^T \mathbf{X}_k - \bar{\mathbf{X}}_k^T \mathbf{D} \bar{\mathbf{X}}_k$ , la matriz de covarianzas dentro de los grupos, en cada una de las ocasiones.

Un estudio se define como una terna compuesta por la matriz de medias y las métricas de las filas y las columnas. Esto es  $(\bar{\mathbf{X}}_k, \mathbf{M}_f, \mathbf{D})$ , para  $k = 1, \dots, K$ , donde  $\mathbf{D}$  y  $\mathbf{M}_f$  son las métricas de las columnas (variables) y las filas (grupos), respectivamente; para esta última, más adelante se considerarán dos alternativas.  $\mathbf{R}_k = \bar{\mathbf{X}}_k^T \mathbf{D} \bar{\mathbf{X}}_k$  es el objeto representativo del estudio  $k$ ; contiene los productos cruzados entre las variables con respecto a la métrica  $\mathbf{D}$ . El producto escalar de Hilbert Schmidt para dos objetos  $\mathbf{R}_k$  y  $\mathbf{R}_{k'}$ , con respecto a las métricas  $\mathbf{M}_f$  y  $\mathbf{D}$  para las filas y las columnas respectivamente (Escoufier 1973), se define como

$$\langle \mathbf{R}_k | \mathbf{R}_{k'} \rangle_{HS} = \text{tr} \left( \mathbf{M}_f \bar{\mathbf{X}}_k^T \mathbf{D} \bar{\mathbf{X}}_k \mathbf{M}_f \bar{\mathbf{X}}_{k'}^T \mathbf{D} \bar{\mathbf{X}}_{k'} \right)$$

Este producto escalar permite establecer una medida de proximidad entre los dos estudios. Se denota por  $\mathbf{S}^R$  la matriz de tamaño  $K \times K$  que contiene los productos escalares de Hilbert Schmidt para cada par de ocasiones. La diagonal de esta matriz contiene las normas de los objetos al cuadrado, esto es,  $\|\mathbf{R}_k\|^2 = \langle \mathbf{R}_k | \mathbf{R}_k \rangle_{HS}$ . Cuando las normas de cada estudio son muy diferentes, es muy conveniente normalizarlas dividiendo cada objeto entre su norma. Si los estudios están normados, entonces el producto escalar coincide con el coeficiente de correlación  $RV$  (Escoufier 1980), y se define como

$$RV \langle \mathbf{R}_k | \mathbf{R}_{k'} \rangle = \frac{\langle \mathbf{R}_k | \mathbf{R}_{k'} \rangle}{\sqrt{\langle \mathbf{R}_k | \mathbf{R}_k \rangle \langle \mathbf{R}_{k'} | \mathbf{R}_{k'} \rangle}} \quad (1)$$

En general, se emplearán los estudios normados. Este coeficiente toma valores entre 0 y 1; para matrices definidas positivas, puede entenderse como un coseno entre matrices (Abdí 2007). Por esto, algunas veces se llama también coeficiente de correlación matricial.

Como se ha supuesto que existe una métrica para las filas  $\mathbf{M}_f$  común a todos los estudios, se proponen las siguientes opciones para estimarla:

- $\mathbf{M}_f = \mathbf{S}^{-1}$ , donde  $\mathbf{S} = \frac{1}{K} \sum_{k=1}^K \mathbf{S}_k$ , con  $\mathbf{S}_k = \frac{1}{I-G} (\mathbf{X}_k^T \mathbf{X} - \bar{\mathbf{X}}_k^T \mathbf{D} \bar{\mathbf{X}}_k)$ , la matriz de covarianzas dentro de los grupos en cada una de las ocasiones. Es necesario suponer además que  $\mathbf{S}$  es definida positiva.
- Utilizar un consenso basado en STATIS-ACT como matriz común. Esto es, si el objeto representativo de cada estudio es  $S_k$ , la matriz de covarianzas dentro de los grupos, puede calcularse una matriz de covarianzas dentro de los grupos consenso utilizando el propio método STATIS-ACT, y definiendo los productos escalares entre las matrices de covarianzas dentro de los grupos como  $\langle \mathbf{S}_k | \mathbf{S}_{k'} \rangle_{HS} = \text{tr}(\mathbf{S}_k \mathbf{S}_{k'})$ . Si se organizan los productos escalares (o las correlaciones vectoriales) en una matriz  $\mathbf{S}^S$ , la matriz de covarianzas dentro de los grupos consenso se calcula como  $\mathbf{M}_f = \mathbf{S}^{-1}$ , con  $\mathbf{S} = \sum_{k=1}^K \alpha_k \mathbf{S}_k$ , donde las ponderaciones se obtienen como  $\alpha_k = \frac{1}{\sqrt{\phi_1}} \left( \sum_{\tau=1}^K \sqrt{S_{\tau\tau}^S} \right) P_{k1}$ , donde  $p_{k1}$  es la  $k$ -ésima componente de la primera componente principal de la matriz  $\mathbf{S}^S$  asociada al valor propio más grande  $\phi_1$  y  $S_{\tau\tau}^S$  es la norma cuadrática del  $\tau$ -ésimo objeto. En este caso, el objeto es la variabilidad dentro de los grupos, mientras que el objeto definido antes es la variabilidad entre grupos.

## 2.2. La interestructura: similaridades globales entre las ocasiones

El objetivo del análisis de la *interestructura* es decidir si existe o no una estructura común entre los  $K$  objetos representativos  $\mathbf{R}_k$ . El análisis de las similaridades entre los objetos se obtiene de la representación en componentes principales de la matriz  $\mathbf{S}^R$ . Las coordenadas  $\mathbf{E}$  de los puntos se obtienen de la descomposición espectral  $\mathbf{S}^R = \mathbf{L}\mathbf{\Gamma}\mathbf{L}^T$ ; se toma  $\mathbf{E} = \mathbf{L}\mathbf{\Gamma}^{1/2}$ . De este análisis surge una gráfica en la que cada vector corresponde a un estudio; si estos vectores tienen longitudes similares y los ángulos entre ellos son pequeños, existe una estructura común (figura 1 de la aplicación).

