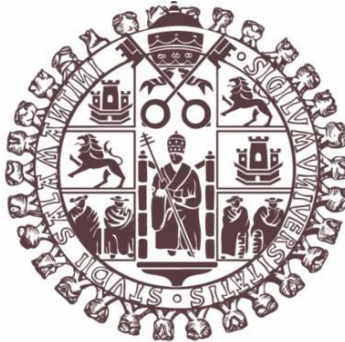


**UNIVERSIDAD DE SALAMANCA**

Departamento de Estadística

Máster en Análisis Avanzado de Datos Multivariantes

Trabajo Fin de Máster



**EVALUACIÓN DE LA CALIDAD DEL AGUA EN GAMBOA Y  
PARAISO DEL EMBALSE GATÚN, EN LA CUENCA  
HIDROGRÁFICA DEL CANAL DE PANAMÁ**

AUTOR

**GONZALO ISAAC CARRASCO OBERTO**

TUTORAS

**M<sup>a</sup> Purificación Galindo Villardón**

**M<sup>a</sup> Carmen Patino Alonso**

2015



**VNiVERSiDAD  
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Dpto. de Estadística

Universidad de Salamanca

**DRA. M<sup>a</sup> PURIFICACIÓN GALINDO VILLARDÓN**

*Profesora Titular del Departamento de Estadística de la Universidad de Salamanca*

**DRA. M<sup>a</sup> CARMEN PATINO ALONSO**

*Profesora Ayudante Doctora del Departamento de Estadística de la Universidad de Salamanca*

---

CERTIFICAN que **D. Gonzalo Isaac Carrasco Oberto** ha realizado en la Universidad de Salamanca, bajo su dirección, el trabajo que para optar título de Máster en Análisis Avanzado de Datos Multivariantes, presenta con el título **Evaluación de la calidad del agua en Gamboa y Paraíso del embalse Gatún, en la Cuenca hidrográfica del canal de Panamá**, autorizando expresamente su lectura y defensa.

Y para que conste, firman el presente certificado en Salamanca a 13 de julio de 2015.

M<sup>a</sup> Purificación Galindo Villardón

M<sup>a</sup> Carmen Patino Alonso

**Evaluación de la calidad del agua en Gamboa y Paraíso del embalse Gatún, en la  
Cuenca hidrográfica del canal de Panamá**



**UNIVERSIDAD  
DE SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Dpto. de Estadística    Universidad de Salamanca

Trabajo para optar al título de Máster en  
Análisis Avanzado de Datos Multivariantes  
por la Universidad de Salamanca.

Presenta:

**Gonzalo Isaac Carrasco Oberto**

**Salamanca**

**2015**

## **AGRADECIMIENTOS**

Gracias a Dios porque más que pedirte, tengo que agradecerte.

A la Directora de este Trabajo Fin de Master Dra. M<sup>a</sup> Purificación Galindo Villardón que me brindo la oportunidad de cursar el Master en Análisis Avanzados de Datos Multivariantes, así como por sus enseñanzas, consejos, asesoramiento prestado para la realización de este Trabajo Fin de Máster.

A la Directora de este Trabajo Fin de Master Dra. M<sup>a</sup> Carmen Patino Alonso por sus sugerencias y colaboración en los requerimientos que he tenido, durante este trabajo.

A la Licenciada Marisela del C Castillo G., por su orientación en el tema de Calidad del Agua en la Cuenca Hidrográfica del Canal de Panamá.

A mi familia, la cual me ha apoyado en todo tiempo.

A mis colegas y compañeras de cursos Mitzi, Carmen, Estelina y Aurora por su dedicación y compromiso en estos años de estudio.

# ÍNDICE

RESUMEN .....	1
INTRODUCCIÓN .....	2
1. MÉTODOS PARA ANALIZAR TABLAS DE DE DOS Y TRES VÍAS ..	5
1.1 MÉTODOS PARA ANALIZAR TABLAS DE DOS VÍAS .....	5
1.1.1 MÉTODOS BILOT .....	5
1.1.1.1 GH-BILOT O CMP-BILOT (COLUMN METRIC PRESERVING).....	8
1.1.1.2 JK-BILOT O RMP-BILOT (ROW METRIC PRESERVING) .	10
1.1.1.3 SQRT -BILOT o BILOT -SIMÉTRICO .....	11
1.1.1.4 HJ-BILOT o RCMP BILOT .....	12
1.1.1.5 AVANCES DEL BILOT .....	15
1.1.1.5.1 BILOT CLÁSICOS: GH BILOT Y JK BILOT .....	16
1.1.1.5.2 HJ-BILOT .....	16
1.1.1.5.3 BILOT ROBUSTO .....	18
1.1.1.5.4 MANOVA BILOT .....	20
1.1.1.5.5 MODELOS AMMI GGE BILOT .....	21
1.1.1.5.6 BILOT AMMI (EFECTOS PRINCIPALES ADITIVOS Y ANÁLISIS DE INTERACCIÓN MULTIPLICATIVA).....	22
1.1.1.5.7 GGE BILOT .....	23
1.1.1.5.8 GGEBILOTGUI PACKAGE .....	24
1.1.1.5.9 MODELO AMMI VERSUS GGE BILOT .....	24
1.1.1.5.10 BILOT LOGÍSTICO .....	26
1.1.1.5.11 BILOT LOGÍSTICO EXTERNO .....	26
1.1.1.5.12 BILOT PARA DATOS COMPOSICIONALES.....	27
1.1.1.5.13 VERSIÓN INFERENCIAL DEL BILOT .....	29
1.1.1.5.14 BILOT PARA TABLAS DE TRES VÍAS .....	31
1.1.1.5.15 BILOT DE CARLIER Y KROONENBERG.....	32
1.1.1.5.16 METABILOT .....	33
1.1.1.5.17 MULTIBILOT .....	34
1.1.1.5.18 BILOT DINÁMICO .....	35
1.1.1.5.19 TUCKER3.....	35
1.2 MÉTODOS PARA ANALIZAR TABLAS DE TRES VÍAS.....	37
1.2.1 STATIS/STATIS DUAL .....	37
1.2.1.1 AVANCES DEL STATIS .....	40

2.	APLICACIÓN A DATOS REALES.....	50
2.1	LA CALIDAD DEL AGUA EN PANAMÁ.....	50
2.2	ANTECEDENTES .....	52
2.3	OBJETIVOS.....	55
2.4	MUESTRA.....	56
2.5	DESCRIPCIÓN DE LA MUESTRA.....	60
2.6	CARACTERIZACIÓN MULTIVARIANTE DE LA INFLUENCIA DE VARIABLES FÍSICO-QUÍMICAS Y BIOLÓGICAS EN LA CALIDAD DEL AGUA: APLICACIÓN AL EMBALSE GATÚN (PANAMÁ).....	62
2.6.1	CLUSTERS SEGÚN LOS PUNTOS DE MUESTREO .....	69
2.7	STATIS DUAL COMO HERRAMIENTA EVALUATIVA PARA LA VALORACIÓN DE CALIDAD DEL AGUA.....	72
2.7.1	LA INTERESTRUCTURA .....	73
2.7.2	COMPROMISO .....	77
2.8	COMPARACIÓN DE LAS HERRAMIENTAS ESTADÍSTICAS UTILIZADAS: HJ-BIPLLOT Y STATIS DUAL.....	85
2.9	CONCLUSIONES.....	87
3.	BIBLIOGRAFÍA.....	88

## **RESUMEN**

---

La calidad del agua es un tema sensitivo y de preocupación a nivel mundial. Procesos naturales como cambios en la precipitación, erosión, desgaste de materiales incrustantes, etc., además de la influencia antropogénica urbana, industrial, actividades agrícolas, incremento en el consumo de fuentes de agua degradan las aguas superficiales.

La calidad del agua está definida por una serie de características físicas, químicas y biológicas. En los conjuntos de datos de calidad de agua, las variables o grupos de variables se relacionan, y desde magnitudes diferentes presentan similitudes. Esto podría ser porque más de una se rige por el mismo principio, permitiendo evaluar el funcionamiento en un sistema acuático. En el presente estudio se evaluaron las variables: temperatura, pH, transparencia, turbiedad, nitratos, ortofosfatos, fósforo y nitrógeno total, clorofila a, radiación solar, oxígeno disuelto y microcistinas o toxina que determinan la calidad de agua en los sitios de muestreo, Gamboa (643343E; 1007468N) y Paraíso (651144E; 997879N). Se realizaron mediciones en campo y se colectaron muestras (puntos de muestreo) para análisis de laboratorio. Los muestreos se realizaron durante los meses febrero a mayo y julio a diciembre de 2009. Un total de doce variables de calidad de agua fueron medidas.

Los resultados obtenidos a través de la aplicación de métodos de análisis multivariantes de dos y de tres vías tales como el HJ-biplot y Statis, logran la conformación de cluster de puntos de muestreo de acuerdo a los meses de la estación climática de la región, además de reflejar asociaciones de variables de composiciones físico-químicas y biológicas y de indicar las diferencias de los sitios, donde se efectuaron las colectas de las muestras.

**Palabras clave:** calidad de agua, BIPLLOT, STATIS.

## **INTRODUCCIÓN**

---

El presente trabajo es fruto de la colaboración del departamento de Estadística, los Programas de Maestría en Estadística Aplicada, y el de Maestría en Microbiología Ambiental (Facultad de Ciencias Naturales, Exactas Y Tecnología), el laboratorio de la Unidad de Calidad de Agua, División de Agua del Canal de Panamá y el departamento de Estadística (Universidad de Salamanca).

La Cuenca Hidrográfica del Canal es uno de los elementos hidrológicos más importante de la República de Panamá. La cantidad y calidad del agua está condicionada por el estado de los bosques y del entorno natural; por tanto, su administración requiere del conocimiento de la “calidad natural del agua” y de la dinámica del agua a través del ciclo hidrológico; lo que define las características propias que la hacen apta o no para su uso.

