

## Análisis multivariante: Un recorrido por las técnicas de reducción de dimensiones

## Multivariate analysis: A tour through dimension reduction methods

Mariela González-Narváez, Omar Ruiz-Barzola y Ana Nieto-Librero

Recepción: 22/07/2020 Aceptación: 30/09/2020 Publicación: 20/10/2020

**Abstract** This article integrates analyzes, and applies the main dimension reduction techniques, developed to analyze multivariate data. The proposed objective of this work was to bring out the wide variety of options and highlight the advantages they offer according to the type of analysis intended and the nature of the data under analysis. The article covers the method for studying a two and three dimensional matrix, the method for studying relationships between two and three dimensional arrays; in addition, the graphs resulting from applying each method are presented, environmental and biological data taken by INOCAR in Esmeraldas, Manta, La Libertad and Puerto Bolívar were used.

**Keywords** dimensionality reduction, INOCAR, multivariate analysis.

**Resumen** En este artículo se integra, analiza y aplican las principales técnicas de reducción de dimensiones, desarrolladas para analizar datos multivariantes. El objetivo propuesto de este trabajo fue poner de manifiesto la amplia variedad de opciones y resaltar de ellas las ventajas que ofrecen conforme al tipo de análisis que se pretenda y la naturaleza de los datos objeto del análisis. El artículo cubre el método

---

Mariela González-Narváez, Ph.D.(c) en Estadística Multivariante Aplicada  
Docente, ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias de la Vida (FCV), Campus Gustavo Galindo Km. 30.5 Vía Perimetral P.O. Box 09-01-5863, Guayaquil, Ecuador, e-mail: [margonzal@espol.edu.ec](mailto:margonzal@espol.edu.ec),

 <https://orcid.org/0000-0002-8747-5257>

Omar Ruiz-Barzola, Ph.D. en Estadística y Optimización  
Docente, ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias de la Vida (FCV), Facultad de Ciencias Naturales y Matemáticas (FCNM), Campus Gustavo Galindo Km. 30.5 Vía Perimetral P.O. Box 09-01-5863, Guayaquil, Ecuador, e-mail: [oruiz@espol.edu.ec](mailto:oruiz@espol.edu.ec),

 <https://orcid.org/0000-0001-8206-1744>

Ana Belén Nieto-Librero, Ph.D. en Matemáticas  
Profesora Ayudante, Universidad de Salamanca, Facultad de Medicina, Dpto. de Estadística, Campus Miguel de Unamuno, Salamanca, España, e-mail: [ananieto@usal.es](mailto:ananieto@usal.es),

 <https://orcid.org/0000-0001-6643-247X>

para el estudio de una matriz de dos y tres dimensiones, el método para estudiar relaciones entre matrices de dos y tres dimensiones; además se presentan los gráficos resultantes de aplicar cada método, para ello fueron empleados datos ambientales y biológicos tomados por el INOCAR en: Esmeraldas, Manta, La Libertad y Puerto Bolívar.

**Palabras Claves** análisis multivariante, INOCAR, reducción de dimensiones.

## 1 Introducción

Consideremos inicialmente la descripción precisa del análisis multivariante de datos dado por Cuadras (2019): “El análisis multivariante forma parte de la estadística y del análisis de datos que estudia, analiza, representa e interpreta los datos que resultan de observar más de una variable sobre una muestra de individuos”(p.11). Luego añade: “La información estadística en análisis multivariante es de carácter multidimensional; por lo tanto, la geometría, el cálculo matricial y las distribuciones multivariantes juegan un papel fundamental”(p.11).

Los datos estructurados, generalmente están contenidos en una matriz multivariante, de los cuales se desea representar el grado de discrepancia entre individuos o entre variables, a través de medidas como correlaciones, distancias o similitudes.

Por otro lado, las técnicas de reducción de dimensiones son aplicadas en el análisis multivariante para representar un espacio de  $p$  variables a través de un espacio  $k$  dimensional, con  $k < p$ , de tal forma que puedan ser visualizadas todas las variables a la vez, así como sus relaciones, grado de aportación, etc.

Por este motivo, el presente trabajo hace un recorrido metodológico por las principales técnicas de reducción de dimensiones, desarrolladas para analizar datos multivariantes.

El recorrido inicia con el método para el estudio de una matriz de dos dimensiones; continúa con el método para estudiar relaciones entre matrices de dos dimensiones; luego se revisa la metodología para el análisis de matrices de tres dimensiones y finalmente se revisa la metodología desarrollada para el estudio de la relación entre matrices de tres dimensiones.

Para cada método, se presenta un gráfico resultante de su aplicación en datos ambientales y biológicos (orden de especies de fitoplancton-algas marinas). Los datos corresponden a un muestreo que realizó INOCAR en cuatro estaciones del perfil costero de Ecuador continental: Esmeraldas, Manta, La Libertad y Puerto Bolívar, en cada una de ellas se analizó la parte ambiental y biológica (especies), los datos fueron estructurados en dos matrices respectivamente y en sus filas se registraron los valores promedios (0-20m) de las mediciones realizadas mensualmente (excepto febrero) (02, 03, ..., 12) en cada estación de muestreo.

La matriz ambiental contiene en sus columnas 7 variables físico-químicas: temperatura (T), oxígeno disuelto (OD), nitrito (NO<sub>2</sub>), nitrato (NO<sub>3</sub>), fosfato (PO<sub>4</sub>), silicato (SiO<sub>4</sub>) y salinidad (S), mientras que la matriz de especies contiene la identificación de 11 órdenes del fitoplancton (O1, O2, ..., O11). Los meses permiten

identificar la época climática de invierno-lluviosa (diciembre a mayo) y verano-seca (junio a noviembre).

## 2 Análisis Multivariante de Matrices de dos dimensiones o Tablas de Dos-Vías

Los métodos clásicos de análisis multivariante como el Análisis de Componentes Principales (PCA), el Análisis de Correspondencias (CA) o los métodos Biplot, se centran en estudiar el comportamiento de los datos que están contenidos en una matriz de dos dimensiones  $X_{[I,P]}$  (tablas de 2-vías), cuya estructura está formada por  $I$  filas (individuos) y  $P$  columnas (variables), tal como se indica en (1) y se representa en la Figura A 1.a.

$$X_{[I,P]} = \{x_{ij} \mid i = 1, \dots, I; j = 1, \dots, P\} \quad (1)$$

### 2.1 Análisis de Componentes Principales - ACP

Es un método multivariante que permite reducir la dimensionalidad de los datos, reteniendo la mayor cantidad posible de la información (variabilidad). Este método fue desarrollado por Pearson (1901) y mejorado con aportaciones de Hotelling (1933, 1936). Su enfoque está orientado en la ordenación de los individuos (filas) de una matriz en un espacio de baja dimensión. Su análisis, está basado en la Descomposición de Valores Singulares (DVS) a partir de una matriz de covarianzas (matriz centrada por variables) o de correlaciones (matriz centrada y escalada por variables).

$$X = U\Sigma V^T \quad (2)$$

La DVS permite factorizar la matriz  $X$  rectangular ( $I \geq J$ ), descrita en (2), en una terna de matrices  $U\Sigma V$ , tales que  $U \in \mathbb{R}^{I \times r}$ ,  $V \in \mathbb{R}^{P \times r}$  y  $\Sigma \in \mathbb{R}^{r \times r}$  que es una matriz rectangular diagonal de rango  $r \leq \min\{I, P\}$ . Las columnas de las matrices  $U$  y  $V$  son ortogonales y contienen a los vectores singulares izquierda (filas) y derecha (columnas) de la matriz  $X$  respectivamente, mientras que la diagonal principal de la matriz  $\Sigma$  contiene a los valores singulares  $\sigma_i$  de  $X$  tal que  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$  (positivos). Simultáneamente, las columnas de  $U$  son los vectores propios de  $XX^T$  y las columnas de  $V$  los de  $X^T X$ , a ambos productos de matrices están asociados los valores propios  $\sigma_i^2$ .