## 2.3. Compromiso: búsqueda de un sistema de referencia común

Para encontrar un sistema de referencia común para todos los estudios, se construye un objeto compromiso denotado por  $\mathbf{R}_c$ , de la misma naturaleza de los objetos, y que además sea un buen resumen de estos. Se toma el objeto  $\mathbf{R}_c$  más correlacionado a todos los objetos  $\mathbf{R}_k$  en el sentido del producto escalar de Hilbert-Schmidt; este objeto es la media ponderada de los objetos  $\mathbf{R}_k$ , como lo describe (Lavit 1988), esto es:

$$\mathbf{R}_c = \sum_{k=1}^K \beta_k \mathbf{R}_k \quad (2)$$

Los valores de  $\beta_k$  están dados por

$$\beta_k = \frac{1}{\sqrt{\gamma_1}} \left( \sum_{\tau=1}^K \sqrt{S_{\tau\tau}} \right) l_{k1} \quad (3)$$

donde  $l_{k1}$  es la  $k$ -ésima componente de la primera componente principal de la matriz  $\mathbf{S}^R$  asociada al valor propio más grande  $\lambda_1$  y  $S_{\tau\tau}$  es la norma cuadrática del  $\tau$ -ésimo objeto. Los objetos ideales, cuyos productos escalares están contenidos en  $\mathbf{R}_c$ , se denominan *variables compromiso*.

La representación gráfica del compromiso se obtiene por la descomposición espectral de  $\mathbf{R}_c \mathbf{M}_f$ , siempre que  $\mathbf{M}_f$  sea definida positiva. Dado que esta matriz no es simétrica, ha de obtenerse la descomposición

$$\mathbf{M}_f^{1/2} \mathbf{R}_c \mathbf{M}_f^{1/2} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T \quad (4)$$

despejando  $\mathbf{R}_c$  en la ecuación anterior, se tiene

$$\mathbf{R}_c = \mathbf{M}_f^{-1/2} \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T \mathbf{M}_f^{-1/2} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \quad (5)$$

donde,

$$\mathbf{V} = \mathbf{M}_f^{-1/2} \mathbf{Q} \quad \text{y} \quad \mathbf{V}^T \mathbf{M}_f \mathbf{V} = \mathbf{I}$$

Las coordenadas de las variables compromiso se pueden encontrar como las filas de la matriz

$$\mathbf{B}^c = \mathbf{V} \mathbf{\Lambda}^{1/2} = \mathbf{M}_f^{-1/2} \mathbf{Q} \mathbf{\Lambda}^{1/2} \quad (6)$$

## 2.4. Representación Biplot inducida por el STATIS dual canónico: intraestructura

En la metodología clásica del STATIS dual se representan las trayectorias de las variables en todas las ocasiones, proyectando los objetos de cada ocasión sobre la estructura compromiso como

$$\mathbf{B}_k^c = \mathbf{R}_k \mathbf{M}_f \mathbf{V} \mathbf{\Lambda}^{-1/2} \quad (7)$$

para  $j = 1, \dots, J$  y  $k = 1, \dots, K$ . La fila  $j$  de la matriz  $\mathbf{B}_k^c$  representa la imagen de la variable  $j$  del estudio  $k$  sobre la estructura compromiso y la conexión de los puntos puede interpretarse como la trayectoria de dicha variable. Con esta representación, puede compararse la posición de las variables compromiso con relación a cada una de las ocasiones individuales. En general, las proyecciones tienen baja calidad de representación; por esta razón se han suprimido en las últimas versiones de software especializado, como el SPAD.

En este trabajo se propone una representación biplot del método basada en la proyección de las matrices de medias de los tratamientos en todas las ocasiones, sobre el espacio compromiso. El biplot proporciona una aproximación de las matrices de datos completas y una trayectoria para cada uno de los tratamientos

en cada ocasión sobre el subespacio generado por las variables compromiso, que pueden considerarse un espacio de representación común.

Para construir el biplot, se definen las siguientes matrices de concatenaciones:

$$\bar{\mathbf{X}}_b = \begin{bmatrix} \bar{\mathbf{X}}_1 \\ \vdots \\ \bar{\mathbf{X}}_k \end{bmatrix} \quad \text{y} \quad \bar{\mathbf{X}}_b^* = \begin{bmatrix} \sqrt{b_1} \mathbf{D}^{1/2} \bar{\mathbf{X}}_1 \\ \vdots \\ \sqrt{b_1} \mathbf{D}^{1/2} \bar{\mathbf{X}}_k \end{bmatrix}$$

Las dos matrices anteriores cumplen la siguiente relación:

$$\bar{\mathbf{X}}_b^* = \mathbf{D}_{\beta} \mathbf{D}_D \bar{\mathbf{X}}_b \quad (8)$$

con

$$\mathbf{D}_D = \begin{bmatrix} \mathbf{D}^{1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{D}^{1/2} \end{bmatrix} \quad \text{y} \quad \mathbf{D}_{\beta} = \begin{bmatrix} \mathbf{D}_{\beta_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{D}_{\beta_k} \end{bmatrix}$$

donde  $\mathbf{D}_{\beta_k} = \sqrt{\beta_k} I_G$  y  $\mathbf{D} = \text{diag}(I_1, \dots, I_G)$ . Puede demostrarse que la matriz compromiso puede escribirse como

$$\mathbf{R}_c = (\bar{\mathbf{X}}_b^*)^T \bar{\mathbf{X}}_b^* \quad (9)$$

En el análisis de la intraestructura en el método STATIS dual canónico, la matriz compromiso  $\mathbf{R}_c$  se diagonalizó con relación a la métrica dada a través de la matriz  $\mathbf{M}_f$ . Esta descomposición se obtiene de la ecuación (4) y puede escribirse como:

$$\mathbf{M}_f^{-1/2} (\bar{\mathbf{X}}_b^*)^T \bar{\mathbf{X}}_b^* \mathbf{M}_f^{-1/2} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$$

Esta expresión induce un biplot para  $\bar{\mathbf{X}}_b^* \mathbf{M}_f^{1/2}$ , que se obtiene realizando la descomposición

$$\bar{\mathbf{X}}_b^* \mathbf{M}_f^{1/2} = \mathbf{P} \mathbf{L}^{1/2} \mathbf{Q}^T$$

donde  $\mathbf{P} = \bar{\mathbf{X}}_b^* \mathbf{M}_f^{-1/2} \mathbf{Q}$ . Además, reemplazando en la expresión anterior  $\bar{\mathbf{X}}_b^*$  y despejando  $\bar{\mathbf{X}}_b$ , se tiene

$$\bar{\mathbf{X}}_b = (\mathbf{D}_D^{-1} \mathbf{D}_{\beta}^{-1} \mathbf{P}) (\mathbf{M}_f^{-1/2} \mathbf{Q}^{1/2})^T \quad (10)$$

La ecuación (10) puede escribirse como

$$\bar{\mathbf{X}}_b = \mathbf{A} (\mathbf{B}^c)^T$$

donde

$$\mathbf{A} = \mathbf{D}_D^{-1} \mathbf{D}_{\beta}^{-1} \mathbf{P} \quad \text{y} \quad \mathbf{B}^c = \mathbf{M}_f^{-1/2} \mathbf{Q} \mathbf{\Lambda}^{1/2}$$

Los marcadores  $\mathbf{B}^c$  son los utilizados para las variables compromiso del método STATIS dual canónico, dados en la ecuación (6).