En los últimos años, el planeta ha experimentado cambios rápidos con relación al clima. Estos cambios, se han convertido en una de las principales preocupaciones mundiales. En Panamá hay un vínculo histórico con la administración del recurso hídrico desde que se concibió la idea de construir un canal interoceánico y es en el territorio de la Cuenca Hidrográfica del Canal donde se almacena el agua que permite el abastecimiento de agua para las ciudades más pobladas del país, para la operación del Canal de Panamá y para muchas otras actividades económicas, siendo el embalse Gatún la principal reserva de agua para la operación del Canal. El agua de esta zona facilita el movimiento transoceánico de un 6 por ciento del comercio mundial. Por ello, conscientes de esta situación nos planteamos el estudio de la evolución de la calidad del agua y de los parámetros asociados con ésta. Dicho estudio produce gran cantidad de datos cuyo manejo puede ser difícil mediante técnicas estadísticas convencionales. La naturaleza compleja que supone estudiar la calidad del agua requiere encontrar modelos simples para identificar las variables que tengan una mayor influencia en la calidad de la misma, por esta razón el uso de técnicas de análisis multivariante será de gran ayuda para



encontrar asociaciones y conclusiones que nos ayuden en la determinación del estado de la calidad del agua mediante indicadores biológicos y físico-químicos.

Así la estructura del trabajo es la siguiente:

- En el capítulo I se hace una revisión bibliográfica exhaustiva de los métodos para analizar tablas de dos y tres vías, que nos servirá para determinar las técnicas estadísticas más adecuadas a nuestros datos.
- En el capítulo II se contextualiza el objeto de estudio mediante un marco teórico en el que centraremos la calidad del agua en Panamá. A continuación se recoge todo lo relativo al tratamiento de datos, es decir, la forma en que se han recogido y preparado los datos para posibilitar el uso de las herramientas multivariantes, se exponen los resultados obtenidos, se hace una comparativa de las técnicas estadísticas utilizadas HJ-biplot y Statis-dual y finalmente se recogen las principales conclusiones del trabajo y bibliografía utilizada.

“Cuando bebes agua, acuérdate de la fuente”.

# **CAPÍTULO I**

## **MÉTODOS PARA ANALIZAR TABLAS DE DOS Y TRES VÍAS**

# 1. MÉTODOS PARA ANALIZAR TABLAS DE DE DOS Y TRES VÍAS

---

## 1.1 MÉTODOS PARA ANALIZAR TABLAS DE DOS VÍAS

Los métodos de análisis multivariante trabajan tradicionalmente con matrices de dos vías, las cuales contienen valores de  $p$  variables observadas en  $n$  individuos. Para poder caracterizar los individuos en función de las variables observadas, es necesario representar los individuos en un subespacio de dimensión reducida, generalmente de dimensión 2. Por ejemplo, si aplicamos un Análisis de Componente Principales, Biplot (GABRIEL, 1971), o HJ-Biplot (GALINDO-VILLARDÓN, 1986) si lo que queremos es representar conjuntamente individuos y variables.

### 1.1.1 MÉTODOS BILOT

Los métodos biplot fueron propuestos por GABRIEL (1971) como representaciones gráficas de datos multivariantes, es decir, al igual que un diagrama de dispersión muestra la distribución conjunta de dos variables, un biplot representa tres o más variables (GABRIEL y ODOROFF, 1990).

Las filas de la matriz son representadas mediante puntos (marcadores fila) y las columnas con vectores (marcadores columna), siguiendo la terminología introducida por el autor.

La *fundamentación teórica* de los biplot clásicos, se basa en la aproximación de la matriz de datos  $\mathbf{X}$  por una de menor rango  $q$ , siendo  $q < r$ , a través de la descomposición en valores singulares de  $\mathbf{X}$ . Se realiza luego, una factorización en matrices de marcadores filas y de marcadores columnas de manera tal que el producto escalar entre los marcadores aproxime 'lo mejor posible' los valores de  $\mathbf{X}$ .

Si la matriz  $\mathbf{X}$  es de *rango dos*, es posible lograr una representación *exacta* sobre el plano, en otro caso se necesitarán tantos ejes como rango tenga  $\mathbf{X}$ , para lograr un ajuste perfecto.

Cuando el rango de  $\mathbf{X}$  es *mayor a dos*, entonces la factorización biplot garantiza la representación gráfica *aproximada* de dicha matriz.

**DEFINICIÓN:** Un BILOT para una matriz de datos  $\mathbf{X}$  ( $I \times J$ ) es una representación gráfica mediante marcadores (vectores):  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_I$  para las as de  $\mathbf{X}$  y  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_J$  para las columnas de  $\mathbf{X}$ , de forma tal que el producto interno aproxime el elemento  $x_{ij}$  de la matriz de partida lo mejor posible.

Si consideramos los marcadores  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_I$  como filas de una matriz  $\mathbf{A}$  y los marcadores  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_J$  como filas de una matriz  $\mathbf{B}$ , entonces podemos escribir:

$$\mathbf{X} \cong \mathbf{A}\mathbf{B}'$$

Tanto los marcadores  $\mathbf{a}_i$  para las filas, como los marcadores  $\mathbf{b}_j$  para las columnas estarán representados en un espacio de dimensión  $q \leq r$ , siendo  $q$  el número de ejes retenidos y  $r$  el rango de  $\mathbf{X}$ .

La factorización biplot garantiza la representación gráfica aproximada de la matriz, ya que cada  $x_{ij}$  puede reconstruirse de la siguiente forma:  $x_{ij} = \mathbf{a}_i' \mathbf{b}_j$  ( $i = 1, 2, \dots, I; j = 1, 2, \dots, J$ ) obtenida al efectuar el producto interno de la fila  $i$  de  $\mathbf{A}$  y la columna  $j$  de  $\mathbf{B}'$ , admitiendo una representación gráfica a través de la proyección ortogonal de  $\mathbf{a}_i$  sobre  $\mathbf{b}_j$  y viceversa.

Considerando la aproximación BILOT de los elementos de la matriz  $\mathbf{X}$  y aplicando la definición de producto escalar tenemos que:

$$X_{ij} = \mathbf{a}_i' \mathbf{b}_j = \text{Pr oy}(\mathbf{a}_i / \mathbf{b}_j) (\text{signo}) \mathbf{b}_j, \text{ Siendo:}$$

$|| \text{Proy}(\mathbf{a}_i / \mathbf{b}_j) || =$  longitud del segmento que vá desde el origen de la representación hasta el punto  $\mathbf{a}_i$  (longitud de la proyección de  $\mathbf{a}_i$  sobre  $\mathbf{b}_j$ ).

$|| \mathbf{b}_j ||$  = módulo de  $\mathbf{b}_j$  ( longitud del segmento que une al origen con el extremo del vector  $\mathbf{b}_j$ ).

### **Obtención de los marcadores**

El método para la obtención de los marcadores de los diferentes métodos biplot; más conocido es el propuesto por ECKART y YOUNG (1939) que puede encontrarse también en YOUNG y HOUSEHOLDER (1938), GABRIEL (1971), GREENACRE (1984), entre otros autores.

El método está basado en la Descomposición en valores y vectores singulares de la matriz  $\mathbf{X}$  :  $\mathbf{X} = \mathbf{UDV}'$

Siendo :

$\mathbf{X}$  (I x J): la matriz de datos

r: es el rango de  $\mathbf{X}$  (  $r \leq \min (I, J)$  )

$\mathbf{U}$  (I x r) la matriz cuyas columnas contienen los vectores propios de  $\mathbf{XX}'$

$\mathbf{V}$  ( J x r) la matriz cuyas columnas contienen los vectores propios de  $\mathbf{X}'\mathbf{X}$ ,

$\mathbf{D}$  (r x r) es una matriz diagonal que contiene a los valores propios de  $\mathbf{X}$ .

Debe cumplirse que  $\mathbf{U}'\mathbf{U}=\mathbf{V}'\mathbf{V}=\mathbf{I}$ , es decir, las columnas de  $\mathbf{U}$  y  $\mathbf{V}$  son ortonormales, esta propiedad asegura la unicidad de la factorización.

'Siendo:

$$\mathbf{X} = \mathbf{AB}' = \mathbf{UDV}'$$

La selección de distintas factorizaciones conduce a distintos tipos de marcadores y por ende a distintos BIPLLOTS. Sin embargo, es importante

aclarar que, la interpretación del BILOT siempre se realiza a partir de los productos escalares, independientemente de la factorización elegida.

En el caso de la *métrica identidad*, se pueden elegir los marcadores de acuerdo a los distintos valores de  $\gamma$  en la siguiente descomposición:

$$\mathbf{A}=\mathbf{U}\mathbf{D}^\gamma \quad \mathbf{B}=\mathbf{V}\mathbf{D}^{1-\gamma} \quad (1)$$

Dependiendo del valor seleccionado para  $\gamma$  ( $\gamma = 0, 1, 1/2$ ) tenemos los BILOTS CLÁSICOS de GABRIEL (1971): el GH-Biplot, el JK-Biplot y el SQRT-Biplot, respectivamente.

### ***Propiedades de los marcadores***

Greenacre en 1984, introduce una nueva terminología para el GH y el JK BILOT, denominándolos respectivamente CMP (**C**olumn **M**etric **P**reserving) y RMP (**R**ow **M**etric **P**reserving), aludiendo al hecho que con el primer tipo se preserva la métrica para las columnas y con el segundo tipo la métrica para las filas. Esta terminología ha sido aceptada por Gabriel, razón por la cual en las referencias más recientes se opta por la misma.

#### ***1.1.1.1 GH-BILOT O CMP-BILOT (COLUMN METRIC PRESERVING)***

Cuando en la expresión  $\mathbf{A}=\mathbf{U}\mathbf{D}^\gamma \quad \mathbf{B}=\mathbf{V}\mathbf{D}^{1-\gamma}$   $\gamma = 0$  se trata del GH-BILOT o CMP-BILOT (Column Metric Preserving), luego tendremos lo siguiente:

$$\mathbf{X}=\mathbf{A} \mathbf{B}^\top = (\mathbf{U} \mathbf{D}^\gamma) (\mathbf{D}^{1-\gamma} \mathbf{V}^\top) = (\mathbf{U}) (\mathbf{D} \mathbf{V}^\top)$$

Luego la factorización en el CMP BILOT corresponde a la elección de los marcadores tal que:

$$\mathbf{A} = \mathbf{U} \mathbf{B} = \mathbf{V} \mathbf{D}$$

Como trabajamos con la métrica identidad en el espacio de las filas,  $\mathbf{A}$  debe verificar que:  $\mathbf{A}' \mathbf{A} = \mathbf{I}$  (por lo que la representación es única, excepto por rotaciones).