*Ambiente:* El componente 1 clasifica las observaciones según la época climática: verano (derecho) e invierno (izquierdo), el componente 2 según la localización geográfica a las estaciones del sector norte (cuadrante superior) y las localizadas en el sector sur (cuadrante inferior). Ordena de manera conjunta las mediciones de verano de Esmeraldas (estación con temperaturas más cálidas en el perfil) y Man-

ta y lo aproxima al vector de temperatura. Explicación de la variabilidad cercana al 76.5 %. Las variables mostraron una relación directa entre la temperatura y el oxígeno e inversa con las demás, Figura 1(a.1). *Especies*: La componente clasifica las filas según la época verano (izquierda) e invierno (derecha), mientras que la componente 2 parece clasificarlos según el tipo de estación oceánicas (superior) y estuarios (inferior). Ubica la abundancia de la mayoría de las especies con los meses de invierno. Explicación de la variabilidad aproximadamente del 61.5 %. Las variables mostraron una relación entre la abundancia de las especies del orden de Gymnodiniales (O10) y Cyclotrichiida (O11), Figura 1(a.2).

## 2.2 Biplots

El método Biplot nace de la aportación de Gabriel (1971) y consiste en la representación gráfica de la estructura de grandes matrices de datos. Si una matriz de partida -con individuos en las filas y variables en las columnas- es de rango superior, el método la representa de manera aproximada como una matriz de rango 2. El Biplot se caracteriza por mostrar distancias o agrupaciones entre individuos y correlaciones entre variables.

El Biplot se fundamenta en la DVS de la matriz de partida, realiza una factorización en matrices de marcadores filas y de marcadores columnas. Se compone de dos métodos: GH-Biplot y JK-Biplot, el primero se caracteriza por representar de la mejor manera posible las variables, pero con una menor calidad de representación de los individuos y el segundo se caracteriza por representar de mejor manera los individuos, pero con una menor calidad de representación de las variables.

Sea  $X_{[I,P]}$  la matriz definida en (1), de rango  $r$ , que puede ser factorizada de la siguiente manera:

$$X \cong GH^T \quad (3)$$

donde  $G$  es una matriz de dimensión  $(I \times r)$  y  $H$  de dimensión  $(P \times r)$ , donde cada elemento  $x_{ij} = g_i^T h_j$  puede ser representado por el producto de la  $i$ -ésima fila de  $G$  y la  $j$ -ésima columna de  $H^T$ . Los vectores  $g_i$  son considerados los efectos filas y los  $h_j$  los efectos columna. Sea  $X$  tal como se definió en (2); se definen las matrices  $JK$  y  $GH$  como en (4) y (5), respectivamente

$$\mathbf{JK}: \quad J = U\Sigma \quad \text{y} \quad K = V \quad (4)$$

$$\mathbf{GH}: \quad G = U \quad \text{y} \quad H = V\Sigma \quad (5)$$

A partir de esta metodología Galindo-Villardón (1986) realiza una aportación a los métodos Biplot, al proponer una representación simultánea con máxima calidad de representación para los marcadores filas  $j_1, \dots, j_I$  y los marcadores columnas  $h_1, \dots, h_P$  en un mismo sistema de referencia de ejes factoriales; este método se lo conoce como HJ-Biplot, tal que:

$$\mathbf{HJ} : H = U\Sigma \quad \text{y} \quad J = V\Sigma \quad (6)$$

siendo  $\sigma$  el valor singular de la matriz  $\Sigma$ .

A través de métricas se pueden interpretar relaciones entre los ejes factoriales con las variables y con los individuos (Galindo-Villardón y Cuadras, 1986). Además, de contribuciones relativas del factor al elemento y del elemento al factor, así como la calidad de representación (QLR, por sus siglas en inglés) de filas y columnas, tal como se muestra en la Tabla 1.

**Tabla 1:** Contribuciones del HJ-Biplot

Contribución	Ecuación
Relativa a la traza del elemento $i$ -ésimo (fila), representa la variabilidad total explicada por el individuo.	$CRT_i = \frac{\sum_{k=1}^r J_{ik}^2}{\sum_{k=1}^r \sigma_k^2} \quad (7)$
Relativa del elemento $i$ -ésimo (fila) al factor $l$ , representa la variabilidad del factor explicada por el individuo.	$CRE_i F_l = \frac{j_{il}^2}{\sigma_l^2} \quad (8)$
Del factor $l$ al elemento $i$ -ésimo (fila), representa la parte de la variabilidad de cada individuo explicada por el factor.	$CRF_l E_i = \frac{j_{il}^2}{\sum_{k=1}^r J_{ik}^2} \quad (9)$
Relativa a la traza del elemento $j$ -ésimo (columna), representa la variabilidad total explicada por la variable.	$CRT_j = \frac{\sum_{k=1}^r h_{jk}^2}{\sum_{k=1}^r \sigma_k^2} \quad (10)$
Relativa del elemento $j$ -ésimo (columna) al factor $l$ , representa la variabilidad del factor explicada por la variable.	$CRE_j F_l = \frac{h_{jl}^2}{\sigma_l^2} \quad (11)$
Del factor $l$ al elemento $j$ -ésimo (columna), representa la parte de la variabilidad de cada variable explicada por el factor.	$CRF_l E_j = \frac{h_{jl}^2}{\sum_{k=1}^r J_{jk}^2} \quad (12)$
Calidad de representación de la $i$ -ésima (fila).	$CLRE_i P_{I-II} = CRF_l E_i + CRF_{II} E_i \quad (13)$
Calidad de representación de la $j$ -ésima (columna).	$CLRE_j P_{I-II} = CRF_l E_j + CRF_{II} E_j \quad (14)$

**Fuente:** Elaboración propia

*Ambiente:* Similar al ACP el eje 1 clasificó las filas por época y el eje 2 por ubicación geográfica. Se observa mejor representadas ( $> QLR$ ) las mediciones de julio-octubre (Esmeraldas) y diciembre (Manta) y están asociadas con temperatura, siendo la temperatura la variable con mejor calidad de representación. Explicación del 76.5 % de la inercia, Figura 1(b.1). *Especies:* Eje 1 clasifica según la época y el eje 2 según el tipo de estación, las mediciones mejor representadas ( $> QLR$ ) son las de febrero a mayo para Puerto Bolívar a la que se asocia la abundancia del grupo de especies *Thalassionematales* (O9). La calidad de representación de los individuos (filas) está representada por el tamaño. Para este análisis se utilizó la herramienta Multibplot (Vicente-Villardón, 2016), Figura 1(b.2).

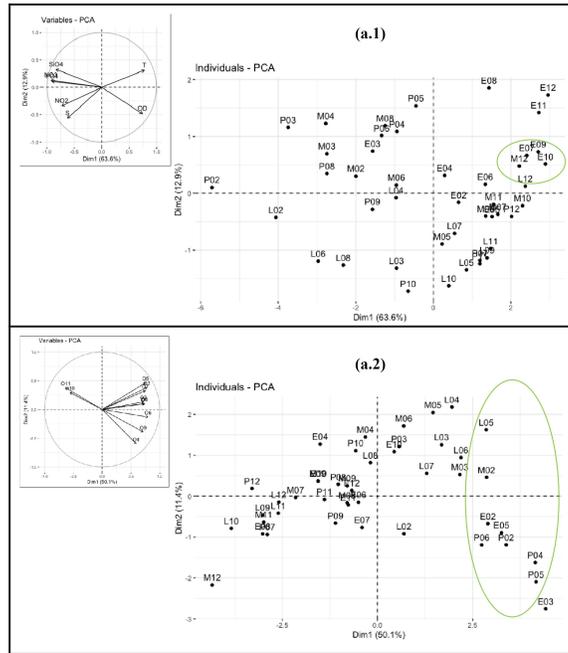
### 2.3 Análisis de Correspondencias – AC

Es un método de ordenación que estudia las asociaciones (relaciones de dependencia) entre las categorías de dos variables cualitativas. Permite representar como puntos en un espacio vectorial de baja dimensión las filas y columnas de una tabla de contingencia (Benzécri, 1973; Hill, 1973). Se inicia con la hipótesis de que no existe una relación de dependencia entre ambas variables cualitativas.