Para la representación en dimensión reducida  $s$ , es suficiente tomar las primeras columnas de  $A$  y de  $B^c$ , denotadas por  $\mathbf{A}_s$  y  $(\mathbf{B}^c)_s$ ; así se obtiene una aproximación a bajo rango de  $\bar{\mathbf{X}}_b$ , que puede escribirse como

$$\tilde{\mathbf{X}}_b = (\mathbf{A}_s)(\mathbf{B}^c)_s^T \quad (11)$$

Si  $s = 2$ , en esta última ecuación, la representación en el plano es del tipo GH o RMP biplot, de acuerdo con la terminología utilizada por Gabriel (1971).

Por tanto, se construyó un biplot para los datos originales, donde los marcadores para las variables coinciden con las coordenadas de las variables compromiso del método STATIS dual canónico y se tienen marcadores para los grupos de todas las  $K$  tablas de datos.

Con la representación en el plano principal, en el espacio de las variables promedio de todos los estudios, pueden representarse las trayectorias de cada uno de los grupos; se trata de conectar en forma secuencial ascendente los puntos correspondientes a los marcadores de cada uno de estos. Esta forma de conectarlos permite clasificar estas trayectorias de acuerdo con su forma y relacionarlas con las variables compromiso que dan una interpretación a sus movimientos en el plano.

Una de las medidas de la calidad de representación global de  $\tilde{\mathbf{X}}_b$  es a través del porcentaje de relación:

$$\frac{\text{tr}(\tilde{\mathbf{X}}_b^T \tilde{\mathbf{X}}_b)}{\text{tr}(\bar{\mathbf{X}}_b^T \bar{\mathbf{X}}_b)} \times 100$$

que puede interpretarse como el porcentaje de variabilidad de los datos, explicada por los primeros ejes.

La calidad de representación definida en la ocasión  $k$  puede definirse como

$$\frac{\text{tr}(\tilde{\mathbf{X}}_{b(k)}^T \tilde{\mathbf{X}}_{b(k)})}{\text{tr}(\bar{\mathbf{X}}_b^T \bar{\mathbf{X}}_b)} \times 100$$

donde  $\tilde{\mathbf{X}}_{b(k)}$  es la parte de  $\tilde{\mathbf{X}}_b$  correspondiente a la  $k$ -ésima tabla.

Es posible un razonamiento análogo para definir la calidad de representación de una media de un grupo en una ocasión y para la trayectoria completa.

### 3. Diseño de medidas repetidas doblemente multivariante

El resultado de cada una de las unidades que configuran el diseño es una respuesta  $J$  dimensional en  $K$  ocasiones, puede describirse dentro del contexto del modelo lineal general como

$$\mathbf{X}_{(I \times KJ)} = \mathbf{Z}_{(I \times G)} \mathbf{B}_{(G \times KJ)} + \mathbf{U}_{(I \times KJ)} \quad (12)$$

donde  $\mathbf{X}$  es la matriz de respuestas de orden  $(I \times KJ)$ ,  $\mathbf{Z}$  la matriz de diseño de orden  $(I \times G)$  y con  $\text{rango}(\mathbf{X}) = G$ ,  $\mathbf{B}$  la matriz de parámetros no aleatorios (por

ejemplo medidas de la población de orden  $(G \times KJ)$  y  $\mathbf{U}$  la matriz de errores de orden  $(I \times KJ)$ . Las respuestas son ordenadas por columnas conforme a las variables dependientes y, dentro de estas, conforme a las ocasiones de medida.

Si se denota por  $\mathbf{u}'_i$  el vector de errores aleatorios correspondientes al sujeto  $i$ -ésimo, se supone que

$$\mathbf{u}'_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \quad (13)$$

La matriz  $\boldsymbol{\Sigma}$  es de orden  $(KJ \times KJ)$  y definida positiva. El hecho de que la matriz  $\boldsymbol{\Sigma}$  no dependa de  $i$ , indica que todos los vectores de errores aleatorios  $\mathbf{u}$  tienen la misma matriz  $\boldsymbol{\Sigma}$  y, por tanto, son homogéneas. Sin embargo, puede ocurrir que no todos los vectores de errores tengan la misma matriz  $\boldsymbol{\Sigma}$ , en cuyo caso se sugiere hacer el ajuste de los grados de libertad (Greenhouse & Geisser 1959, Huynh & Feldt 1976).

Para contrastar que  $h$  funciones estimables tienen un valor específico, la hipótesis nula  $H_0$  se expresa como

$$H_0 : \mathbf{C}'\mathbf{B}\mathbf{A} = 0 \quad (14)$$

La matriz  $\mathbf{C}'$  de dimensiones  $(h \times G)$  con  $\text{rango}(\mathbf{C}) = h$  se usa para definir  $h$  contrastes entre los grupos de tratamientos; matriz  $\mathbf{A}$  de dimensiones  $(KJ \times l)$  con  $\text{rango}(\mathbf{A}) = l$  se usa para definir  $l$  contrastes entre las diferentes ocasiones de medida para cada una de las  $J$  variables dependientes. A los primeros se les denomina contrastes entre sujetos; a los segundos, contrastes intra sujetos. La matriz  $\mathbf{B}$  de orden  $G \times KJ$  es la matriz de parámetros no aleatorios.

Los estadísticos usados para probar las hipótesis de interés son funciones de las raíces características de la matriz  $\mathbf{H}\mathbf{E}^{-1}$ , donde  $\mathbf{H}$  y  $\mathbf{E}$  son las matrices de sumas de cuadrados y productos cruzados correspondientes a la hipótesis y al error, y se obtienen como

$$\mathbf{H} = (\mathbf{C}'\hat{\mathbf{B}}\mathbf{A})'[\mathbf{C}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{C}]^{-1}(\mathbf{C}'\hat{\mathbf{B}}\mathbf{A}) \quad \text{y} \quad \mathbf{E} = \mathbf{A}'\mathbf{X}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{X}\mathbf{A}$$

Puede demostrarse (Vallejo et al. 1998) que

$$\mathbf{H} \sim W_t(\nu_h, \mathbf{A}'\boldsymbol{\Sigma}\mathbf{A}, (\mathbf{A}'\boldsymbol{\Sigma}\mathbf{A})^{-1}\boldsymbol{\Phi}) \quad \text{y} \quad \mathbf{E} \sim W_t(\nu_e, \mathbf{A}'\boldsymbol{\Sigma}\mathbf{A}, \mathbf{0})$$

donde  $W_t$  denota una distribución Wishart,  $t = KJ$ ,  $\nu_h = h$  y  $\nu_e = I - J$ .