Luego:  $\mathbf{X}' \mathbf{X} = (\mathbf{A}\mathbf{B}')' \mathbf{A}\mathbf{B}' = \mathbf{B}\mathbf{A}'\mathbf{A}\mathbf{B}' = \mathbf{B}\mathbf{B}'$ .

Es decir:  $\mathbf{X}' \mathbf{X} = \mathbf{B}\mathbf{B}'$

Por lo tanto este BILOT, preserva la métrica euclídea usual entre las columnas pero no entre las filas, obteniéndose así para las columnas, una alta calidad de representación.

El nombre de este BILOT (**GH-BILOT**), se debe a que GABRIEL adoptó la notación  $\mathbf{G}$  para los marcadores fila y  $\mathbf{H}$  para los marcadores columna.

$\mathbf{G} = \mathbf{U} \quad \mathbf{H} = \mathbf{V} \mathbf{D}$

Al introducir un factor de escala, la matriz  $[(1/n-1)\mathbf{X}'\mathbf{X}]$  coincide exactamente con la matriz de covarianzas, y si se designan los marcadores como:

$\mathbf{A} = \sqrt{n-1} \mathbf{U} \quad \mathbf{B} = (1/\sqrt{n-1})(\mathbf{V} \mathbf{D})$

Obtenemos el biplot que se denomina Biplot de Componentes Principales, donde los productos escalares entre los marcadores columna, reproducen la estructura de las covarianzas entre variables.

### ***Calidad de representación del GH-biplot***

La calidad de representación global se calcula como el cociente entre la suma de cuadrados de  $\mathbf{X}(q)$  y la suma de cuadrados de los elementos de  $\mathbf{X}$ .

Dado que la suma de cuadrados de  $\mathbf{X}$  es igual a la traza de  $(\mathbf{X}'\mathbf{X})$ , y como:

$$Tr(\mathbf{X}' \mathbf{X}) = \sum_{k=1}^r \lambda_k = \sum_{k=1}^r \alpha_k^2$$

Donde  $\alpha_k$ , son los  $q$  mayores valores singulares asociados a los  $\lambda_k$  valores propios de la matriz  $\mathbf{X}'\mathbf{X}$  ( $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_q$ ).

Por lo que la calidad de representación global es:

$$CRG = \left( \frac{\sum_{k=1}^q \alpha_k^2}{\sum_{k=1}^r \alpha_k^2} \right) \times 100$$

De donde la calidad de representación de las columnas es:

$$CRC = \left( \frac{\sum_{k=1}^q \alpha_k^4}{\sum_{k=1}^r \alpha_k^4} \right) \times 100$$

Por lo tanto, la calidad de representación de las columnas de la matriz  $\mathbf{X}$  es óptima, en cambio para las filas es baja.

#### 1.1.1.2 JK-BILOT O RMP-BILOT (ROW METRIC PRESERVING)

Si  $y = 1$  en la expresión  $\mathbf{A} = \mathbf{U}\mathbf{D}^y$   $\mathbf{B} = \mathbf{V}\mathbf{D}^{1-y}$  se define el JK-BILOT o RMP-BILOT (Row Metric Preserving), luego tendremos lo siguiente:

$$\mathbf{X} = \mathbf{A}\mathbf{B}' = (\mathbf{U}\mathbf{D})(\mathbf{V}')$$

La elección de los marcadores es:

$$\mathbf{A} = \mathbf{U}\mathbf{D} \quad \mathbf{B} = \mathbf{V}$$

En este tipo de Biplot se impone la métrica  $\mathbf{B}'\mathbf{B} = \mathbf{I}$  en el espacio de las filas de la matriz  $\mathbf{X}_{(I \times J)}$ .

Considerando el tipo de factorización y la métrica tenemos:

$$\mathbf{X}\mathbf{X}' = \mathbf{A}\mathbf{B}'(\mathbf{A}\mathbf{B}')' = \mathbf{A}\mathbf{B}'\mathbf{B}\mathbf{A}' = \mathbf{A}\mathbf{A}'$$

Luego:

$$\mathbf{X}\mathbf{X}' = \mathbf{A}\mathbf{A}'$$



De allí el nombre de este tipo de BIPILOT (RMP-BIPILOT); el cual preserva la métrica euclídea usual entre las filas pero no entre las columnas, obteniéndose alta calidad de representación para las filas.

A este biplot Gabriel lo denominó JK-biplot porque utilizó **J** para denotar la matriz de marcadores fila y **K** para la matriz de marcadores columna.

$$\mathbf{J}=\mathbf{UD} \quad \mathbf{K}=\mathbf{V}$$

### ***Calidad de representación del JK-biplot***

La calidad de representación para las filas es *óptima*, en cambio para las columnas es *baja*.

#### **1.1.1.3 SQRT -BIPILOT o BIPILOT -SIMÉTRICO**

Si en la expresión  $\mathbf{A}=\mathbf{UD}^y$   $\mathbf{B}=\mathbf{VD}^{1-y}$ , se elige  $y= 1/2$ , se obtiene lo que GABRIEL denominó

SQRT-BIPILOT o BIPILOT-SIMÉTRICO, donde:

$$\mathbf{X} = \mathbf{A} \mathbf{B}' = (\mathbf{U} \mathbf{D}^{1/2})(\mathbf{D}^{1/2} \mathbf{V}')$$

Este BIPILOT asigna papeles simétricos tanto a las filas como a las columnas de **X**.

La elección de los marcadores es:

$$\mathbf{A} = \mathbf{U} \mathbf{D}^{1/2}$$

$$\mathbf{B} = \mathbf{V} \mathbf{D}^{1/2}$$

En este caso dicha factorización no es única ya que se verifica que  $\mathbf{A}'\mathbf{A} \neq \mathbf{I}$   $\mathbf{B}'\mathbf{B} \neq \mathbf{I}$ , razón por la cual este tipo de biplot solo se usa cuando el objetivo del análisis consiste fundamentalmente en la aproximación de los elementos de la

matriz  $\mathbf{X}$ , como en el caso de la diagnosis de modelos en tablas de contingencia donde el papel de filas y columnas es simétrico GOWER (1990).

#### **1.1.1.4 HJ-BILOT o RCMP BILOT**

El HJ-biplot a diferencia de los anteriores fue propuesto por GALINDO-VILLARDÓN (1985, 1986). Es una representación gráfica multivariante de marcadores fila y columna, elegidos de tal forma que puedan superponerse en el mismo sistema de referencia con máxima calidad de representación.

El HJ-biplot, tiene como objetivo principal la interpretación *simultánea* de las relaciones entre individuos (filas) y variables (columnas) de la tabla de datos  $\mathbf{X}$ , no siendo, a diferencia del GH-biplot y del JK-biplot, la búsqueda de la aproximación de los elementos de la matriz de datos  $\mathbf{X}$ .

Los elementos de la matriz  $\mathbf{X}$  están centrados por filas y columnas, por lo que la métrica introducida en el espacio de las filas es equivalente a la inversa de la matriz de covarianzas entre variables, mientras que en el espacio de las columnas la métrica es equivalente a la inversa de la matriz de dispersión entre individuos.

Si la DVS de  $\mathbf{X}$  es:  $\mathbf{X} = \mathbf{U} \mathbf{D} \mathbf{V}'$

Se elige, la siguiente factorización:  $\mathbf{A} = \mathbf{U} \mathbf{D}$      $\mathbf{B} = \mathbf{V} \mathbf{D}$

Dado de que en el HJ-biplot se puede hacer una representación simultánea de filas y columnas se lo denomina también RCMP-biplot (Row Column Metric Preserving).

GALINDO-VILLARDÓN (1986) lo denominó HJ-BILOT, ya que siguiendo la nomenclatura de Gabriel adoptó la siguiente notación para las matrices de marcadores fila y columna:  $\mathbf{J} = \mathbf{U} \mathbf{D}$      $\mathbf{H} = \mathbf{V} \mathbf{D}$

En el HJ-biplot, la representación de filas y columnas de la tabla de datos es simétrica, y se obtiene la misma calidad de representación tanto para los individuos (filas) como para las variables (columnas) (GALINDO-VILLARDÓN y CUADRAS, 1986). El HJ-biplot permite interpretar las posiciones de las filas, de las columnas y las relaciones fila-columna a través de los factores, como en el caso del Análisis Factorial de Correspondencias (BENZECRI, 1973; GREENACRE, 1984) pero a diferencia de éste puede ser aplicado a cualquier tipo de datos (no solo frecuencias).

También se pueden interpretar los ejes factoriales y su relación con las variables y con los individuos, a través de medidas o índices (GALINDO-VILLARDÓN y CUADRAS, 1986) a saber: las contribuciones relativas del factor al elemento y del elemento al factor.

En este biplot como los marcadores para las columnas coinciden con los marcadores columnas del GH-biplot y los marcadores para las filas coinciden con los del JK-biplot.

Los marcadores fila y columna se pueden representar en el mismo sistema de referencia.

Las relaciones entre las dos nubes que propone GALINDO-VILLARDON (1986) son a las del Análisis Factorial de Correspondencias, esto es: las relaciones que ligan a los vectores propios  $\mathbf{U}$  y  $\mathbf{V}$  son:  $\mathbf{U}=\mathbf{XVD}^{-1}$  y  $\mathbf{V}=\mathbf{X}'\mathbf{UD}^{-1}$ .

Por lo tanto las coordenadas para las filas son medias ponderadas de las columnas, donde las ponderaciones son los valores originales en la matriz  $\mathbf{X}$ . Lo mismo ocurre con las coordenadas de las columnas respecto de las filas.