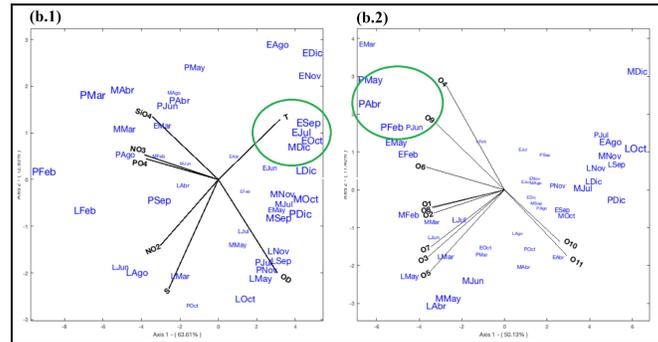
Este método se centra en analizar los perfiles obtenidos como la frecuencia relativa de los elementos filas o columnas con respecto a su total marginal. Las “masas” son valores asignados a cada perfil (fila y columna) calculado como el cociente entre el marginal (fila o columna) y el total global; la distancia ji-cuadrado que es una distancia ponderada que asigna una ponderación menor a las frecuencias más comunes y una ponderación mayor a las menos comunes siendo utilizada para medir las distancias entre los perfiles; y la inercia que representa la variabilidad total de los perfiles en el espacio vectorial.

Cada elemento fila y columna aporta con una contribución a la inercia de un eje factorial y cada eje factorial contribuye a la explicación de la inercia de una variable. Las categorías con altas contribuciones permiten interpretar el correspondiente eje factorial y son las que están mejor representadas (mayor interés) en el gráfico, mientras que los puntos ubicados cerca al origen tienen una baja representación.

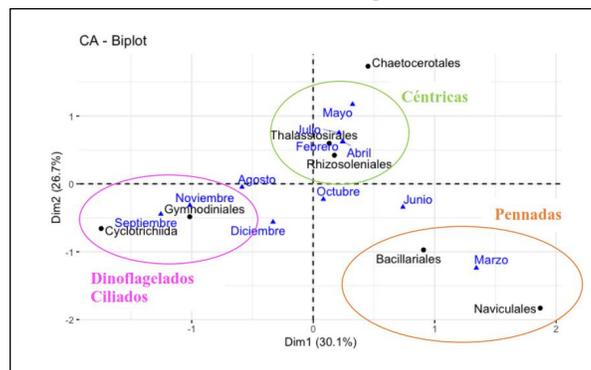
*Especies:* Se asocian los meses del muestreo con el orden de las especies de mayor abundancia. Se observa que el método asoció los últimos meses del año (época seca) con el grupo de los dinoflagelados y ciliados, en marzo (lluvioso) asocia con el orden de las diatomeas pennadas y los meses abril, mayo (lluvioso) y julio (seca) con el orden de las diatomeas céntricas. Total, de inercia proyectada 56.8 %, Figura 1(c)



(a) Análisis de Componentes Principales (ACP). (a.1) Ambiente y (a.2) Especies



(b) Análisis HJ-Biplot



(c) Análisis de Correspondencias. Matriz Especies

**Figura 1:** Aplicación de métodos clásicos de Análisis Multivariante

Fuente: Elaboración propia

### 3 Análisis de la Relación entre Dos Matrices Multivariantes de Dos-Vías

Para realizar estos análisis se parte de dos conjuntos de datos, ordenados en las matrices  $X$  e  $Y$ , observados sobre los mismos  $I$  individuos; de esta forma se puede estudiar la relación existente entre ambas matrices (Ver Figura A 1.b).

Las técnicas que permiten realizar este análisis son:

- Análisis de Correlación Canónica (CANCOR) que consiste en relacionar dos conjuntos de variables de tal manera que trata de encontrar dos nuevas variables compuestas  $U$  y  $V$ , las cuales son combinaciones lineales de  $X$  y  $Y$  respectivamente, que además poseen la máxima correlación entre ellas (Hotelling, 1936).
- Análisis de la Redundancia (RDA) es una técnica de ordenación generalizada del ACP que estudia la relación entre un conjunto de variables y otro de variables instrumentales (Rao, 1964).
- Análisis Canónico de Correspondencias (CCA) es una extensión del AC que a partir de combinaciones lineales de las variables ambientales (independientes) calcula los ejes de ordenación asegurando de esta manera la relación directa entre las variaciones del ambiente y de las especies (análisis del gradiente directo) (Ter Braak, 1986).
- Análisis de la CO-Inercia (COIA) que busca ejes que maximicen la covarianza entre los individuos de las dos tablas (Dolédéc y Chessel, 1994).

El CCA y el RDA son métodos asimétricos cuyo enfoque del análisis está orientado a considerar a las especies como respuesta a los forzamientos ambientales; mientras que la CO-Inercia y el CANCOR son métodos simétricos debido a que las dos matrices desempeñan un rol similar en el análisis. El CCA, RDA y CANCOR incluyen en uno de sus pasos el análisis de regresión, por lo que requieren que la cantidad de datos sea mayor al del número de variables, lo que no ocurre con la CO-Inercia (Thioulouse, Simier, y Chessel, 2004).

#### 3.1 Análisis Canónico de Correspondencias – CCA:

Es una técnica que permite analizar la relación de la composición de las comunidades con las variaciones ambientales (conocidas). Este método se compone de los siguientes pasos (Ter Braak, 1986):

Inicia con dos matrices multivariantes  $X_{[I,P]}$  y  $Y_{[I,Q]}$ , las variables en  $X$  pueden ser cuantitativas o nominales y las variables en  $Y$  pueden ser cuantitativas o dicotómicas;  $P < I$ ,  $P$  y  $Q$  variables e  $I$  individuos.

Resume la variación principal en la data  $Y$  por el método de ordenación (Gaussiano):

$$E(y_{ik}) = c_k \exp\left[-1/2 (z_i - u_k)^2 / t_k^2\right] \quad (15)$$

$E(y_{ik})$  es el promedio esperado de los  $y_{ik}$  en el individuo  $i$  que tiene puntuación  $z_i$  en el eje de ordenación,  $c_k$  es el parámetro para la especie  $k$ ;  $u_k$  (óptimo) el valor de  $z$  para el cuál alcanza su máximo y  $t_k$  el valor de tolerancia.

Como siguiente paso, relaciona los ejes de ordenación a las variables en  $X$ ,

$$z_i = b_0 + \sum_{j=1}^P b_j x_{ij} \quad (16)$$

Calcula puntajes para las variables en  $Y$  mediante el promedio ponderado de los puntajes de los individuos:

$$\lambda u_k = \sum_{i=1}^I y_{ik} \frac{z_i}{y_{+k}} = 1; \quad y_{+k} \text{ es el total de las variables en } Y \quad (17)$$

Calcula puntajes de los nuevos sitios mediante el promedio ponderado de los puntajes de las especies:

$$Z_i^* = \sum_{k=1}^Q y_{ik} \frac{u_k}{y_{i+}}; \quad y_{i+} \text{ es el total de los lugares} \quad (18)$$

Calcula coeficientes de regresión, utilizando regresión múltiple ponderada de las puntuaciones de los individuos en las variables ambientales:

$$b = (X^T R X)^{-1} X^T R Z^* \quad (19)$$

Calcula nuevas puntuaciones para los individuos:

$$z = X b \quad (20)$$

Normaliza las puntuaciones de los individuos:

$$\sum_i y_{i+} z_i = 0 \quad y \quad \sum_i y_{i+} z_i^2 = 1 \quad (21)$$

Finaliza al encontrar convergencia, caso contrario retorne a la Eq. (17)

*Ambiente-Especie:* Inercia total 85 %. Temperatura y oxígeno disuelto están relacionados de manera directa e inversamente con los nutrientes, las mediciones de las 4 estaciones para los meses de julio a diciembre los asocia con temperatura y oxígeno. Las especies no presentan una buena representación en el análisis, están situadas cercanas al origen, Figura 2(a).

### 3.2 CO-Inercia - COIA:

Se basa en encontrar la estructura en común (co-estructura) entre un par de matrices de datos multivariantes. Chessel y Mercier (1993) se refirieron al análisis de la co-estructura como: El análisis de la COIA de dos ACP normalizados que es equivalente al análisis Inter-battery de Tucker (1958), que lo propone como una alternativa del análisis canónico para estudiar las relaciones entre los valores de  $n$  individuos con dos baterías de pruebas (psicometría).

Este método realiza un análisis simultáneo de dos matrices  $X$  e  $Y$ . Inicialmente analiza cada matriz de manera independiente y obtiene ejes principales que representan la dirección de los vectores que maximizan la variabilidad proyectada (inercia), las unidades de observación pueden ser ordenadas sobre estos ejes tal como en los análisis de ACP y AC. A partir de este análisis se obtienen nuevos ejes tal que la covarianza entre los dos nuevos conjuntos de puntuaciones proyectada sea máxima, es decir una máxima correlación (Dolédec y Chessel, 1994). Dray, Chessel, y Thioulouse (2003) presentan las bondades del método comparado con el Análisis de Correspondencia Canónico.