$$\boldsymbol{\Phi} = (\mathbf{C}'\mathbf{B}\mathbf{A})'[\mathbf{C}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{C}]^{-1}(\mathbf{C}'\mathbf{B}\mathbf{A}) \quad \text{y} \quad \boldsymbol{\Sigma} = (I - J)^{-1}\mathbf{X}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{X}$$

La hipótesis de ausencia de interacción entre los grupos y las ocasiones de medida se prueba definiendo las matrices de contrastes  $\mathbf{C}'$  y  $\mathbf{A}$  como

$$\mathbf{C}'_{(G-1) \times G} = [\mathbf{I}_{I-G} \quad \cdots \quad -\mathbf{1}] \quad \text{y} \quad \mathbf{A}_{t \times (K-1)J} = \mathbf{I}_J \otimes \begin{bmatrix} \mathbf{I}_{K-1} \\ \vdots \\ -\mathbf{1} \end{bmatrix}$$

En esta última ecuación,  $\otimes$  denota el producto de Kronecker. Para probar la hipótesis nula multivariante de igualdad de las ocasiones de medida (modelo



no aditivo y medias no ponderadas), la matriz  $\mathbf{A}$  se define como en la ecuación anterior.

Ambas hipótesis se prueban usando la aproximación  $F$  de Rao (1951) a la Lambda de Wilks (Wilks 1932), como sigue:

$$F = \frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}} \left( \frac{\nu_2}{\nu_1} \right)$$

con  $s = [(m^2\nu_h^2 - 4)/(m^2 + \nu_h^2 - 5)]^{1/2}$ ,  $\nu_2 = \{[\nu_e - (m - \nu_h + 1)/2]s - (m\nu_h - 2)/2\}$ ,  $\nu_1 = m\nu_h$ ,  $m$  igual a la dimensión tanto de  $\mathbf{E}$  como de  $\mathbf{H}$  y  $\Lambda = |\mathbf{E}|/|\mathbf{E} + \mathbf{H}|$ . Cada una de estas hipótesis se rechaza con un nivel de significancia  $\alpha$  si  $F > F_{1-\alpha; \nu_1; \nu_2}$ , donde  $F_{1-\alpha; \nu_1; \nu_2}$  es el  $100(1 - \alpha)$ -ésimo percentil de la distribución  $F$  con  $\nu_1$  y  $\nu_2$  grados de libertad.

Otros estadísticos utilizados para contrastar las hipótesis anteriores son la traza de Pillai, la traza de Hotelling o la mayor raíz de Roy (Johnson & Wichern 2007).

## 4. Análisis de los resultados

### 4.1. Descripción de los datos

Durante 1998 y 1999 se realizó un estudio enmarcado dentro del proyecto SWALE (*Shallow wetland lake function and restoration in a changing European environment*) financiado por la Unión Europea (Stephen et al. 2004), cuyo objetivo central fue la determinación de los principales factores que regulan las condiciones ecológicas de los lagos poco profundos europeos. Se realizó un experimento variando tanto el estado trófico como la densidad poblacional de los depredadores en diferentes mesocosmos en lagos poco profundos de seis localidades de la Unión Europea (Suecia, Finlandia, Holanda, Reino Unido, León-España y Valencia-España). Se contrastan así dos explicaciones alternativas acerca de la dinámica trófica de estos sistemas: una supone que el factor principal es la disponibilidad de nutrientes para consumo de los depredadores; la otra sugiere que es el conjunto de poblaciones de depredadores el que controla las relaciones ecológicas entre las distintas comunidades en los lagos.

La serie de experimentos paralelos que se desarrollaron en los cinco países adscritos al proyecto SWALE, identificado como IME (*the international Mesocosm Experiment*), tenía como objetivo estudiar el impacto que provocan las variaciones en la densidad de peces y el exceso de nutrientes en las redes tróficas de los lagos y comparar este impacto en un gradiente longitudinal europeo. Los datos utilizados en este estudio se obtuvieron en el experimento realizado en el lago Sentiz en la localidad leonesa de Valdepolo durante el verano de 1998.

El estudio se realizó durante 10 semanas, del 9 de junio al 10 de agosto: 9 semanas con aplicación de tratamientos y una semana control. Para el desarrollo del mismo se utilizaron varios mesocosmos sometidos a la acción de dos factores:

- Gradiente trófico: adición semanal de diferentes concentraciones de nutrientes (nitrógeno y fósforo). Se establecieron cuatro niveles, uno de los cuales se mantuvo sin enriquecimiento nutricional.
- Densidad poblacional de ciprínidos: mantenimiento de diferentes densidades poblacionales de depredadores. Se consideraron tres niveles, el primero con ausencia de peces. La especie utilizada fue *Rutilus arcasii*, endémica de la Península Ibérica.

Los tratamientos se aplicaron combinando tres densidades de peces (0 g, 4 g y 20 g de masa fresca  $m^{-2}$ ) y cuatro niveles de nutrientes (N-P en mg L<sup>-1</sup>:0-0, 1-0.1, 5-0.5 y 10-1), es decir, se establecieron doce tratamientos diferentes, lo que supuso la instalación de 36 limnocorales. Cada limnocorral constituye una unidad experimental y se diseñó como un cilindro construido con polietileno, de 1 m de altura, rematado en sus extremos por aros de 1 m de diámetro, de lo que resulta un volumen aproximado de 0.8 m<sup>3</sup>. Los limnocorales se instalaron a lo largo de la laguna en zonas donde la profundidad, durante el periodo experimental, no fuese inferior a un 1 m, y la cobertura de los macrófitos (*Myriophyllum alterniflorum*) fuese elevada y homogénea.

Para efectos de este trabajo, se consideraron 4 tratamientos, teniendo en cuenta solo el efecto de la variación en la concentración de nutrientes, denotados por N0, N1, N2, N3 y 9 réplicas para cada uno de los tratamientos, ya que el efecto de los peces no resultó significativo. Se analizaron los datos de 4 semanas: la segunda, la cuarta, la sexta y la novena, denotadas por s2, s4, s6 y s9. Las variables fisicoquímicas medidas fueron 10, pero para los objetivos de este estudio las que van a considerarse son **PH**, alcalinidad total (**ALK**), fósforo total (**TP**), fósforo soluble (**SRP**), nitrógeno en nitrato (**NO3**), nitrógeno en amonio (**NH4**) y conductividad (**CON**).