El objetivo de los métodos BILOT es encontrar una matriz  $\mathbf{X}(q)$  de rango  $q$ , que aproxime lo mejor posible a la matriz de datos  $\mathbf{X}$  de rango  $r$ , siendo  $q < r$ . Si la matriz  $\mathbf{X}$  es de *rango dos*, es posible lograr una representación *exacta* sobre el plano, en otro caso se harán sólo aproximaciones  $\mathbf{X}$ , y en esos casos, que es

lo que sucede en la mayoría de las veces es importante evaluar la calidad de la representación.

### ***Calidad de representación del HJ-biplot***

Respecto a la calidad de representación este biplot posee una importante ventaja respecto a los anteriores ya que las filas y columnas tienen ambas una calidad de representación óptima.

### ***Índices o medidas de bondad de aproximación***

Los métodos biplot, como todo método del análisis multivariante, presentan los resultados del análisis en un subespacio que, generalmente, es de dimensión dos, aunque la configuración inicial sea de dimensión mayor.

Evidentemente, se hace necesario valorar la calidad de la representación en dimensión reducida. Para ello es importante tener en cuenta las propiedades de los marcadores, la interpretación geométrica y ciertos índices o medidas de bondad de la aproximación.

### ***Contribución relativa del factor al elemento (variable)***

Es el índice o medida que mide la relación entre los ejes de la representación biplot y cada una de las variables observadas y se simboliza CRFE.

La CRFE mide la parte de variabilidad de cada una de las variables explicada por el factor y permite conocer las variables que son las responsables de la ordenación de los individuos sobre las proyecciones en cada uno de los ejes biplot.

### ***Contribución del factor al elemento fila ó calidad de representación (CLR)***

Es el cociente entre el cuadrado de las coordenadas del  $i$ -ésimo punto fila sobre el eje  $\alpha$ , y la suma de cuadrados de las coordenadas de los  $I$  puntos fila sobre los  $q$  ejes ( $q < r$ ).

### ***Interpretación geométrica***

Estas contribuciones tienen una interpretación geométrica, ya que es precisamente el coseno al cuadrado del ángulo que forma dicho elemento con el factor. Cada punto, unido con el centro de gravedad, forma un ángulo con su proyección en el espacio de los  $q$  primeros ejes, de modo que el coseno al cuadrado de dicho ángulo es igual a la suma de las contribuciones relativas de dichos factores al elemento, ya que se denomina calidad de la representación.

#### ***1.1.1.5 AVANCES DEL BILOT***

CUBILLA-MONTILLA (2014), presenta las contribuciones al método BILOT, transcurridas más de cuatro décadas desde su formación.

Desde sus orígenes, los métodos biplot GABRIEL (1971), se han constituido en métodos de representación gráfica por excelencia. De esta manera, en la Estadística Multivariante, se suman cada día nuevas investigaciones que contribuyen aún más al desarrollo y aplicación de los métodos biplot con sus diversas variantes.

El biplot se ha utilizado tradicionalmente con fines descriptivos y también en la diagnosis de modelos (BRADU & GABRIEL, 1974, 1978). En la actualidad, los investigadores continúan realizando estudios utilizando la metodología BILOT y combinando ésta con otras técnicas de naturaleza clásica, como por ejemplo, el análisis de varianza y el análisis de componentes principales, entre otros.

Desde el punto de vista teórico, como práctico, surgen nuevas técnicas de análisis; y en consecuencia, líneas de investigación cuyos resultados quedan recogidos en un biplot.

De manera resumida se presenta las distintas contribuciones al biplot.

#### 1.1.1.5.1 BIPILOT CLÁSICOS: GH BIPILOT Y JK BIPILOT

Desde sus orígenes, los métodos biplot (GABRIEL, 1971) se han constituido en métodos de representación gráfica por excelencia. Estas contribuciones, denominadas GH-biplot y JK-biplot deben su nombre a la notación utilizada para ambos marcadores. Además de los nombres, estos métodos varían entre sí, porque presentan diferencias en cuanto a la bondad de ajuste.

El GH-biplot, debe su nombre a que Gabriel adaptó la notación G para simbolizar la matriz de los marcadores fila y H para la matriz de los marcadores columna.

Por su parte, en el JK-biplot, Gabriel representó los marcadores fila con la notación J y los marcadores columna con H.

GREENACRE (1984), establece una nueva terminología para el GH-biplot, denominada CMP-biplot (Column Metric Preserving) en torno al hecho de que preserva la métrica para las columnas; y, al JK-biplot, le denomina RMP-biplot (Row Metric Preserving), en función de que éste preserva la métrica para las filas.

Entre algunas de las aplicaciones de los biplots clásicos, se pueden citar, STRAUSS et al. (1979), GABRIEL & ODOROFF (1990), en el campo de la medicina; y TSIANCO & GABRIEL (1981, 1984), en el campo de la metodología.

#### 1.1.1.5.2 HJ-BIPILOT

Como una alternativa para optimizar los métodos biplot por GABRIEL (1971), GALINDO- VILLARDON (1986) plantea la técnica multivariante denominada HJ-BIPILOT; contribución que logra el objetivo de representar con máxima calidad de representación (GALINDO- VILLARDON & CUADRAS, 1986) y en forma simultánea las filas y columnas, sobre un mismo sistema de

coordinadas. De esta manera, es posible interpretar conjuntamente las relaciones entre individuos (filas) y variables (columnas).

En el HJ-biplot, facilita la interpretación de las posiciones de filas, columnas y relaciones filas-columnas, a través de los ejes; como lo hace el Análisis Factorial de Correspondencia (BENZECRI, 1973; GREENACRE, 1984).

Numerosos estudios reflejan la aplicación del HJ-biplot, desde el punto de vista teórico como práctico, en diferentes áreas del conocimiento. Desde su origen, en 1985 se han dado importantes desarrollos y aplicaciones en campos como la Economía, Educación, Genética, Meteorología, Sociología, Psicología, Biología, Política, Minería de datos, entre tantos otros.

Como ejemplos de aplicación del HJ-biplot, se presentan algunas contribuciones, con diferentes objetivos, de un gran número de estudios:

#### **Analysis of the sustainable society index worldwide: A study from the Biplot perspective**

GALLEGO-ÁLVAREZ et al. (2014) analizan, a través del HJ-biplot, la existencia de diferencias en las puntuaciones del índice de sostenibilidad obtenidas por una muestra de países según zona geográfica.

#### **The recent changes in the Portuguese Farm's Land Uses: An application of an Approach based on HJ-Biplot**

XAVIER et al. (2012) presentan un enfoque basado en el HJ-biplot para identificar las tendencias principales en términos de uso de la tierra a nivel municipal y generar políticas de uso de tierras.

#### **HJ-Biplot como herramienta de inspección de matrices de datos bibliométricos**

En este estudio, DÍAZ-FAES et al. (2013) caracterizan la producción científica de centros propios y mixtos, en relación a una serie de indicadores bibliométricos de impacto y colaboración científica.

#### **Phytoplankton and periphyton in ponds with Nile tilapia (*Oreochromis niloticus*) and bocachico (*Prochilodus magdalenae*)**

GARCÍA-GONZÁLEZ et al. (2012) evalúan, mediante el HJ-BILOT, el efecto del perifiton sobre la concentración (org/L) y el índice de diversidad algal Shannon Weaver (DI) del fitoplancton.

#### **El HJ-Biplot como herramienta en el análisis de grupos de discusión**

CABALLERO-JULIA (2011) demuestra la importancia del HJ-biplot como herramienta de análisis en grupos de discusión; y lo lleva a la práctica con datos de grupos de discusión en el contexto de la calidad de vida en ludopatía.

#### **Caracterización Multivariante de los perfiles de las mujeres en situación laboral irregular: el caso de salamanca**

GALINDO-VIILARDON et al. (2011) realizan un estudio en el que se identifican y comparan, a través del HJ-biplot, los perfiles socioeconómicos de las mujeres salmantinas bajo condiciones de irregularidad laboral, con las mujeres que se desenvuelven en un marco de actividad laboral regular.

##### **1.1.1.5.3 BILOT ROBUSTO**

Tradicionalmente, la descomposición en valores y vectores singulares se aproxima mediante el Análisis de Componentes Principales de la matriz  $X'X$ . No obstante, HERNÁNDEZ (2005) plantea que ambos procedimientos, tanto la descomposición en valores singulares, como el análisis de componentes principales, son susceptibles a la presencia de valores atípicos (outliers) y



propone una aproximación de los métodos biplot a través de modelo bilineales, al cual ha denominado BILOT ROBUSTO.

Este método, se constituye en una alternativa al biplot clásico, bajo la presencia de valores atípicos que conlleven un comportamiento discordante.

La idea del Biplot Robusto (HERNANDEZ, 2011) es obtener estimadores para la matriz de marcadores, así como también para la matriz de cargas a partir de un modelo bilineal (GOLLOB, 1968).

En síntesis, el biplot robusto presenta una mejora con respecto a los métodos Biplot, ya que resiste la presencia de valores extremos o atípicos, con lo cual se puede realizar un mejor análisis exploratorio de los datos. No obstante, el uso del biplot robusto es limitado, a continuación citamos algunas de sus aplicaciones.

#### **Análisis del Índice Metropolitano de la Calidad del Aire (IMECA) en una zona de México.**

HERNANDEZ & GALINDO- VILLARDÓN (2006) exploran el IMECA, con indicios de valores atípicos, para representar mediante un biplot robusto, los niveles de contaminación del aire en la Zona Metropolitana de México.

#### **Estudio comparativo del contenido nutrimental de los suelos en bosques de diferentes edades dentro del ecosistema húmedo tropical del Ecuador.**

TIPÁN & ADRIAN (2014) comparan las correlaciones entre nutrientes del suelo y las hojas, mediante el biplot robusto, para medir las relaciones entre la edad del bosque y la economía de hojas en la disponibilidad de nutrientes en suelos de bosques húmedos tropicales.