A continuación, se describe el modelo matemático del método propuesto (Dolédec y Chessel, 1994; Dray et al., 2003):

Sean dos matrices normalizadas (centradas y escaladas):

$$X_{[I,P]} \text{ e } Y_{[I,Q]} \quad (22)$$

sus respectivos tripletes son:

$$(X, D_P, D_I) \text{ y } (Y, D_Q, D_I) \quad (23)$$

tal que las matrices diagonales  $D_I$  contiene los pesos  $(1/I)$  para los  $I$  individuos-observaciones (filas),  $D_P$  y  $D_Q$  son los pesos (unitarios) asociados a las  $P$  y  $Q$  variables (columnas) respectivamente.

Cada triplete es diagonalizado, tal que:

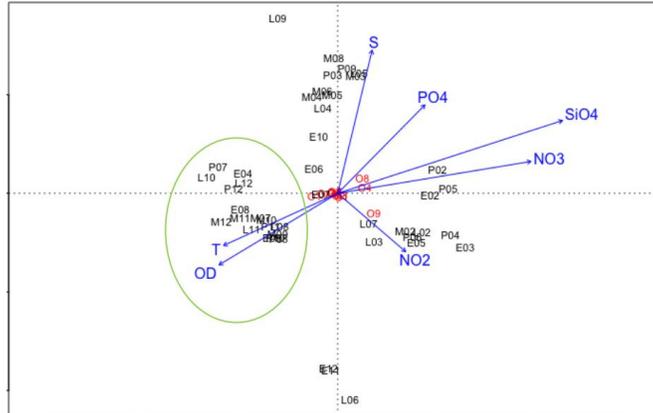
$$(X, D_P, D_I) = RNC; \quad R \in \mathbb{R}^{n \times r}, \quad V \in \mathbb{R}^{p \times r} \text{ y } N \in \mathbb{R}^{r \times r} \quad (24)$$

siendo  $N$  la matriz que contiene los valores propios  $(v_1, \dots, v_r)$  almacenados en el vector  $v$ ,  $R$  contiene las puntuaciones de las  $I$  filas y  $C$  las puntuaciones de las  $P$  filas en los  $r$  ejes, y

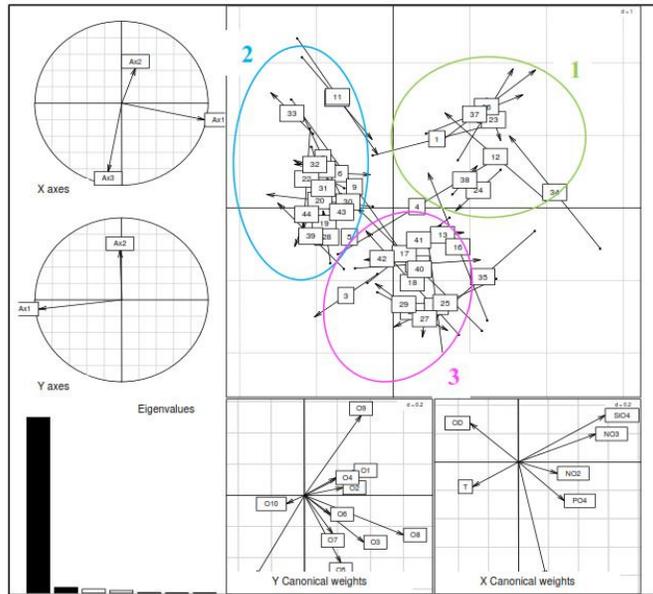
$$(Y, D_Q, D_I) = RMC; \quad R \in \mathbb{R}^{n \times s}, \quad V \in \mathbb{R}^{p \times s} \text{ y } N \in \mathbb{R}^{s \times s} \quad (25)$$

siendo  $M$  la matriz que contiene los valores propios  $(u_1, \dots, u_s)$  almacenados en el vector  $u$ ,  $R$  contiene las puntuaciones de las  $I$  filas y  $C$  las puntuaciones de las  $Q$  filas en los  $s$  ejes.

$\xi$  y  $\psi$  son matrices columnas que contienen  $I$  coordenadas generadas por las proyecciones del espacio multidimensional asociado con  $X$  y  $Y$  en los vectores  $u$  (normalizado por  $D_P$ ) y  $v$  (normalizado por  $D_Q$ ) respectivamente,



(a) Análisis Canónico de Correspondencias (ambiente-especie)



(b) Análisis de CO-Inercia (ambiente-especie)

**Figura 2:** Aplicación de Métodos CCA y COIA  
Fuente: Elaboración propia

$$\xi = XD_P u \quad y \quad \psi = YD_Q v \quad (26)$$

La CO-inercia se la obtiene como:

$$H(u, v) = \xi^T D_I \Psi \cong (X^T D_I Y, D_P, D_Q) \quad (27)$$

Su medida de covarianza es:

$$Cov(\xi, \Psi) = Corr(\xi, \Psi) \times VAR(\xi) \times VAR(\Psi) \quad (28)$$

$$Cov(\xi, \Psi) = Corr(\xi^T D_I \Psi, YD_Q v) \times (VAR(\xi^T D_I \Psi))^{1/2} \times (VAR(YD_Q v))^{1/2} \quad (29)$$

*Relación Ambiente-Especie:* Total de inercia proyectada 95.93 %. Relación directa entre Temperatura y Oxígeno e inversa con los nutrientes. Relación directa entre los grupos Gymnodiniales (O10) y Cyclotrichiida (O11) e inversa con las demás. Se forman 3 agrupaciones: 1) color verde, están agrupados los meses lluviosos principalmente febrero para las 4 estaciones de muestreo; 2) color celeste, los meses entre la culminación de la época lluviosa e inicio de la seca (abril, mayo, junio, julio) principalmente para las 3 estaciones excepto Puerto Bolívar (agosto-octubre) y 3) color rosado, los últimos meses del año (octubre-diciembre) en las 4 estaciones, Figura 2(b).

#### 4 Análisis Multivariante de Matrices de Tres-Vías

Hay ocasiones que se tiene la necesidad de estudiar el comportamiento de un evento, teniendo en consideración las características de las unidades de observación (filas) explicadas por múltiples variables (columnas) y su evolución en el tiempo o su comportamiento ante diferentes situaciones (condición). Es decir, los datos se encuentran estructurados como un conjunto de matrices de tres dimensiones o cubo de datos (Ver Figura A 1.c).

La matriz de tres dimensiones  $X_{[I,P,K]}$  está compuesta por los elementos descritos en (30):

$$\{x_{ijk} | i = 1, \dots, I; \quad j = 1, \dots, P; \quad k = 1, \dots, K\} \quad (30)$$

La matriz o tabla de tres vías se originó con Ledyard R Tucker y fueron Douglas Carroll y Richard Harshman quienes realizaron aportaciones en el desarrollo de esta metodología. Fue a partir de 1976 que aparecen los métodos que estudiarían la estabilidad de la estructura del conjunto de matrices, y son reconocidos como la familia STATIS: STATIS, STATIS DUAL y X-STATIS o también conocido como Análisis Triádico Parcial (PTA).

#### 4.1 *Structuration des Tableaux à Trois Indices de la Statistique - STATIS:*

Este método fue desarrollado por L'Hermier des Plantes (1976) y recibió importantes contribuciones de Robert y Escoufier (1976); Lavit (1988); Lavit, Escoufier, Sabatier, y Traissac (1994).

STATIS permite calcular la parte estable entre las matrices (tablas) de datos que contienen los mismos  $I$  individuos (filas) con diferentes conjuntos de variables (columnas) evaluados en  $K$  condiciones (momentos o situaciones). Este método se encarga de resaltar la posición relativa de los individuos (Ver Figura A 1.d).