Los datos inicialmente se centraron para cada variable. Para el análisis estadístico se usaron programas escritos en **MATLAB**, específicamente para este propósito, puesto que no existe software apropiado en los paquetes estándar.

El objetivo general puede plantearse como responder a la pregunta que ha cambiado en cuanto a las variables y a los efectos de los tratamientos, durante las 4 semanas del estudio, teniendo en cuenta que las mediciones se realizaron para 4 grupos, con 9 réplicas cada uno. Esto puede formularse como dos objetivos: investigar el cambio producido en las medidas de las 7 variables a lo largo de las 4 semanas, y garantizado el cumplimiento del primer objetivo, investigar si las posiciones de las medias de los grupos han cambiado a través de las 4 semanas y, si lo han hecho, buscar las variables que expliquen dicho cambio.

Para conseguir los objetivos mencionados, los resultados del análisis se obtuvieron por dos métodos: la metodología del **STATIS** dual canónico y las medidas repetidas doblemente multivariantes.

## 4.2. Resultados obtenidos a través del STATIS dual Canónico

Para el análisis, se tendrán en cuenta las etapas descritas para el STATIS dual canónico, esto es, análisis de la interestructura, sistema de referencia común y biplot inducido por el STATIS dual canónico. La transformación utilizada para los datos fue restar media de las columnas; la media de las matrices dentro en todas las ocasiones se utilizó como métrica para las filas. Para el análisis estadístico se usaron programas escritos en MATLAB para este propósito puesto que no existe un software apropiado en los paquetes estándar.

El primer objetivo del análisis se planteó como investigar el cambio producido en las medidas de las 7 variables a lo largo de las 4 semanas; este objetivo se plantea en el contexto del STATIS dual canónico como analizar la existencia de un sistema de covariación común entre las variables. El segundo objetivo planteado es investigar si las posiciones de las medias de los grupos han cambiado a través de las 4 semanas y, si lo han hecho, buscar las variables que expliquen dicho cambio. Para conseguir este objetivo, en el contexto del método, se encuentra un sistema de referencia común y se representan conjuntamente las variables promedio y las trayectorias de las medias de los grupos utilizando el biplot inducido.

### 4.2.1. Análisis de la interestructura

El análisis de componentes principales de la matriz  $\mathbf{S}^R$ , de los coeficientes  $RV$  y de la representación en el plano principal de los puntos correspondientes a las 4 semanas (denotados como s2, s4, s6, s9) proporciona la representación euclídea de la interestructura y permite decidir si existe o no una estructura de covariación común en los datos de las 4 semanas. En la tabla 1 se tiene la matriz  $\mathbf{S}^R$ ; en la figura 1 se muestra representación de las 4 semanas en el primer plano principal de la interestructura.

TABLA 1: Coeficientes de correlación  $RV$  entre las semanas.

	s2	s4	s6	s9
s2	1.000	0.813	0.583	0.431
s4	0.813	1.000	0.755	0.640
s6	0.583	0.755	1.000	0.816
s9	0.431	0.640	0.816	1.000

El porcentaje de inercia explicada por el eje 1 es 75.71%, muy alto comparado con el eje 2, que es 16.69%, lo cual muestra que al analizar las diferentes ocasiones hay una parte común en el primer eje de la interestructura. La correlación de los puntos correspondientes a las 4 semanas, con los ejes 1 y 2 que aparecen en la tabla 2, muestra que todas las ocasiones tienen correlación más alta con el eje 1. Se observa además que la longitud de los vectores es similar y los ángulos entre estos vectores son pequeños. Puede concluirse que existe una estructura de covariación común entre las variables de las 4 semanas, esto es, se cumple el primer objetivo.

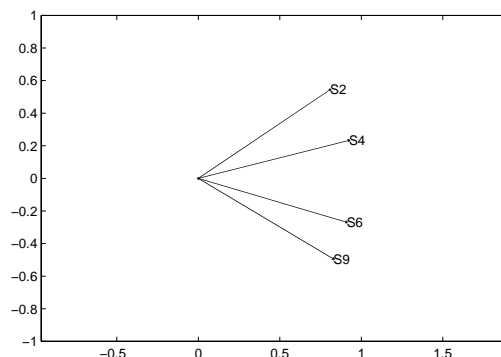


FIGURA 1: Representación de las 4 ocasiones en el primer plano principal de la interestructura.

TABLA 2: Coordenadas y calidad de representación por semana en el primer plano principal de la interestructura.

Semanas	Coordenadas		Calidad de representación	
	Eje 1	Eje 2	Eje 1	Eje 2
s2	0.809	0.544	65.443	29.609
s4	0.925	0.233	85.562	5.432
s6	0.911	-0.269	82.916	7.239
s9	0.830	-0.495	68.927	24.484

#### 4.2.2. Sistema de referencia común: representación de la estructura compromiso

La estructura compromiso es un sistema de referencia común que permite representar las 4 ocasiones, que se obtuvieron utilizando los resultados descritos en las ecuaciones (2) y (3), ésta se representa en la figura 2. La inercia explicada en el primer plano principal es 88.17%. Todas las variables tienen alta calidad de representación en dicho plano y toman valores en el intervalo [84.61, 97.215] (tabla 2). La calidad de representación se interpreta como la parte de la variación total de cada variable que es explicada por cada eje, y se calcula como el coseno al cuadrado del ángulo del vector correspondiente a cada variable y los ejes factoriales. En el eje 1 se observa que este tiene alta correlación positiva con las variables CON, PH, NO<sub>3</sub>, TP, SRP y NH<sub>4</sub>; por tanto, puede interpretarse como un gradiente de la eutrofización del sistema, esto es, se espera que si se tiene alta concentración de nutrientes y alta presencia de algas, su ubicación se tenga en el lado izquierdo del plano. El eje 2 tiene correlaciones altas con ALK y mediana con PH, que son negativas, y con NO<sub>3</sub> que es positiva; este eje puede interpretarse como un gradiente de la dureza de los sistemas generados en cada unidad experimental.