#### 1.1.1.5.4 MANOVA BILOT

El Manova Biplot de una vía (GABRIEL, 1972), así como también el llamado biplot canónico (VICENTE-VILLARDÓN, 1992) son propuestas orientadas a la obtención de representaciones ponderadas de la matriz de medias de los grupos, basadas en el biplot. El biplot canónico o manova biplot se concibe como una representación gráfica del análisis discriminante que incorpora información simultánea sobre los grupos y las variables; y, en consecuencia, separa los grupos con máximo poder discriminante a partir de las variables originales. Además de esta particularidad, la técnica permite realizar pruebas estadísticas en el contexto inferencial, sobre las variables canónicas.

AMARO et al. (2004) idean una generalización del manova biplot de dos vías (GABRIEL, 1972), en la cual la matriz de parámetros del modelo manova, se vale de cuatro representaciones biplot que recogen las diferentes fuentes de variación: Biplot Total, Biplot Interacción, Biplot Filas, Biplot Columnas.

Las aplicaciones del manova biplot, en comparación con los métodos clásicos y el HJ-biplot son escasas; quizás por el contexto en que debe ser aplicado este modelo, ya que entre otras condiciones, se requiere una estructura de grupos para su aplicación. Algunas de las publicaciones se detallan a continuación.

**Canonical Biplot Statistical Analysis to detect the Magnitude of the effects of Phosphates Crystallization aging on the color in Siliceous conglomerates.**

IÑIGO et al. (2013) aplicaron el biplot canónico en el estudio estadístico de los cambios en el color producido por dos tipos de procedimientos de envejecimiento artificial en cuatro variedades de conglomerados silíceos que tradicionalmente se han utilizado en la construcción y renovación en los edificios históricos de Zamora (España). Los resultados del estudio permitieron diferenciar el efecto de magnitud en el color de la superficie.

### **Manova Biplot para la composición de azúcares neutros en *Medicago Arborea* L.**

VICENTE-VILLARDON et al. (2007) muestran la aplicación del método manova biplot en un estudio de la composición de azúcares neutros, medios cromatografía de gases, en callos embriogénicos y no embriogénicos de *Medicago Arborea* L. Con estos resultados se precisa la importancia del método en la diferenciación y regeneración celular de plantas.

### **Role of Canonical Biplot method in the study of building stones; an example from Spanish monumental heritage.**

VARAS et al. (2005) aprovechan el biplot canónico para determinar el origen, la distribución, la conservación y el estado de la contaminación de los diferentes tipos de piedras utilizadas en la construcción y restauración de monumentos de gran interés histórico.

### **Manova Biplot para arreglos de tratamientos con dos factores basados en Modelos Lineales Generales Multivariantes**

AMARO et al. (2004) tomaron los datos de Seber (1984) para determinar el efecto de una infección de hongos sobre el crecimiento de un determinado tipo de césped, aplicando el método manova biplot.

#### **1.1.1.5.5 MODELOS AMMI GGE BILOT**

Usualmente, un gran número de genotipos se someten a pruebas, en diferentes ambientes y estaciones, lo que hace difícil determinar el patrón de respuesta genotípica, sin la ayuda de una visualización gráfica de los datos YAN et al. (2001).

Esta respuesta o reacción de los genotipos en condiciones ambientales diferentes es el resultado de la relación entre el genotipo (G) y el ambiente (E), y es denominada interacción genotipo ambiente (GEI).

Para evaluar genotipos probados en diferentes ambientes, se realizan análisis de adaptabilidad y estabilidad. La adaptabilidad es la capacidad del genotipo de responder positivamente a los efectos ambientales, garantizando un alto nivel de productividad (BECKER, 1981). La estabilidad está relacionada con el sostenimiento o subsistencia de la productividad en diferentes entornos. Existen diversas metodologías para estudiar la adaptabilidad y estabilidad fenotípica, las cuales difieren en cuanto a sus conceptos y procedimientos biométricos de estimación. Dentro de las metodologías planteadas, podemos destacar el análisis AMMI (additive Main Effects and Multiplicative Interaction) propuesto por MANDEL (1971) y el GGE Biplot (YAN et al., 2000). Ambas metodologías, se basan en los gráficos Biplot, por tanto, permiten representar gráficamente una matriz de datos.

#### 1.1.1.5.6 BILOT AMMI (EFECTOS PRINCIPALES ADITIVOS Y ANÁLISIS DE INTERACCIÓN MULTIPLICATIVA)

El modelo AMMI, estudiado también por ZOBEL et al. (1988) y validado por GAUCH JUNIOR & ZOBEL (1989) y CROSSA et al. (1990), combina el análisis de regresión lineal con el análisis de componentes principales (PCA), bajo los supuestos de que los efectos principales (genotipo y ambiente) son de naturaleza aditiva y la interacción genotipo x ambiente es de naturaleza multiplicativa.

El modelo AMMI primero realiza un Análisis de Varianza (ANOVA) con dos factores a partir de la matriz de medias para calcular los principales efectos aditivos de genotipo y ambiente. Luego, realiza un Análisis de Componentes Principales (ACP) sobre los residuos de este modelo aditivo para estudiar los efectos multiplicativos de la interacción (GAUCH, 1988).

El modelo AMMI genera un gráfico de dos dimensiones (biplot) en el que se pueden observar las diferencias entre ambientes (columnas), el grado de interacción de los genotipos (filas) con el ambiente, la estabilidad y las adaptaciones específicas de algunos genotipos a determinados ambientes.

#### 1.1.1.5.7 GGE BILOT

El biplot GGE es un modelo similar al Modelo AMMI, pero los términos lineales de genotipos no se consideran individualmente y se adicionan al término multiplicativo de la interacción genotipo x ambiente.

El GGE BILOT está basado en el modelo SREG (Modelo de Regresión de Sitios).

El modelo del GGE BILOT conserva unidos G y GE, y particiona GGE en dos términos multiplicativos.

La descomposición en valores singulares en el análisis GGE BILOT permite calcular los componentes principales, y también proporciona una medida de la variabilidad capturada por cada una de las componentes (YAN, 2002; YAN & TINKER, 2006). La primera componente, se encuentra altamente correlacionada con el efecto principal del genotipo; por tanto, representa la proporción del rendimiento que se debe sólo a las características del genotipo. La segunda componente representa la parte del rendimiento debido a la interacción genotipo-ambiente.

El BILOT GGE permite examinar, mediante la interacción genotipo x ambiente, la capacidad de discriminar y la representatividad de los ambientes de prueba como una medida conveniente para definir mega-ambientes homogéneos. (YAN et al., 2000; YAN & HUNT, 2002; YAN & KANG, 2003).

YAN (2001) señala que este gráfico permite identificar al genotipo ideal como aquel con alta puntuación en el primer eje del componente principal que está

asociada a altos rendimientos; y las puntuaciones cercanas a cero en el segundo eje del componente principal, relaciona con buena estabilidad el genotipo a través de los ambientes contrastantes.

#### 1.1.1.5.8 GGEBILOTGUI PACKAGE

FRUTOS, et al. (2013) hacen una propuesta computacional, en lenguaje R, como recurso interactivo para ejecutar las principales funciones de los BILOT CLÁSICOS y el GGE BILOT.

Este programa de computación que viene organizado en un paquete denominado GGEBILOTGUI; es una propuesta interactiva, no comercial y de código abierto, que se ofrece como una alternativa al software comercial disponible. Es una herramienta útil para la representación gráfica de datos procedentes de diferentes áreas; en particular, de gran utilidad para genetistas, a fin de estudiar los rendimientos de genotipos y la interacción entre el genotipo y el medio ambiente.

En el GGEBILOTGUI, los genotipos son representados como "puntos" y los ambientes como "vectores".

#### 1.1.1.5.9 MODELO AMMI VERSUS GGE BILOT

El uso en conjunto de las metodologías AMMI y GGE BILOT torna más eficaz la exploración de la interacción genotipo ambiente. Estudios recientes demuestran que cuando ambas metodologías son utilizadas adecuadamente, se obtienen resultados eficientes (CAMARGO-BUITRAGO et al., 2011; YAN, 2011). No obstante, algunos autores mantienen posiciones encontradas en cuando al uso de una u otra metodología.

GAUCH (2006) critica el modelo GGE BIPLLOT porque este no separa los efectos del G y GE, como lo hace el análisis AMMI. Sin embargo, YAN et al. (2007) discrepan con fitotecnistas que muestran interés en estudios basados en G y GE simultáneamente.

Algunas aplicaciones:

### **Métodos estadísticos multivariantes en el estudio de la interacción genotipo ambiente en caña de azúcar.**

GROSS et al. (2012) aplican AMMI para comparar su eficiencia en el estudio de la interacción genotipo ambiente y la estabilidad fenotípica en cultivares de caña de azúcar. A la vez comparan estos resultados con los obtenidos mediante el Análisis de Coordenadas Principales y el Análisis de Regresión de Sitios.

### **Genotype X Environment Interaction in Canola ( Brassica napus L). Seed Yield in Chile.**

ESCOBAR et al. (2011) diseñan un experimento que explora la interacción genotipo x ambiente en raps en Chile, utilizando el modelo SREG (regresión de sitios).

### **Análisis gráfico mediante biplot del comportamiento de híbridos de maíz.**

IBÁÑEZ et al. (2006) utilizan el modelo de regresión de sitios para evaluar la variabilidad en el rendimiento del grano de maíz y determinar los patrones de respuesta entre híbridos y ambientes.

### **Modelos con término multiplicativo. Aplicación en el análisis de la interacción genotipo ambiente.**

VARELA & CASTILLO (2005) consideran el rendimiento de 10 variedades de papa, en tres localidades, durante tres años, para evaluar la interacción genotipo ambiente.

#### 1.1.1.5.10 BILOT LOGISTICO

Como un procedimiento alternativo ante escenarios en los que se cuenta con datos binarios (VICENTE-VILLARDÓN, 2001; VICENTE-VILLARDÓN et al., 2003, 2006) proponen un nuevo enfoque biplot, denominado Biplot Logístico, basado en un modelo en el cual las coordenadas de los individuos y las variables están previstas para obtener respuestas de tipo logístico.

VICENTE-VILLARDON et al. (2006) fijan los parámetros del biplot logístico en un esquema iterativo que alterna A y B.