La definición matemática del método es:

Se tiene la matriz de tres dimensiones construida a partir de  $I$  individuos (los mismos en cada matriz),  $P$  variables diferentes y  $K$  condición.

$$X_{[I,P,K]} \quad (31)$$

Para cada matriz se tienen los tripletes que contienen la matriz diagonal  $D_I$  con los pesos para las filas y  $D_{p,k}$  los pesos para las columnas de cada  $k$  matriz.

$$(X_1, D_I, D_{p,1}), (X_2, D_I, D_{p,2}), \dots, (X_k, D_I, D_{p,k}) \quad (32)$$

Se calcula la matriz:

$$W_k = (X_k D_{p,k} X_k^T); \text{ para } k = 1, \dots, K \quad (33)$$

La matriz de producto interno entre los individuos de la  $k$  matriz, conocido como el operador de Escoufier.

$$W_k D_I \quad (34)$$

El STATIS se compone de tres elementos obtenidos de forma secuencial, que son:

##### i. *Inter-estructura*

Se calcula la matriz de coeficientes de correlación vectorial RV, tal como:

$$W_k = (X_k D_{p,k} X_k^T D_I) \text{ y } W_l = (X_l D_{p,l} X_l^T D_I); \text{ para } k \text{ y } l = 1, \dots, K \quad (35)$$

$$COVV(X_k, X_l) = Trace(X_k D_{p,k} X_k^T D_I X_l D_{p,l} X_l^T D_I) \quad (36)$$

Como operador de Escoufier:

$$COVV(X_k, X_l) = Trace(W_k D_I W_l D_I) \quad (37)$$

La varianza vectorial:

$$VARV(X_k) = COVV(X_k, X_k) = Trace(W_k D_I W_k D_I) = Trace(W_k D_I)^2 \quad (38)$$

$$RV : RV(X_k, X_l) = \frac{COVV(X_k, X_l)}{\sqrt{VAR(X_k)} \sqrt{VAR(X_l)}}; \quad \text{tal que } 0 \leq (X_k, X_l) \leq 1 \quad (39)$$

Se factoriza la matriz  $RV_{[K,K]}$ , tal que sus valores propios  $\lambda_K$  y su vector propio normado  $U_K$  proporcionan las puntuaciones de las matrices  $A$  que son representadas en un círculo de correlaciones

$$A = U_K \lambda_K^{1/2} \quad (40)$$

### ii. *Compromiso o consenso*

El primer vector propio de la inter-estructura  $u_{k=1}^T(\alpha_1, \alpha_2, \dots, \alpha_K)$ , tal que  $\sum_{i=1}^K \alpha_i^2 = 1$ , los  $\alpha_i$  son los pesos de las matrices iniciales  $X_k$ .

Matriz compromiso:

$$WD_I = \sum_{k=1}^K \alpha_k \frac{W_k D_I}{\sqrt{VAR(X_K)}} \quad (41)$$

Al factorizar la matriz  $WD_I$  se tienen los valores propios  $\Lambda$  y vectores propios  $V$ , tal que  $(V^T IV) = 1$ . Las coordenadas de los  $I$  individuos para representarlos sobre los ejes del plano factorial, se los obtiene como:

$$L = WD_I V \Lambda^{1/2} \quad (42)$$

### iii. *Intra-estructura*

Consiste en proyectar las variables de cada matriz sobre el compromiso:  $C_k = X_k^T D_I V$  y los individuos

$$L_k = W_k D_I V \Lambda^{1/2} \quad (43)$$

La aplicación del método solo se realizará para el método PTA, debido a la estructura que tienen los datos empleados en este trabajo.

## 4.2 STATIS DUAL

El STATIS Dual (L'Hermier des Plantes, 1976) se caracteriza por enfatizar las posiciones relativas de las variables. A partir de  $K$  matrices que poseen las mismas  $P$  variables que son medidas en diferentes conjuntos de  $I$  individuos (Ver Figura A 1.e).

A diferencia del STATIS, el STATIS DUAL en vez de calcular la matriz de  $W_k$  de los productos interno entre los individuos calcula la matriz  $C_k$  de covarianzas entre variables.

$$C_k = (X_k^T D_{I,k} X_k); \quad \text{para } k = 1, \dots, K \quad (44)$$

La matriz de covarianzas entre variables de la  $k$ -ésima matriz es

$$C_k D_P \quad (45)$$

Se calcula la matriz de correlaciones vectoriales entre las  $C_k$ , utilizando el producto interno de Hilbert-Schmidt (HS):

$$C = (C_k C_l) = \text{trace}(C_k D_P C_l D_P); \quad \text{para } k \text{ y } l = 1, \dots, K \quad (46)$$

El STATIS-DUAL realiza los mismos pasos que el STATIS y proyecta sobre un plano factorial la matriz compromiso para las variables.

### 4.3 Análisis Triádico Parcial - PTA:

También conocido como X-STATIS (Jaffrenou, 1978) consiste en encontrar la proporción de variabilidad entre las variables que dependen en situaciones o tiempo. El PTA requiere que las  $K$  matrices tengan igual número de  $P$  variables e igual número de  $I$  individuos entre ellas, formando un cubo de datos (Ver Figura A 1.f).

PTA no trabaja con operadores como es el caso de STATIS, ya que trabaja directamente con las matrices. Este método consiste en realizar un análisis STATIS a las  $X_k$  matrices.

Similar al STATIS, este método se compone de: la Inter-estructura que expresa la importancia de cada matriz en el análisis, el Compromiso que manifiesta la estructura en común entre las matrices y la Intra-estructura que proyecta los individuos y las variables de cada matriz en el espacio bidimensional del compromiso, permite analizar la representación de cada matriz de forma independiente.

Se inicia con una matriz de tres dimensiones, que contiene los mismos  $I$  individuos y  $P$  variables en  $K$  condición.

$$X_{[I,P,K]} \quad (47)$$

Para cada matriz se tienen los tripletes que contienen la matriz diagonal  $D_I$  con los pesos para las filas y  $D_P$  los pesos para las columnas de cada  $k$ -matriz.

$$(X_1, D_I, D_P), (X_2, D_I, D_P), \dots, (X_k, D_I, D_P) \quad (48)$$

Se calcula el producto interno entre las  $k$ -matrices, tal como:

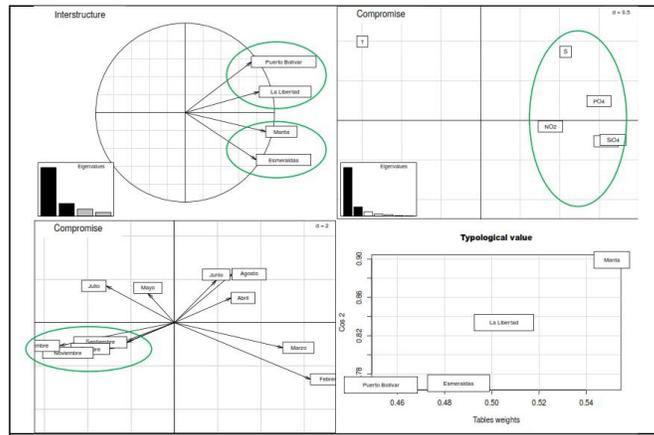
$$COVV(X_k, X_l) = \text{Trace}(X_k^T D_I X_l D_P) = \text{Trace}(X_l^T D_I X_k D_P) \quad (49)$$

Varianza vectorial se la calcula como:

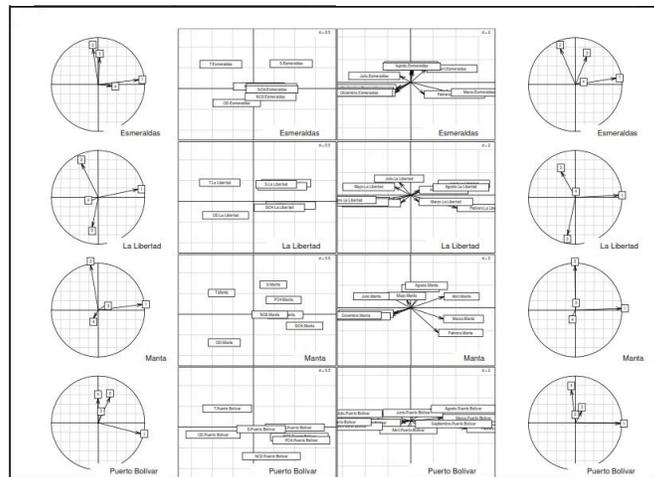
$$VARV(X_k, X_k) = \text{Trace}(X_k^T D_I X_k D_P) \quad (50)$$

El RV (coeficiente de correlación vectorial):

$$RV(X_k, X_l) = \frac{COVV(X_k, X_l)}{\sqrt{VAR(X_k)} \sqrt{VAR(X_l)}} \quad (51)$$