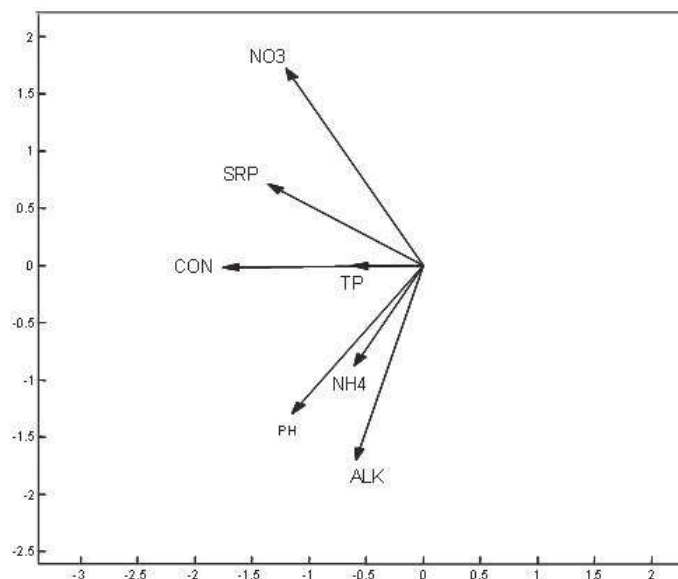


FIGURA 2: Representación de las variables compromiso en el plano principal del sistema de referencia común del STATIS dual canónico.

TABLA 3: Coordenadas y calidad de representación de las variables compromiso para los dos primeros ejes.

Variable	Coordenadas		Contribuciones por ejes (%)		Contribuciones acumuladas (%)
	Eje 1	Eje 2	Eje 1	Eje 2	Ejes 1-2
PH	-4.002	-2.469	<b>61.289</b>	23.321	84.610
ALK	-2.060	-3.240	27.211	<b>67.349</b>	94.561
NH4	-2.096	-1.669	<b>44.086</b>	27.942	72.028
NO3	-4.199	3.287	<b>54.491</b>	33.389	87.880
SRP	-4.757	1.360	<b>79.621</b>	6.512	86.133
TP	-2.199	-0.011	<b>69.604</b>	0.002	69.606
CON	-6.142	-0.032	<b>97.212</b>	0.003	97.215

#### 4.2.3. Biplot inducido por el STATIS dual canónico

En la figura 3 se muestra la representación biplot con los datos originales para el STATIS dual canónico, en el que, sobre el espacio de las variables canónicas, las variables compromiso o promedio son representadas como vectores y para las trayectorias de las medias de los tratamientos en las 4 semanas, el nombre del tratamiento es el correspondiente al inicio de la trayectoria. El porcentaje de la inercia total explicada en el primer plano principal es 88.17%.

En la tabla 2, se tiene la calidad de representación, en el primer plano factorial, de las variables compromiso, que coincide con las analizadas en la representación compromiso. En la tabla 3, se observaron las coordenadas y la calidad de representación de las variables compromiso en los dos primeros ejes. En la tabla 4, se tiene

la calidad de representación en el primer plano factorial de las medias en cada semana; en general, estas son altas para todos los tratamientos. En la tabla 4, se tienen las calidades de representación de las trayectorias de los tratamientos en el primer plano factorial del biplot; puede verse que están por encima de 75.011, es decir, pueden considerarse bien representadas.

Las trayectorias de las medias correspondientes a los nutrientes N0 y N1 tienen las siguientes características: están localizadas en la parte derecha y al centro del plano. De acuerdo con esta ubicación, se espera que estos sistemas tengan bajos valores en las variables CON, PH, NO<sub>3</sub>, TP, SRP y NH<sub>4</sub>, lo cual significa que son sistemas poco eutrofizados. En cambio, la trayectoria de las medias correspondientes al nutriente N3 está a la izquierda del gráfico y con tendencia hacia la parte superior; de acuerdo con esta localización, se espera que asuman valores altos para las variables CON, PH, NO<sub>3</sub>, TP, SRP y NH<sub>4</sub>. Esto significa que se trata de sistemas altamente eutrofizados, y que van en aumento con el tiempo. La trayectoria de las medias correspondientes al nutriente N2 está ubicada en la parte central e inferior, con valores altos de ALK, PH y NH<sub>4</sub>, y bajo NO<sub>3</sub>.

En general, el eje 1 puede interpretarse como un gradiente de la cantidad de nutriente añadido. A la derecha, se tienen las trayectorias correspondientes a bajos nutrientes; en el lado opuesto, las correspondientes a altas cantidades de nutrientes. El eje 2 caracteriza la dureza que se presenta en los diferentes sistemas obtenidos en los distintos limnocorrales.

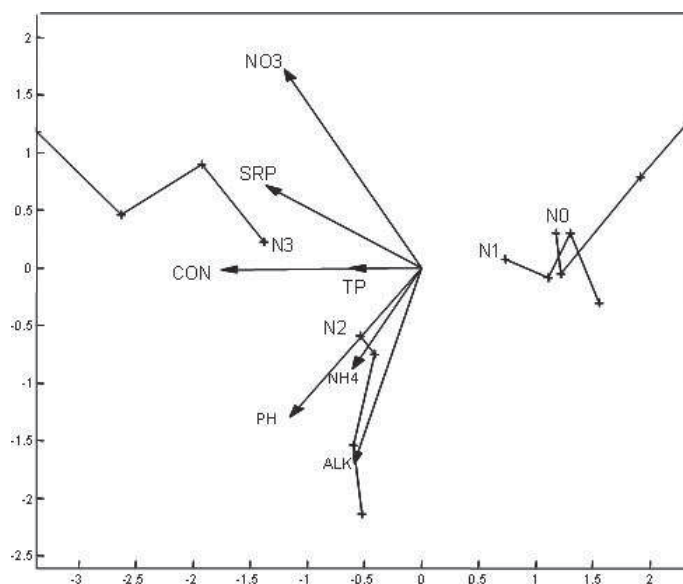


FIGURA 3: Representación de los datos originales en el Biplot inducido por el STATIS dual canónico.

TABLA 4: Calidad de representación de las medias de los tratamientos en cada ocasión en el biplot del STATIS dual canónico.

Tratamiento	Semana 2		Semana 4		Semana 6		Semana 9	
	Eje 1	Ejes 1-2	Eje 1	Ejes 1-2	Eje 1	Ejes 1-2	Eje 1	Ejes 1-2
N0	47.531	50.595	83.324	83.492	82.500	96.373	73.276	94.817
N1	61.858	62.453	86.419	86.943	79.241	83.371	65.918	68.506
N2	29.104	64.766	17.982	79.281	12.117	93.001	5.162	92.699
N3	70.936	72.801	75.705	92.112	93.780	96.621	83.525	93.903

TABLA 5: Calidad de representación para las trayectorias de las medias (medias  $\times$  ejes).