A y B son las matrices que contienen los marcadores para las filas y columnas de la matriz de datos  $X_{n \times p}$ , en la cual las filas corresponden a los individuos, y las columnas miden atributos o datos cualitativos que se asocian a variables binarias.

En esencia, un conjunto de parámetros es introducido, mientras que el otro se mantiene fijo, y este procedimiento se repite hasta que la probabilidad converge a un grado de precisión deseado.

Como ejemplo se cita, la investigación, “el uso de información externa en aproximación Biplot”, desarrollada por CARDENAS & GALINDO-VILLARDON (2003) en la cual ajustan un biplot logístico con información externa, a datos dicotómicos correspondientes a granjas. Los investigadores realizan una predicción de presencia-ausencia de cada variable sobre las granjas presentes en el estudio.

#### 1.1.1.5.11 BILOT LOGÍSTICO EXTERNO

Siguiendo la línea de pensamiento de VICENTE-VILLARDON et al. (2006), surge una variante al biplot logístico, DEMEY et al. (2008), que combina en un mismo algoritmo, el análisis de coordenadas principales y la regresión logística, para construir la técnica conocida como un Biplot Logístico Externo. La



propuesta se basa en el hecho de que la regresión en el procedimiento alternativo para datos binarios, no es más que una regresión logística que se puede articular convenientemente, a la configuración obtenida a partir del Análisis de Coordenadas Principales.

En base a su experiencia, los autores de esta propuesta, señalan que este método es más simple, más accesible y confiere resultados similares al biplot logístico con en diferentes escenarios.

En efecto, GALLEGO-ÁVAREZ & VICENTE-VILLARDON (2012) comprueban mediante la técnica de biplot logístico externo, el creciente interés de las empresas por atender los problemas de emisión de gases de efecto invernadero (aplicación medioambiental).

VEIGA DE CABO & MARTÍN-RODERO (2011) prueban mediante esta técnica, el grado de implementación de las tecnologías web 2.0 en revistas científicas médicas de acceso abierto; como la intención de fomentar la innovación, el crecimiento económico y el desarrollo social (aplicación en innovación). Por su parte VICENTE-GALINDO et al. (2010) aplicaron esta metodología en el campo de la economía; concluyendo que este método es altamente efectivo para los empresarios y/o administradores, que necesitan contar con herramientas prácticas y apropiadas para proyectar sus acciones y decisiones empresariales (aplicación en la economía).

#### 1.1.1.5.12 BILOT PARA DATOS COMPOSICIONALES

Los datos composicionales están presentes en diferentes campos y/o disciplinas (RODRÍGUEZ & SÁNCHEZ, 2013) como por ejemplo, en Biología para caracterizar la fauna existente en un determinado hábitat; en Economía, para identificar partidas con el fin de identificar la inversión de los recursos económicos; en Psicología, partiendo de la inteligencia global, es posible analizar cierto tipo de inteligencia y sus correspondientes variables.

En este sentido, en la década del 80, AITCHISON (1982, 1986) propuso una metodología para el análisis de datos composicionales, trabajo que ha recibido valiosos aportes y reformulaciones de otros autores intentando afrontar las dificultades de interpretación de las magnitudes absolutas de las partes. No obstante, AITCHISON (1997) revela algunos principios básicos a los que debe responder el análisis de datos composicionales. Estos principios son: invariancia por escala y coherencia subcomposicional.

A partir del año 2000 se promueven diversos aportes en los aspectos formales del análisis (PAWLOWSKY-GLAHN & EGOZCUE 2001, BILLHEIMER et al. 2001, AITCHISON et al. 2002), que han permitido un mejor tratamiento de los métodos ya propuestos.

Básicamente, el análisis de datos composicionales puede resumirse en tres pasos:

- La transformación de los datos a coordenadas de tipo log-cociente.
- El análisis estadístico (usual) de dichas coordenadas como variables reales; y
- La interpretación de los modelos obtenidos en las propias coordenadas o volviendo a expresar los resultados en términos de composiciones.

AITCHISON & GREENACRE (2002) llamaron al biplot derivado de la matriz de datos composicionales, "BILOT de variación relativa" en vista de que este constituye la variación en todas las relaciones de los componentes.

Algunas aplicaciones de los datos composicionales:

RODRÍGUEZ & SÁNCHEZ (2013) emplean la metodología para datos composicionales en el análisis de un cuestionario de Calidad de Vida Relacionada con la Salud en pacientes con osteoporosis.

**El análisis composicional en el estudio de la producción y distribución de la cerámica prehispánica.**

GHEZZI (2011) aborda el análisis composicional como método para definir la procedencia, arqueología de la muestra de cerámica prehispánica.

### **Análisis composicional de datos en Ciencias Geoambientales.**

EGOZCUE & PAWLOWSKY (2011) demuestran con un análisis de datos hidroquímicos, uso de (log) cocientes para evitar la interpretación individual de concentraciones desatendiendo su carácter relativo.

### **Análisis estadístico de datos geoquímicos de volcánicas jurásicas del macizo del deseado, provincia de santa Cruz.**

(ALPERIN et al. 2007) para analizar las relaciones entre óxidos y entre unidades volcánicas, examinan la existencia de subpoblaciones a través de gráficos biplot de datos composicionales como herramientas exploratoria.

### **Clasificación no paramétrica de datos composicionales.**

FERNÁNDEZ et al. (2003) mediante la aplicación de la metodología para datos composicionales, realizan una clasificación automática no paramétrica de las 41 comarcas catalanas con el fin de analizar la existencia de grupos de comarcas que sean similares en relación a la distribución de su población activa.

### **The use logratios in subcompositional analysis and geochemical discrimination of metamorphosed limestones.**

THOMAS & AITCHISON (1998) estudian qué óxidos son más efectivos a la hora de discriminar entre dos tipos de calizas.

#### **1.1.1.5.13 VERSIÓN INFERENCIAL DEL BILOT**

Los métodos biplot se constituyen en una poderosa herramienta de representación gráfica, en un espacio de dos dimensiones, para explicar la

variación conjunta de variables e individuos. La interpretación de un biplot se hace a partir de los marcadores filas y marcadores columna, quienes definen parámetros de estimación representados como puntos y vectores en el gráfico biplot. Siendo así, los parámetros se estiman como valores puntuales en el biplot, por lo que no hay acerca de la precisión de los estimadores. Es por ello, que nace la idea de una versión inferencial del biplot para los citados parámetros.

Conocer con qué precisión se han analizado los datos, es tarea de la estadística inferencial. Ahora bien, conocer esta precisión a través de la metodología biplot, no ha sido tarea fácil. EFRON (1979) propuso una metodología para el estudio de los cálculos de la estadística inferencial, a través del método conocido como "BOOTSTRAP".

Recordando que la distribución de muestreo de un estadístico es clave a la hora de realizar análisis inferenciales. BOOTSTRAP es un método de remuestreo utilizado con frecuencia para aproximar el error estándar o la varianza de un estimador, y construir así intervalos de confianza o realizar contrastes de hipótesis sobre parámetros de interés. En lugar de fórmulas o modelos matemáticos abstractos, el Bootstrap simplemente requiere un ordenador capaz de simular un proceso de muestreo aleatorio de los datos.

El BOOTSTRAP no es rígido respecto al cumplimiento de supuestos teóricos, en tal sentido es menos restrictivo que las técnicas convencionales. En su aplicación, en vez de asumir a priori una determinada distribución teórica, se utiliza la muestra original y se generan un gran número de sub-muestras que sirven de base para estimar la distribución de los datos. De esta manera, pueden analizarse datos provenientes de distribuciones desconocidas o incluso abordarse situaciones, frente a los cuales no hay una solución analítica conocida (EFRON y TIBSHIRANI, 1993).

Se pueden encontrar aplicaciones de la metodología Bootstrap en diversas áreas como por ejemplo, el sector financiero, ingeniería, control de calidad,

medicina, ciencias biológicas estadística espacial, series de tiempo, entre otras. Veamos algunas de estas aplicaciones.

### **Efecto del tamaño de muestra y el número de réplicas bootstrap**

RAMÍREZ et al. (2013) evalúan el efecto del tamaño de la muestra y el número de réplicas Bootstrap en las estimaciones de la media, varianza y algunos percentiles para una distribución de probabilidad bajo diferentes medidas de asimetría. Además, presentan algunas recomendaciones para el uso del BOOTSTRAP no paramétrico.

### **Método Bootstrap de factores estratificados**

KIKUCHI (2010) presenta una variante del método BOOSTSTRAP para efectos de calcular el capital de solvencia necesario para cubrir la variabilidad que pudiera presentarse en las reservas de siniestros y los gastos asociados en una reaseguradora, estratificando la variabilidad de los factores por año de desarrollo.

### **Modelos de Clases Latentes en tablas poco ocupadas: una contribución basada en Bootstrap**

ARAYA (2010) determina el poder del Método BOOSTSTRAP No paramétrico en la diagnosis de modelos de clases latentes y propone el BOOSTSTRAP No paramétrico Corregido y el Método de Patrones para el diagnóstico de modelos de clases latentes en tablas poco ocupadas.

#### **1.1.1.5.14 BILOT PARA TABLAS DE TRES VÍAS**

Al considerar biplot para tablas de tres vías, la complejidad de la estructura de los datos aumenta; puesto que se requiere clasificarlos de acuerdo a tres tipos de información: individuos, variables y situaciones.

En las últimas décadas se han desarrollado diferentes métodos para analizar este tipo de tablas, algunos desde la perspectiva anglosajona y otros desde la óptica francesa.

Sin embargo, todos son métodos de reconocida importancia y/o trayectoria, cuyas diferencias se deben básicamente al punto de vista teórico.

El análisis de datos de tres vías tiene su origen en la década de los sesenta con los trabajos de TUCKER (1964), con el análisis de factor de tres modos, actualmente conocido como el análisis de componentes de tres modos (TUCKER, 1966).

Vamos a explorar diferentes técnicas para el tratamiento de tablas de tres vías, exclusivamente aquellas basadas en la proyección sobre la representación biplot.