(a) Inter-estructura y Compromiso



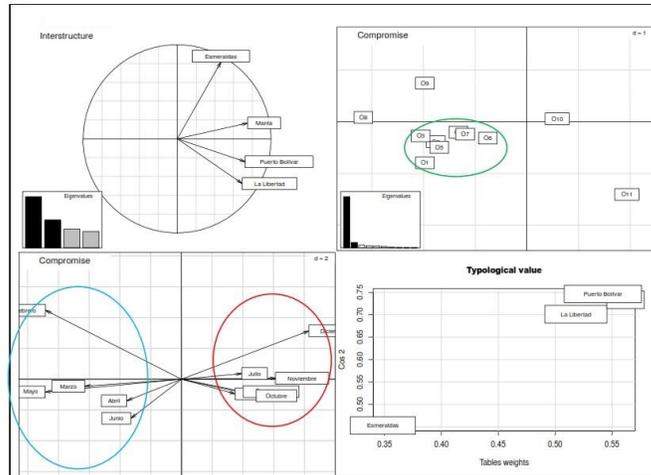
(b) Intra-estructura

**Figura 3:** Análisis PTA. Matriz Ambiente

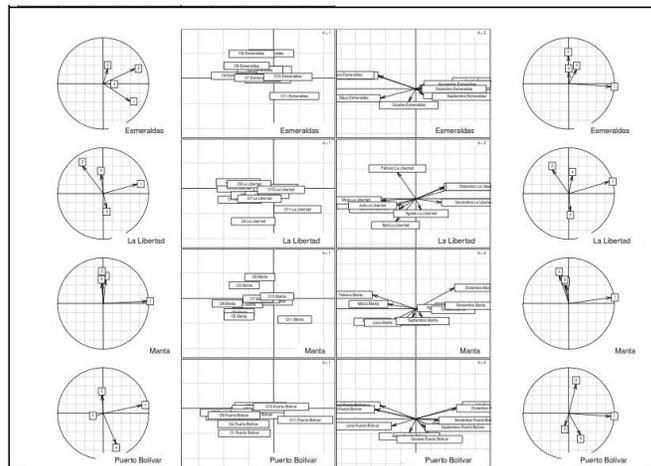
**Fuente:** Elaboración propia

Para obtener la matriz de producto interno se utilizan las  $X_k$  matrices en lugar de los  $W_k$  (STATIS), del mismo modo para calcular el compromiso se utilizan las  $X_k$  matrices para la combinación lineal ponderada.

Los pasos de cálculo para la inter-estructura, compromiso e intra-estructura son similares a los detallados en el STATIS estándar.



(a) Inter-estructura y Compromiso



(b) Intra-estructura

**Figura 4:** Análisis PTA. Matriz Especies

**Fuente:** Elaboración propia

*Ambiente:* En la Figura 3(a), la inter-estructura evidenció el 84.4 % de la variabilidad, expresando dos comportamientos ambientales, uno similar entre las estaciones localizadas al norte y otro entre las estaciones del sur. El compromiso manifestó

un 84.4 % de la estabilidad ambiental a lo largo del espacio y representa de mejor manera lo ocurrido en Manta ( $\cos^2 = 0,90$ ) y La Libertad ( $\cos^2 = 0,83$ ) que son las estaciones con mayor influencia oceánica. La temperatura y el oxígeno muestran una correlación directa e inversa con los nutrientes y salinidad. Los parámetros ambientales se mostraron muy variantes durante el año 2015, ocurriendo una mayor homogeneidad de septiembre a diciembre. Las más altas temperaturas ocurrieron en mayo y julio.

En la Figura 3(b) Manta mostró un comportamiento muy similar al compromiso, en Esmeraldas se registraron los valores más altos de temperatura (vector más largo) y ocurrieron en el mes de julio principalmente, mientras que en Puerto Bolívar se registraron los valores más altos de nitrato y fosfato durante los meses de febrero y marzo (inicios del año).

*Especies:* En la Figura 4(a), la inter-estructura mostró el 69.3 % de la variabilidad, la similaridad biológica entre las estaciones es baja principalmente en Esmeraldas con respecto a las demás. El compromiso manifestó el 86.1 % de la variabilidad en la estabilidad capturada entre las 4 estaciones y representa de mejor manera lo ocurrido en Puerto Bolívar ( $\cos^2 = 0,75$ ) seguido de Manta ( $\cos^2 = 0,73$ ). El orden de las especies con comportamiento más similar fue: en las diatomeas céntricas Chaetocerotales, Hemiaulales, Leptocylindrales, Lithodesmiales, Rhizosoleniales, Thalassiosirales y en las diatomeas pennadas Bacillariales (del O1 al O7 respectivamente). Para el caso de los meses se observa que el método evidenció el comportamiento de la época estacional, a la izquierda los meses de invierno (azul, donde ocurrió la mayor abundancia de las especies antes detalladas) y a la derecha los meses de verano (rojo) que son donde se destacaron las especies del orden Gymnodiniales (O10) y Cyclotrichiida (O11). En la Figura 4(b) Puerto Bolívar muestra un resultado similar al del compromiso, mientras que en Esmeraldas y Manta destacaron las especies del orden Cyclotrichiida (O11) principalmente en el mes de septiembre.

#### 4.4 DISTATIS

Este método fue desarrollado por Abdi, Valentin, Chollet, y Chrea (2007). Es una variante del STATIS estándar y se caracteriza por analizar  $k$ -matrices de distancias evaluadas en los mismos  $I$  individuos. Su base matemática consiste en calcular matrices de producto interno entre las matrices de distancia.

Se inicia con una matriz de distancias de tres dimensiones, que contiene  $I$  filas e  $I$  columnas en  $K$  condiciones. Sea la matriz  $D_{[I,I,K]}$  con sus elementos:

$$\{d_{ijk} | i = 1, \dots, I; j = 1, \dots, I; k = 1, \dots, K\} \quad (52)$$

Se calcula el producto cruz entre matrices para analizarlas con Escalamiento Multidimensional (MDS), previamente cada elemento de las matrices de distancia se eleva al cuadrado y se aplica doble centrado:

$$\begin{aligned}\tilde{s}_{ij} &= -\frac{1}{2} \left\{ (d_{i,j} - \bar{d}_{+,+}) - (\bar{d}_{i,+} - \bar{d}_{+,+}) - (\bar{d}_{j,+} - \bar{d}_{+,+}) \right\} \\ \tilde{s}_{ij} &= -\frac{1}{2} (d_{i,j} - \bar{d}_{i,+} - \bar{d}_{j,+} + \bar{d}_{+,+})\end{aligned}\quad (53)$$

La Eq. (53) también puede ser expresada por la siguiente operación de matrices:

$$\begin{aligned}\Xi_{[I,I]} &= I_{[I,I]} - 1m_{[I,I,1]}^T; \quad \text{tal que } m_i = 1/I \quad (\text{centrado}) \\ \tilde{S}_{[k]} &= -\frac{1}{2} \Xi D_k \Xi^T \quad (\text{matriz de producto cruz})\end{aligned}\quad (54)$$

Posteriormente se normalizan las matrices  $\tilde{S}_{[k]}$  aplicando Análisis factorial múltiple (AFM) al dividir cada elemento de la matriz por su primer valor propio, obteniéndose las matrices normalizadas  $S_k$ .

Calcular coeficientes para la matriz RV:

$$RV(D_k, D_{k'}) = c_{k,k'} = \frac{\text{Trace}\{S_k^T S_{k'}\}}{\sqrt{\text{Trace}\{S_k^T S_k\} \times \text{Trace}\{S_{k'}^T S_{k'}\}}}; \quad \text{para } k \text{ y } k' = 1, \dots, K \quad (55)$$

Los coeficientes de RV se almacenan en la nueva matriz de similaridades  $C$ , que se factoriza para obtener los pesos para el compromiso.

$$C = P\Theta P^T; \quad \text{tal que } P^T P = I \quad (56)$$

El primer vector propio  $p_1$  es re-escalado para obtener los pesos para construir el compromiso, de la siguiente manera:  $\alpha = (1^T p_1)^{-1} \times p_1$

Compromiso:

$$S_+ = \sum_{k=1}^K \alpha_k S_k \quad (57)$$

Se aplica DVS a  $S_+$ :

$$S_+ = V\Lambda V^T \quad (58)$$

Las puntuaciones del compromiso para graficar las observaciones:

$$F = V\Lambda^{1/2} \quad (59)$$

La matriz  $F$  (operador de proyección) definida en (59) transforma las matrices de producto cruz en sus proyecciones sobre el mapa bifactorial del compromiso; por lo tanto, se obtiene la intra-estructura como:

$$F_k = S_k (V\Lambda^{-\frac{1}{2}}) \quad (60)$$

## 5 Análisis de la Relación entre Dos Matrices Multivariantes de Tres-Vías

Es muy común encontrar estudios en los que el investigador posee interés en estudiar la relación entre 2 conjuntos de matrices de tres vías  $X$  y  $Y$  (Ver Figura A 1.g).