Tratamiento	Eje 1	Ejes 1-2
N0	72.307	86.211
N1	72.595	75.011
N2	10.722	88.805
N3	83.647	92.316

### 4.3. Resultados obtenidos a través de las medidas repetidas doblemente multivariantes

Los resultados se obtuvieron a través del SPSS, por el procedimiento GLM, medidas repetidas. Se incluyen en el análisis las mediciones de las 7 variables en los 4 grupos con 9 réplicas por grupo, obtenidas en las 4 semanas. Los resultados se obtuvieron a través del SPSS, por el procedimiento GLM, medidas repetidas. Se incluyen en el análisis las mediciones de las 7 variables en los 4 grupos, con 9 réplicas por grupo obtenidas en las 4 semanas. Las pruebas realizadas hacen referencia a los efectos intra sujetos y entre sujetos. Con la primera para las semanas y para las semanas por grupos y la segunda para los tratamientos. El diseño utilizado fue: intersección + grupo y para el efecto intra sujetos se utilizó las semanas.

#### 4.3.1. Pruebas de efectos intra sujetos

Con estas pruebas se contrastan las siguientes hipótesis multivariadas: los vectores de medidas entre las diferentes semanas son iguales y existe interacción entre las semanas y los tratamientos. Considerando el valor  $p$  de todos los estadísticos de la tabla 6, puede concluirse que, globalmente, las medidas de las variables son diferentes durante las 4 semanas y que no existe interacción entre las semanas y los tratamientos. Los grados de libertad son corregidos, puesto que por la prueba de esfericidad de Mauchly, que se tiene en la tabla 8, se contrasta la hipótesis nula de que la matriz de covarianza error de las variables dependientes transformadas es proporcional a una matriz identidad. Esta hipótesis se rechaza. Las pruebas corregidas se muestran en la tabla 6.

#### 4.3.2. Prueba de efectos entre sujetos

Con esta prueba se contrasta si los vectores de medias entre las diferentes tratamientos son iguales, pero considerando las variables promedio de todas las

TABLA 6: Prueba de efectos intra sujetos. (a) El estadístico es un límite superior para la  $F$  el cual ofrece un límite inferior para el nivel de significación, las pruebas se basan en las variables promediadas.

Efecto intra sujetos		Valor	$F$	Gl de la hipótesis	Gl del error	Significación
Semana	Traza de Pillai	1.104	7.649	21	276	0.000
	Lambda de Wilks	0.126	13.024	21	258	0.000
	Traza de Hotelling	5.180	21.870	21	266	0.000
	Raíz mayor de Roy	4.839	63.597(a)	7	92	0.000
Semana*grupo	Traza de Pillai	1.381	2.622	63	672	0.000
	Lambda de Wilks	0.149	3.276	63	512	0.000
	Traza de Hotelling	2.886	4.044	63	618	0.000
	Raíz mayor de Roy	1.902	20.285(a)	9	96	0.000

semanas. Puesto que el nivel de significación en los diferentes test  $< 0.05$  (tabla 7), puede concluirse que existe una diferencia global entre los vectores de medias de los grupos.

TABLA 7: Prueba de efectos entre sujetos: (a) Estadístico exacto, (b) El estadístico es un límite superior para  $F$ , el cual ofrece un límite inferior para el nivel de significación.

Efecto entre sujetos		Valor	$F$	Gl de la hipótesis	Gl del error	Significación
Intersección	Traza de Pillai	1.000	9831.990(a)	7	26	0.00
	Lambda de Wilks	0.000	9831.990(a)	7	26	0.00
	Traza de Hotelling	2647.070	9831.990(a)	7	26	0.00
	Raíz mayor de Roy	2647.070	9831.990(a)	7	26	0.00
Grupo	Traza de Pillai	1.862	6.540	21	84	0.00
	Lambda de Wilks	0.016	11.430	21	75	0.00
	Traza de Hotelling	14.291	16.786	21	74	0.00
	Raíz mayor de Roy	10.820	43.280(b)	7	28	0.00

Los análisis de varianza para cada variable por separado son una herramienta adicional que permite, desde este punto de vista, identificar las variables responsables de las diferencias entre los grupos. En este caso, como se observa en la tabla 9, existe una diferencia significativa entre los grupos para los análisis con todas las variables, puesto que los valores  $p$  de la prueba son mucho menores que 0.01.

TABLA 8: Prueba de esfericidad de Mauchly.

Efecto intra sujetos	Medida	W de Mauchly	Chi-Cuadrado aproximada	Gl	Significación
Semana	PH	0.791	7.221	5	0.205
	ALK	0.205	48.758	5	0.000
	NH4	0.112	67.380	5	0.000
	NO3	0.157	56.790	5	0.000
	SRP	0.025	113.327	5	0.000
	TP	0.023	115.572	5	0.000
	CON	0.029	108.485	5	0.000



TABLA 9: Pruebas de los efectos entre sujetos por variable.

Fuente	Medida	Suma de cuadrados tipo III	gl	Media Cuadrática	$F$	Significación
Intersección	PH	11820.582	1	11820.582	19621.722	0.000
	ALK	827.425	1	827.425	1175.564	0.000
	NH4	48485713.238	1	48485713.238	70.467	0.000
	NO3	3403.264	1	3403.264	123.121	0.000
	SRP	21683659.687	1	21683659.687	105.068	0.000
	TP	74783075.611	1	74783075.611	93.968	0.000
	CON	25796241.000	1	25796241.000	1316.195	0.000
Grupo	PH	11.856	3	3.952	6.560	0.001
	ALK	71.320	3	23.773	33.776	0.000
	NH4	19659517.645	3	6553172.548	9524.000	0.000
	NO3	4494.352	3	1498.117	54198.000	0.000
	SRP	24987811.396	3	8329270.465	40.359	0.000
	TP	41482588.797	3	13827529.599	17.375	0.000
	CON	2823054.722	3	941018.241	48.013	0.000
Error	PH	19.278	32	0.602		
	ALK	22.523	32	0.704		
	NH4	22018025.944	32	688063.311		
	NO3	884.535	32	27.642		
	SRP	6604092.736	32	206377.898		
	TP	25466691.271	32	795834.102		
	CON	627171.278	32	19599.102		

El primer objetivo de este análisis se planteó como investigar el cambio producido en las medidas de las 7 variables a lo largo de 4 semanas. Al respecto, en el contexto del modelo, puede concluirse que utilizando la prueba de los efectos intra sujetos, se produjo un cambio de las variables en las 4 semanas. Con respecto al segundo objetivo, que se formuló como investigar si las posiciones de las medias de los grupos han cambiado a través de las 4 semanas y, si lo han hecho, buscar las variables que expliquen dicho cambio, se puede afirmar, utilizando la prueba de efectos entre los sujetos, que existen diferencias entre las medias de los tratamientos globales; también es posible analizar este efecto por variables, como se observa en la tabla 9, donde para todas las variables promedio puede inferirse que hay diferencias entre los tratamientos.