De manera resumida se va a presentar diferentes técnicas para el tratamiento de tablas de tres vías, aquellas basadas en la proyección sobre la representación biplot.

#### 1.1.1.5.15 BILOT DE CARLIER Y KROONENBERG

Desde una perspectiva teórica asociada a modelos de tres vías, nace la idea de integrar datos estructurados, a partir de la generalización de la descomposición en valores y vectores singulares, GDVS (CARLIER Y KROONENBERG, 1996). Esta descomposición, apoyada en los modelos de tres vías, TUCKER (1966) y TUCKALS3 (1989); sugiere dos formas de representaciones BILOT para absorber la información contenida en la tabla de datos: BILOT INTERACTIVO" y "BILOT CONJUNTO". Para construir estos Biplot, es necesario utilizar la descomposición de tablas de tres vías asociadas con los modelos de Tucker. El método Tucker3 es una generación del modelo de componentes de dos vías al modelo de componentes de tres vías.

El BILOT INTERACTIVO (BRADU Y GRABIEL, 1978; COX Y GABRIEL, 1982; VAN EEUWIJK Y KROONENBERG, 1998) parte del modelo Tucker3 y se fundamenta en la combinación de dos de los modos (J y K). Este gráfico refleja diferentes comportamientos entre los elementos que corresponden a los resultados en el análisis de una matriz denominada, matriz de enlace. El biplot interactivo permite la visualización de la interestructura de las diferencias tablas de datos.

El biplot conjunto (KROONENBERG, 1983) parte del modelo de Tucker 3 y es un biplot condicional **a uno de los modos**; a través del plano  $G_r$ , los ejes del BILOT conjunto se reajustarán visual de la intraestructura de cada tabla.

#### 1.1.1.5.16 METABILOT

La capacidad de la representación BILOT, ha hecho posible utilizar esta metodología para integrar información de varias matrices de datos y comparar varios biplot. MARTIN-RODRÍGUEZ et al. (2002) formulan el análisis de más de dos BILOT, como resultado de su aplicación al estudio de varios individuos con las mismas variables, en momentos o situaciones diferentes. Esta importante contribución, denominada, "METABILOT", sugiere que a partir de diferentes matrices de datos de un estudio longitudinal, se pueda comparar e integrar subespacios resultantes a partir de los marcadores de un GH-biplot o un JK-biplot, a fin de obtener una máxima similitud entre ellos. El criterio de similitud empleado es el coseno del ángulo formado por los componentes del vector de los marcadores biplot obtenidos.

Encontramos algunos estudios que se han presentado en eventos académico, en donde se ha aplicado el Metabiplot; los cuales citamos a continuación:

### **Metabiplot: Aplicación al estudio de variables económicas en la CEE**

Martín-Rodríguez y Vicente Villardon (2004). Analizan algunas variables económicas que permiten ver en el espacio consenso, la evolución de los países a lo largo de los años y el posicionamiento de las variables.

### **Statis e Metabiplot. Un estudio comparativo**

MORGADO Y GALINDO VILLARDON (2006). Analizan el índice metropolitano de calidad del Aire (IMECA), en México, para comparar las regiones en estudio y estudiar la existencia de una matriz consenso. En el trabajo hacen una comparación con el método STATIS.

### **Metabiplot: Application to the Study of a lake**

MARTÍN-RODRÍGUEZ Y VICENTE VILLARDON (2004) analizan el comportamiento de nutrientes en un lago, mediante la representación en un espacio consenso, después de integrar estructuras y hacer análisis comparativos.

#### 1.1.1.5.17 MULTIBILOT

Los datos de conjuntos múltiples representan un caso particular de datos de tres vías.

Las técnicas de análisis de datos múltiples vías, ampliamente utilizadas en muchas áreas de la investigación, representan una poderosa herramienta para analizar las relaciones entre individuos, variables y situaciones experimentales. Sin embargo, algunas técnicas presentan limitantes en el manejo de las matrices, pues trabajan con productos cruzados; y, en otros casos la limitante es la representación conjunta de individuos y variables, para lo cual suele elegir el momento de óptima representación.



Ante estas limitantes, BACCALÁ (2004), presenta una aplicación alternativa de análisis de datos de conjuntos múltiples, a partir de los Métodos BIPLLOT de GABRIEL (1971). Esta aplicación aborda, en conjunto situaciones distintas que otros métodos plantean por separado. Baccala, propone el método, denominado "MULTIBIPLLOT", que puede ser utilizado en datos de conjuntos múltiples, bajo dos situaciones:

- Varios conjuntos de individuos observados sobre un único conjunto de variables.
- Varios conjuntos de variables observadas sobre un único conjunto de individuos.

#### 1.1.1.5.18 BIPLLOT DINÁMICO

Una nueva corriente ha surgido para el análisis de tablas de tres entra denominada BIPLLOT DINÁMICO (EGIDO, 2013). Esta corriente subyace en el principio general de tablas de tres vías: se tienen  $X_i$  matrices de datos provenientes de "p" variables, referidas a un conjunto de "n" individuos, los cuales hemos representado mediante un cubo.

Finalmente en el biplot dinámico obtenido, se proyecta la trayectoria para cada matriz obtenida del cubo. La distancia de cada uno de los puntos al origen, corresponde a la covarianza entre la variable en el momento de referencia y la situación representada.

#### 1.1.1.5.19 TUCKER3

El modelo Tucker3 descompone el arreglo X en cuatro conjuntos de parámetros, teniendo en cuenta la extracción de diferente número de componentes en cada uno de los modos.

Se inicia con una descomposición completa del arreglo tridimensional  $X$  en tres matrices ortonormales. Los parámetros corresponden a:

- Una matriz  $A$  de dimensión  $I \times P$  que contiene los coeficientes para los componentes del modo individuos;
- Una matriz  $B$  de tamaño  $J \times Q$  que contiene los coeficientes para los componentes del modo variables;
- Una matriz  $C$  de tamaño  $K \times R$  que contiene los coeficientes para los componentes del modo ocasiones; y
- Una matriz central  $G$  de dimensión  $P \times Q \times R$  denominada, matriz de enlace; en ella se encuentran las interrelaciones entre las respectivas direcciones de inercia de cada modo, generad en base a las tres matrices componentes.

El objetivo principal del modelo es estimar los parámetros  $A$ ,  $B$ ,  $C$  y  $G$ . En su expresión general el modelo Tucker3 se define a partir de la generalización del modelo de componentes principales para dos vías.

## 1.2 MÉTODOS PARA ANALIZAR TABLAS DE TRES VÍAS

Cuando los individuos y variables están medidas en distintos lugares o en distintos momentos en el tiempo, se tendrá en cada lugar o en cada momento, una tabla de dos vías y por tanto, la información se presenta en varias tablas de individuos x variables, es decir, cada observación es originada por tres modos: individuos, variables y condiciones/tiempos.

Existen varios tipos de datos de tres vías:

- Datos de Tres Vías: se considera un solo conjunto de individuos, un solo conjunto de variables y un solo conjunto de condiciones.
- Datos de Conjuntos Múltiples:
  - Varios conjuntos de individuos, un solo conjunto de variables y un solo condiciones, en tal caso, en cada condición se miden las mismas variables en diferentes individuos.
  - Varios conjuntos de variables y un mismo conjunto de individuos y condiciones; es decir, en cada condición se miden diferentes variables a los mismos individuos

### 1.2.1 STATIS/STATIS DUAL

El STATIS o STATIS DUAL, incluye un estudio de la inter-estructura, o sea, de la relación entre las distintas condiciones o matrices, la determinación de una matriz compromiso, y un estudio de la intra-estructura, eso es, de las relaciones entre individuos o entre variables dentro de las distintas condiciones o matrices.

El método STATIS (Structuration de Tableaux À Trois Indices de la Statistique) y STATIS DUAL, son métodos de análisis de datos que han sido desarrollados para extraer la información relevante almacenada en tablas de tres vías propuesto por L'HERMIER DES PLANTES (1976), LAVIT (1988), y LAVIT et al. (1994), ESCOUFIER (1973, 1976).

En síntesis, el método STATIS consiste en determinar una matriz compromiso que es la más próxima de todas las matrices originales, y finalmente, todas estas matrices originales son proyectadas en el espacio de componentes principales de matriz compromiso. Estas operaciones, sin embargo, no se hacen directamente con las matrices, sino con operadores.

En el caso donde el interés es comparar los individuos, se aplica el método STATIS; en la situación donde el interés es estudiar la relación entre las variables, se aplica el método STATIS DUAL.

Cuando los individuos y las variables se mantienen en todas las matrices, los dos métodos son aplicables.

El STATIS, se centra en los individuos, partimos de la matriz  $W_t = X_t X_t'$  ó  $W_t = X_t' M X_t'$ , es el producto interno entre individuos.

El STATIS DUAL, se centra en las variables  $C_t = X_t' X_t = X_t' D_t X_t$  es la matriz de correlaciones entre las variables.

Donde:  $X_t$  es la información para cada tiempo o cada grupo

$M_t$ : matriz diagonal de "pesos de las variables en la tabla  $X_t$

$D_t$  : matriz diagonal de "pesos de los individuos" en la tabla  $X_t$

En el STATIS DUAL, la estructura en cada tabla ( $X_t$ ) puede ser capturada por la configuración  $C_t = X_t' X_t$ ; y para comparar estas configuraciones  $C_t$  debemos utilizar un coeficiente de correlación entre matrices que es el producto interno de Hilber-Schmidt (HS):  $\langle C_t | C_t' \rangle = \text{tr}(C_t C_t')$

## **ETAPAS DEL STATIS/STATIS DUAL**

*Primera Etapa:* Estudio de la INTER-ESTRUCTURA.

Una vez que tenemos la matriz de correlaciones podemos obtener los valores y vectores propios de esta matriz para reducir la dimensionalidad y producir una

imagen euclídea, donde cada matriz de partida vendrá representada como punto en un espacio euclídeo de baja dimensión.