Las filas de cada par de matrices deben contener la información de los mismos  $I$  individuos. La serie de datos en cada matriz corresponde a repeticiones que se dan en el tiempo o ante diferentes situaciones, tal que la estructura final consiste en pares de matrices que se repiten a lo largo del tiempo o en diferentes situaciones.

Para analizar este tipo de datos, se revisarán las dos técnicas comúnmente conocidas como *STATICO* y *COSTATIS*. Ambas parten de la idea principal: Encontrar la relación entre dos conjuntos de variables y analizar su estabilidad a través del tiempo o ante diferentes situaciones.

### 5.1 *STATICO*

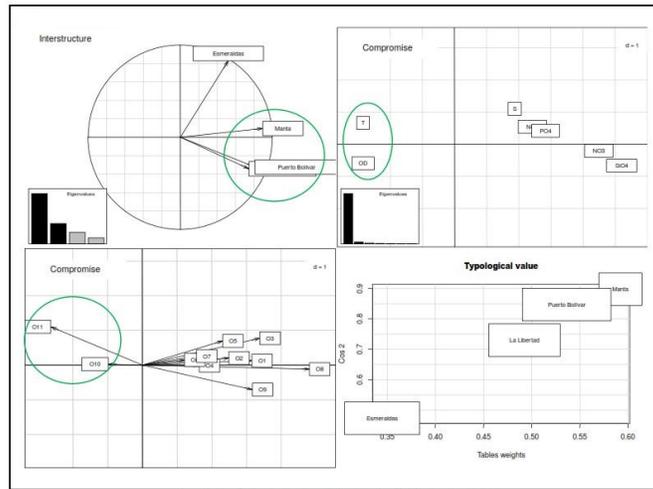
El método *STATICO* (Simier, Blanc, Pellegrin, y Nandris, 1999; Thioulouse et al., 2004) se compone de la combinación de dos métodos: de los métodos *STATIS* (Lavit et al., 1994) para encontrar una tipología estable entre las  $k$ -matrices que poseen al menos una dimensión común y de la *CO-Inercia* (Dolédec y Chessel, 1994) para encontrar entre dos pares de matrices lo que produce la tipología común. Es decir, el método *STATICO* consiste en encontrar la parte estable de la relación entre dos conjuntos de tablas  $X$  y  $Y$ . Este método de acoplamiento simétrico es el resultado de la aplicación de un análisis *PTA* a  $K$ -matrices cruzadas de *CO-Inercia* calculadas en cada repetición.

Los  $I$  individuos deben ser los mismos en cada par de tablas, pero pueden ser diferentes entre pares, el número de variables en la matriz  $X$  y  $Y$  debe ser el mismo en cada  $k$  condición respectivamente.

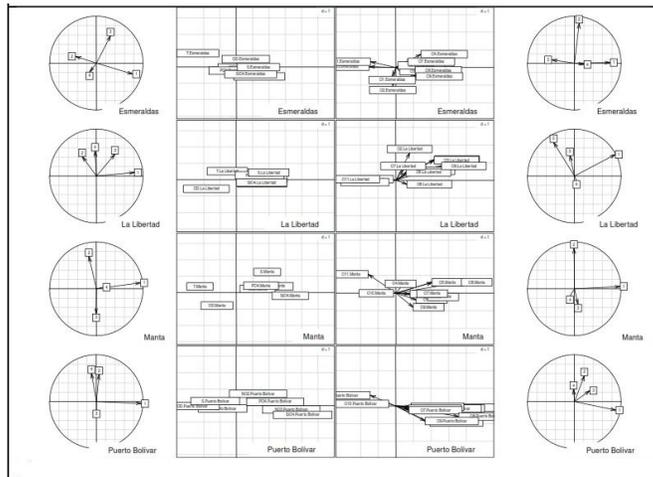
El método *STATICO* se compone de tres pasos (Thioulouse et al., 2004):

- i. Cada matriz inicial que componen la serie de matrices de tres-vías para  $X$  y  $Y$  es analizada por un análisis básico (*ACP*, *AC* o *ACM*) dependiendo del tipo de variable;
- ii. Posteriormente, cada par de matrices en cada  $k$ -condición se analizan de manera conjunta con la técnica de la *CO-Inercia*, obteniéndose una tabla cruzada;
- iii. Finalmente, se aplica *PTA* para analizar la nueva serie de tablas cruzadas (cubo de datos) obtenidos con la *CO-Inercia*.

La base metodológica de los métodos *CO-Inercia* y *PTA* se encuentran detallados en las secciones 3.2 y 4.3 respectivamente.



(a)



(b)

**Figura 5:** Análisis STATICO (ambiente-especie)

**Fuente:** Elaboración propia

En la Figura 5(a), la inter-estructura manifestó el 80.2 % de la variabilidad y evidenció que las estaciones de Puerto Bolívar, La Libertad y Manta presentaron mayor similitud en la relación ambiente-especie, esta relación es mejor representada por el compromiso para Manta ( $\cos^2 = 0,90$ ) seguido de Puerto Bolívar ( $\cos^2 = 0,85$ ). El compromiso explicó el 97.2 % de la variabilidad, en el caso de los parámetros ambientales se destacó una fuerte correlación directa entre la tem-

peratura y el oxígeno e inversa con los nutrientes y la salinidad, mientras que para el caso de las especies los Gymnodiniales (O10) y Cyclotrichiida (O11) mostraron un comportamiento similar entre ellos (asociado con altos valores de temperatura y oxígeno) pero contrario a las demás especies que mostraron preferencias por altos valores en nutrientes y salinidad.

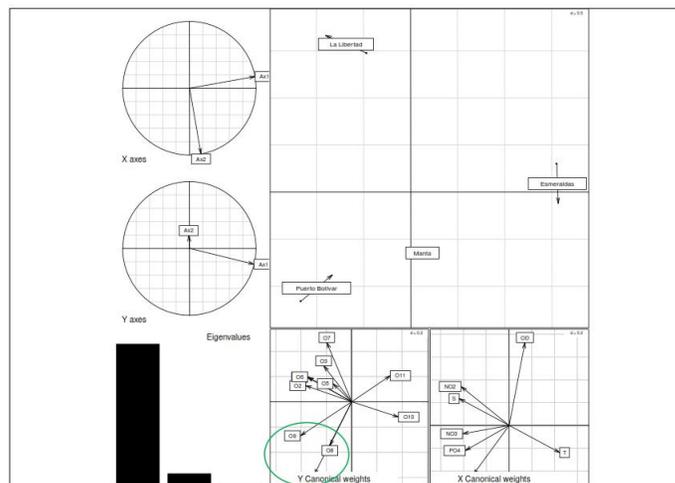
En la Figura 5(b), las estaciones de Manta y Puerto Bolívar mostraron una co-estructura ambiente-especie más similar con el compromiso. Esmeraldas manifestó que la variable temperatura es la de mayor variabilidad (vector más largo).

## 5.2 COSTATIS

El COSTATIS (Thioulouse, 2011) es una alternativa al método STATICO. A diferencia del STATICO, este primero trabaja de forma independiente cada cubo de datos y luego analiza la relación entre las matrices resultantes.

Pasos del COSTATIS:

- i. Realiza dos análisis PTA: uno al conjunto de  $k$ -tablas en  $X$  y otro al conjunto de  $k$ -tablas en  $Y$ ;
- ii. Posteriormente, realiza un análisis de CO-Inercia entre las matrices compromiso  $X_C$  y  $Y_C$  obtenidas en los dos PTA realizados en el paso anterior.



**Figura 6:** Análisis COSTATIS (ambiente-especie)

**Fuente:** Elaboración propia

Tiene la característica que el número de  $k$ -tablas no tiene que ser el mismo para  $X_{[I,P,K]}$  y para  $Y_{[I,Q,K']}$ , es decir  $K$  puede ser  $\neq K'$ ; sin embargo, las  $Q$  variables en  $Y$ , las  $P$  variables en  $X$  y los  $I$  individuos deben ser los mismos para todas las tablas.