## 5. Conclusiones

El método propuesto, STATIS dual canónico, puede verse como una generalización del STATIS dual clásico en el sentido de que utiliza métricas diferentes a la identidad. Los estudios en cada ocasión son las ternas  $(\bar{X}_k, S^{-1}, D)$  para  $k = 1, 2, \dots, K$ , y tiene por tanto las siguientes propiedades:

- El producto escalar de Hilbert Schmidt tiene en cuenta métricas diferentes a la identidad, como se definió en la sección 2.1.

- El análisis de cada estudio lleva a obtener las variables canónicas; no las componentes principales (Takane & Hunter 2001).
- La estructura compromiso puede interpretarse como una estructura de variables canónicas.
- En el biplot para los datos originales, inducido por el método, se representan las variables canónicas compromiso y las trayectorias de las medias en todas las ocasiones.
- Una de las propiedades de este método, mirado desde el punto de vista descriptivo, es que permite obtener una representación gráfica de la evolución de las medias de los grupos en el espacio de las variables canónicas comunes.
- Los únicos supuestos para obtener esta representación son: puede concluirse la existencia de una estructura de variables canónicas comunes, y es válido construir una matriz  $S = \frac{1}{K} \sum_{k=1}^K S_k$ , donde  $S_k$  representa las matrices de covarianza dentro de los grupos en cada estudio para  $k = 1, \dots, K$ .

En el análisis a través del modelo de medidas repetidas doblemente multivariantes, se consideran las siguientes características importantes: se plantea en el contexto del modelo lineal general y permite contrastar las siguientes hipótesis:

- El efecto del tiempo. Para hallar estos contrastes, se crean los perfiles de las  $JK$  variables, esto es, del total de variables medidas, considerando las de cada tiempo como variables diferentes.
- El efecto del tiempo para cada tratamiento. Estos contrastes resultan de crear los perfiles de las  $JK$  variables por tratamiento.
- El efecto de los tratamientos para las variables promediadas entre las ocasiones.
- Para cada una de las variables, realizar diferentes contrastes intra sujetos, por ejemplo los contrastes polinómicos, para analizar la tendencia de las medias a través de las diferentes ocasiones.

Los supuestos del modelo de medidas repetidas doblemente multivariantes son mucho más exigentes, puesto que exigen multinormalidad de los errores aleatorios para cada sujeto y homogeneidad de las matrices de covarianzas. En caso de no cumplirse esta segunda restricción, se sugiere hacer un ajuste de los grados de libertad (Greenhouse & Geisser 1959, Huynh & Feldt 1976). Además, en el contexto de este análisis, es posible estudiar si existe un comportamiento de los datos promedio a través de las diferentes ocasiones. En general, aunque en ambos análisis se estudia el comportamiento de los efectos de los tratamientos, para las variables medidas, estos se diferencian en que el STATIS dual canónico se puede considerar fundamentalmente como un método descriptivo y las medidas repetidas doblemente multivariantes como un método inferencial; por tanto, los supuestos son más exigentes para este último. Es importante resaltar que el STATIS dual

canónico, por tener una representación conjunta de las variables promedio y las trayectorias de las medias de los tratamientos en el espacio de las variables canónicas, es una herramienta gráfica de gran utilidad para analizar la evolución de las diferencias en los efectos de los tratamientos y para identificar las variables que pueden explicarlas. Esto no es posible con el modelo de medidas repetidas doblemente multivariantes.

[Recibido: noviembre de 2007 — Aceptado: octubre de 2008]

## Referencias

- Abdí, H. (2007), *RV Coefficient and Congruent Coefficient*, Encyclopedia of Measurement and Statistics.
- Escoufier, Y. (1973), 'Le traitement des variables vectorielles', *Biometrics* **29**, 751–760.
- Escoufier, Y. (1980), 'L'e analyse conjointe de plusieurs matrices de données', *Biométrie et temps* **58**, 59–76.
- Gabriel, K. R. (1971), 'The Biplot Graphic Display of Matrices with Application to Principal Component Analysis', *Biometrika* **58**, 453–467.
- Greenhouse, S. W. & Geisser, S. (1959), 'On Methods in the Analysis of Profile Data', *Psychometrika* **24**, 95–112.
- Huynh, H. & Feldt, L. S. (1976), 'Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split-Plot Designs', *Journal of Educational Statistics* **1**, 1582–1589.
- Johnson, R. A. & Wichern, D. W. (2007), *Applied Multivariate Statistical Analysis*, sixth edn, Pearson Prentice Hall, Upper Saddle River, NJ.
- Lavit, C. (1988), *Analyse conjointe de tableaux quantitatifs*, Masson, Paris, France.
- L'Hermier des Plantes, H. (1976), *Structuration des tableaux à trois indices de la statistique*, sixth edn, Thèse de 3ème cycle, Université de Montpellier.
- Rao, C. R. (1951), 'An Asymptotic Expansion of the Distribution of Wilks'  $\Lambda$  criterion', *Bull. Inst. Internat. Statist.* **23**(part II), 177–180.
- Stephen, D., Balayla, D. M., Bécares, E., Collings, S. E., Fernández-Aláez, C., Fernández-Aláez, M., Ferriol, M. C., García, P., Gomá, J., Gyllström, M., Hansson, L. A., Hietala, J., Kairesalo, T., Miracle, M. R., Romo, S., Rueda, J., Stahl-Delbanco, A., Svensson, M., Vakkilainen, K., Valentín, M., Van de Bund, W. J., Van Donk, E., Vicente, E., Villena, M. J. & Moss, B. (2004), 'Continental-Scale Patterns of Nutrient and Fish Effects on Shallow Lakes: Introduction to a Pan-European Mesocosm Experiment', *Freshwater Biology* **49**(12), 1517–1524.

- Takane, Y. & Hunter, M. A. (2001), 'Constrained Principal Component Analysis: A Comprehensive Theory', *Appl. Algebra Engrg. Comm. Comput.* **12**(5), 391–419.
- Vallejo-Arboleda, A., Vicente-Villardón, J. L. & Galindo-Villardón, M. P. (2007), 'Canonical STATIS: Biplot Analysis of Multi-Table Group Structured Data Based on Stasis-Act Methodology', *Comput. Statist. Data Anal.* **51**(9), 4193–4205.
- Vallejo, G., Fidalgo, A. M. & Fernández, P. (1998), 'Efectos de la no esfericidad en el análisis de diseños multivariados de medidas repetidas', *Anales de psicología* **14**(2), 249–268.
- Wilks, S. S. (1932), 'Certain Generalizations in the Analysis of Variance', *Biometrika* **24**, 471–494.