Los Gymnodiniales (O10) fueron de mayor abundancia en la estación de Esmeraldas con altas concentraciones de temperatura y bajo en salinidad y nitrito, en Puerto Bolívar se destacaron por su mayor abundancia las órdenes de Lithodesmiales (O4), Naviculales (O8) y Thalassionematales (O9) con altas concentraciones de nitrato, fosfato y silicato. El resto de las especies destacaron su mayor abundancia en La Libertad con altas nitrito y salinidad y bajo en temperaturas. En Manta destacarían las especies mencionadas anteriormente con altos valores de temperatura, silicatos principalmente, es en esta estación donde se observó la más fuerte co-estructura ambiente-especie. El método explica el 98.8 % de la variabilidad con una correlación vectorial  $RV=0.84$ , Figura 6.

## 6 Conclusiones

Conforme han ido evolucionando la disponibilidad de los datos, su volumen y su estructura, han ido apareciendo metodologías multivariantes que apuntan a reducir las dimensiones para su análisis.

El gran legado dejado por Pearson con su propuesta ACP no pierde vigencia, más aún con la propuesta del BIPLLOT, realizada por Gabriel y Galindo.

La familia STATIS ha provisto una variedad de opciones para trabajar en bajas dimensiones varias series de matrices de datos a la vez, haciendo posible su representación gráfica y mejor interpretación.

La incorporación de algoritmos computacionales a estas metodologías seguramente ofrecerá más opciones ante el creciente volumen de datos disponibles.

## 7 Bibliografía

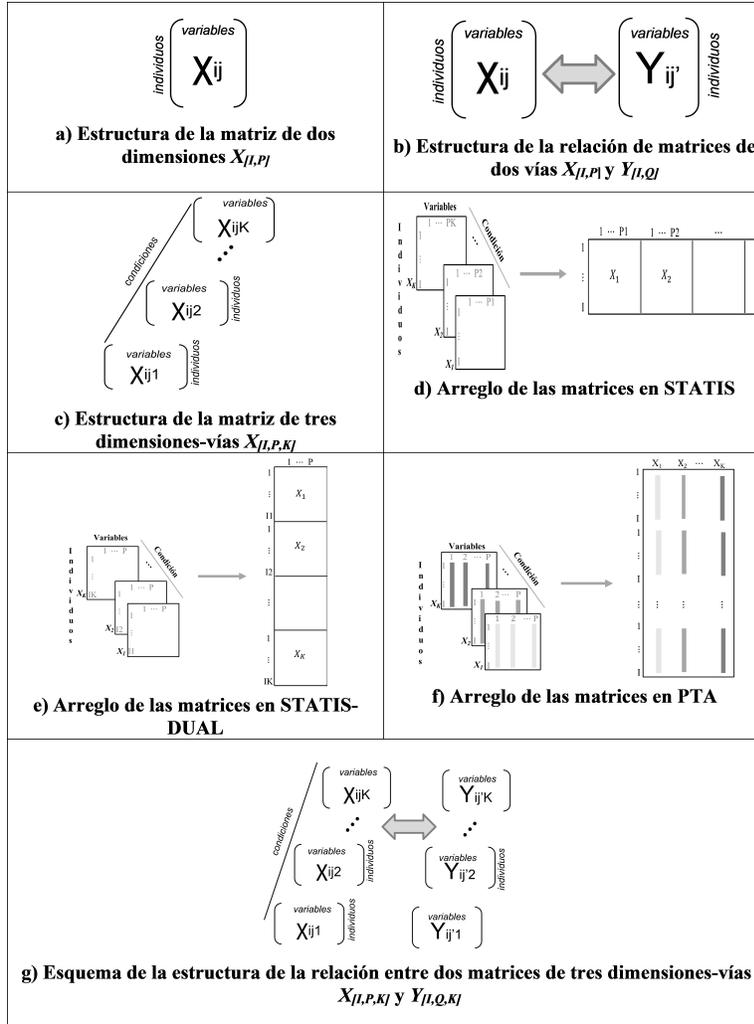
### Referencias

- Abdi, H., Valentin, D., Chollet, S., y Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference*, 18(4), 627–640. doi: 10.1016/j.foodqual.2006.09.003
- Benzécri, J. P. (1973). *L'analyse des données: L'analyse des correspondances* (22.<sup>a</sup> ed.; J. P. Benzécri, Ed.). Dunod. Descargado de <https://books.google.com.ec/books?id=sDTwAAAAAMAAJ>
- Chessel, D., y Mercier, P. (1993). Couplage de triplets statistiques et liaisons espèces-environnement. *Biométrie et Environnement*, 15–44.

- Cuadras, C. M. (2019). *Nuevos métodos multivariantes* (C. Ediciones, Ed.). doi: 10.1017/CBO9781107415324.004
- Dolédec, S., y Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater Biology*, 31(3), 277–294. doi: 10.1111/j.1365-2427.1994.tb01741.x
- Dray, S., Chessel, D., y Thioulouse, J. (2003). CO-inertia analysis and the linking of ecological data tables. *Ecology*, 84(11), 3078–3089. doi: 10.1890/03-0178
- Gabriel, K. R. (1971). The biplot graphic display of matrices with applications to principal component analysis. *Biometrika*, 58(3), 453–467. doi: 10.2307/2334381
- Galindo-Villardón, M. P., y Cuadras, C. M. (1986). Una extensión del método Biplot y su relación con otras técnicas. *Publicaciones de Bioestadística y Biomatemática*, 17.
- Galindo-Villardón, M. P. (1986). Una Alternativa de Representación Simultánea: HJ-Biplot. *Questiio*, 10(1), 13-23.
- Hill, M. O. (1973). Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology*, 61(1), 237–249. doi: 10.2307/2258931
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417-441.
- Hotelling, H. (1936). Simplified calculation of principal components. *Psychometrika*, 1(1), 27–35. doi: 10.1007/BF02287921
- Jaffrenou, P. A. (1978). *Sur l'analyse des familles finies de variables vectorielles: bases algébriques et application à la description statistique* (Université Lyon 1, Ed.). Descargado de <https://books.google.com.ec/books?id=mc4GOgAACAAJ>
- Lavit, C. (1988). *Analyse conjointe de tableaux quantitatifs* (Dunod, Ed.). Descargado de <https://books.google.com.ec/books?id=9dl0AAAACAAJ>
- Lavit, C., Escoufier, Y., Sabatier, R., y Traissac, P. (1994). The ACT (STATIS method). *Computational Statistics and Data Analysis*, 18(1), 97-119. doi: [https://doi.org/10.1016/0167-9473\(94\)90134-1](https://doi.org/10.1016/0167-9473(94)90134-1)
- L'Hermier des Plantes, H. (1976). *Structuration des tableaux a trois indices de la statistique: Theorie et application d'une methode d'analyse conjointe*

- te. Descargado de <https://books.google.com.ec/books?id=gYtVGwAACAAJ>
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11), 559–572. doi: 10.1080/14786440109462720
- Rao, C. R. (1964). The use and interpretation of Principal Component Analysis in applied research. *The Indian Journal of Statistics*, 26(4), 329–358. Descargado de <http://www.jstor.org/stable/25049339>
- Robert, P., y Escoufier, Y. (1976). A Unifying Tool for Linear Multivariate Statistical Methods: The RV-Coefficient. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 25(3), 257–265. Descargado de <http://www.jstor.org/stable/2347233>
- Simier, M., Blanc, L., Pellegrin, F., y Nandris, D. (1999). Approche simultanée de  $K$  couples de tableaux: Application à l'étude des relations pathologie végétale-environnement. *Revue de statistique appliquée*, 47(1), 31–46.
- Ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67(5), 1167–1179. doi: 10.2307/1938672
- Thioulouse, J. (2011). Simultaneous analysis of a sequence of paired ecological tables: A comparison of several methods. *The Annals of Applied Statistics*, 5(4), 2300–2325. doi: 10.1214/10-AOAS372
- Thioulouse, J., Simier, M., y Chessel, D. (2004). Simultaneous analysis of a sequence of paired ecological tables. *Ecology*, 85(1), 272–283.
- Tucker, L. R. (1958). An Inter-battery method of factor analysis. *Psychometrika*, 23(2), 111–136.
- Vicente-Villardón, J. L. (2016). *MULTBILOT: A package for multivariate analysis using Biplots*.

Apéndice



**Figura A 1:** Estructuras y Arreglos Multivariantes  
Fuente: Elaboración propia