*Segunda Etapa: Búsqueda de un COMPROMISO.*

Consiste en resumir las T tablas de datos en una sola denominada "compromiso". El compromiso constituye un resumen global del conjunto de tablas. Esta matriz compromiso debe ser la más correlacionada con todas las configuraciones (en el sentido de correlación vectorial).

La matriz compromiso, debe ser la configuración más correlacionada (en el sentido del producto escalar de Hilbert-Schmidt) con todas y cada una de las configuraciones.

Básicamente la configuración compromiso es la media ponderada de las configuraciones en cada ocasión, donde las ponderaciones son las componentes del primer vector propio de la matriz de correlaciones vectoriales.

*Tercera Etapa: Estudio de la INTRA-ESTRUCTURA.*

El compromiso obtenido en la etapa anterior permite representar las "posiciones compromiso" de cada uno de los elementos (individuos y/o variables) que constituyen las diferentes tablas.

Las posiciones compromiso de los elementos corresponden a las posiciones medias de los mismos.

El estudio de la Intra-estructura es un análisis detallado de las diferentes tablas, ya que permite estudiar las semejanzas o diferencias entre los elementos (individuos o variables) correspondientes a las diferentes tablas.

La representación de los puntos individuos en las diferentes tablas nos permitirá representar una "trayectoria de individuos" o el comportamiento de cada individuo a través de las diferentes tablas.

Si la trayectoria para un individuo está poco extendida, se corresponderá con un individuo cuya “evolución” en las diferentes tablas es similar a la evolución media. Inversamente, una trayectoria de gran amplitud refleja un cambio de estructura del individuo de tabla en tabla, diferente a la evolución media.

### **1.2.1.1 AVANCES DEL STATIS**

VICENTE-GALINDO (2013), presenta del STATIS aspectos relacionados a su situación actual e impacto y desarrollo.

De acuerdo a la situación actual:

- El STATIS es la técnica de Análisis de Datos de tres vías más utilizada en la práctica tanto si se trata de analizar la parte estable de las estructuras como si no; pero no es la técnica óptima cuando se trata de estudiar la parte cambiante de las estructuras y/ las interacciones.
- Todas las alternativas al STATIS trabajan con la correlación vectorial de Hilbert Schmidt; y con pesos relacionados con los factores de carga del primer vector propio de la descomposición espectral de la matriz de correlaciones vectoriales, excepto el Power STATIS.
- La limitación importante del STATIS; radica en que la ponderación utilizada pondera a todas las variables de una matriz que está asociado al compromiso, por un valor inicial que amplifica su valor original, haya o no participado en esa asociación. Sucede a la inversa con las variables de las matrices no asociadas al compromiso.

Con relación al impacto se tiene que:

- El STATIS, a pesar de sus limitaciones, ha despertado un gran

interés, tanto a nivel de desarrollos teóricos como a nivel de aplicaciones prácticas, en campos muy diversos.

- Las aplicaciones del STATIS a distintos campos de la ciencia se cuentan por decenas: en quimiometría, en ecología, en medicina, en calidad de vida relacionada con la salud, en control estadístico de calidad, en biología molecular, etc.
- La primera aplicación del STATIS, que algunos autores dan como cita original donde se introduce el STATIS fue una aplicación al estudio de datos ambientales (pluviosidad), realizado por L'HERMIER DES PLANTES y THIÉBAUT (1977) y publicado en *Revue de Statistique Appliquée*.
- Resaltamos algunas de las más recientes publicaciones en revistas de impacto: PIZARRO et al. (2013); OKSUZ et al. (2013); FERRARO et al. (2013); BENASSENI y DOSSE (2012); ABDI et al. (2012); THIOULOUSE (2011); KRISHNAN et al. (2011); STANIMIROVA et al. (2011); DASZKOWSKI y WALCZAK (2011); WILCOX (2010); STEFANESCU et al. (2010); IACHINE et al. (2010); REGNAULT et al. 2009; GUEBEL et al. (2009); VIVIEN y SUNE (2009); VAN DEUN et al. (2009); POULIOT et al. (2009); ROLLAND et al. (2009); MANUEL-GONZALEZ et al. (2009); FOURNIER et al. (2009); ERSAN et al. (2009); FOGLIATTO y NIAN (2009); SABATIER y VIVIEN (2008); PERRIN et al. (2008); SERGHINI et al. (2008); ARCIDIACONO, et al. (2008).

Los desarrollos del STATIS se pueden clasificar de la siguiente manera:

- De acuerdo a los datos de partida: Cuando todas las tablas de datos contienen información sobre los mismos individuos, en diferentes tiempos o situaciones, se utiliza el X-STATIS o Análisis Parcial Triádico (PTA) (JAFFRENOU, 1978). En este caso se trabaja con las matrices de datos y con operadores. Cuando se trabaja con los mismos sujetos y las mismas variables pero se parte de una matriz de operadores, tenemos el

COVSTATIS (THIOULOUSE, 2011), y el DISTATIS (ABDI y col, 2007). Para el caso en que los individuos tengan estructura de grupos, VALLEJO-ARBOLEDA et al, (2007) desarrollan el CANOSTATIS.

- De acuerdo a los pesos asignados al crear la matriz consenso: BENASSENI & BENNANI-DOSSE, (2012) llaman Power-STATIS a una versión de STATIS en la cual el peso que se da a cada matriz difiere de la propuesta de L'HERMIER.
- Si se tiene en cuenta información externa: t+1 STATIS (SAUZAY y col., 2006) consideran las t matrices que se integran en el STATIS y una matriz más con información externa sobre los individuos. En 2004, Vivien y Sabatier presentan una extensión del STATIS para datos multibloque, a la que llaman DO-ACT para dos tablas multibloque que no necesariamente tienen que tener la misma longitud. En 2008, Sabatier y Vivien proponen el STATIS-4. Si se dispone de pares de tablas en t situaciones o tiempos diferentes, STATICO (STATIS y COINERCIA), SIMIER et al., (1999); THIOULOUSE et al., (2004) y más recientemente COSTATIS (COINERCIA Y STATIS), THIOULOUSE, (2011), suponen interesantes alternativas que permiten estudiar las co-estructuras.

Los métodos abordan la construcción de la matriz consenso de diferentes formas:

a) En algunos métodos, para la construcción del consenso se trabaja con la yuxtaposición de las matrices de datos originales ponderadas de diferentes formas y es, en general, esta ponderación la que los diferencia. Dichas ponderaciones surgen a partir de las matrices de productos cruzados construidas con las matrices originales o de T análisis factoriales realizados en cada una de las tablas. El método Statis y Statis Dual (L'HERMIER des PLANTES, 1976); el Análisis Factorial Múltiple (ESCOFIER, y PAGÈS, 1984, 1990); el método Statico (SIMIER et al., 1999), son ejemplos de este tipo de metodología.

Estos métodos realizan un ACP de Dos Vías ponderado, de las tablas originales.



b) Existen otros métodos en los que el consenso se construye con las configuraciones originadas por alguna técnica multivariante (por ejemplo ACP o Biplot) aplicada a las matrices de datos originales, y la configuración consenso es la “más próxima” a las T configuraciones iniciales. Las distintas formas de integrar las configuraciones resultantes es lo que origina los distintos métodos. Como ejemplo de estos métodos podemos mencionar las Meta-Componentes (KRZANOWSKI, 1979, 1982) y los Meta-Biplot (MARTÍN- RODRIGUEZ et al., 2002), entre otros.

Dentro de este último grupo incluimos los métodos que a partir de análisis parciales en las diferentes tablas construyen un sistema de referencia y encuentran el compromiso maximizando la covarianza de las coordenadas de las proyecciones de los elementos filas de cada una de las tablas, sobre dicho sistema de referencia.

Los métodos relacionados con este grupo son: el Análisis de la Co-Inercia (CHESSEL y MERCIER, 1993), para el caso de tener sólo dos tablas y el Análisis de Co-Inercia Múltiple (CHESSEL y HANAFI, 1996).

c) Por último están los métodos que concatenan las matrices de productos cruzados y realizan una DVS de Tres Modos o Vías, que no es una generalización de la DVS de Dos Vías. Estos métodos tienen como objetivo lograr la mejor aproximación de los datos de partida, esto es, lograr el mejor ajuste en el sentido de los mínimos cuadrados.

Como ejemplo podemos citar los métodos derivados del modelo de Tucker (TUCKER, 1966, 1972).

Como hemos visto, están los métodos que utilizan la DVS de Dos Vías yuxtaponiendo las matrices iniciales ponderadas y tratándolas como si fueran de dos modos; los que utilizan una DVS de Dos Vías en cada una de las T tablas iniciales e integran las configuraciones resultantes y los que utilizan la DVS de Tres Vías concatenando las matrices de productos

cruzados.

Se presentan de manera resumida los siguientes métodos:

### **ANÁLISIS TRIÁDICO**

El Análisis Triádico (AT), también conocido como X-STATIS, fue propuesto por JAFFRENOU (1978). Algunos detalles del método y una aplicación interesante vienen en THIOULOUSE y CHESSEL (1987).

El método es aplicable a conjuntos de T matrices de datos, con los mismos individuos y las mismas variables observados en T ocasiones o condiciones.

Sigue el mismo esquema de Statis, se desarrolla entre etapas: inter-estructura, compromiso, intra-estructura, pero a diferencia de Statis, el AT no utiliza operadores, trabaja directamente con las matrices.

Aunque el método sea más restrictivo que el Statis, ya que supone las mismas J variables, tiene las ventajas de ser más sencillo de entender por los usuarios y no pierde la información original al no trabajar con operadores y permite más representaciones.

### **COVSTATIS y DISTATIS**

Si los operadores que integramos son matrices de covarianzas tomadas sobre los mismos individuos se conoce con el nombre de COVSTATIS (THIOULOUSE, 2011).

Si se parte de varias matrices de distancias definidas sobre los mismos objetos, en diferentes tiempos o condiciones, recibe el nombre de DISTATIS (ABDI y col., 2007) y es presentado por los autores como una generalización del método MDS clásico, para el caso de tablas de tres

vías.

## **CANOSTATIS**

Para el caso en que los individuos tengan estructura de grupos, VALLEJO-ARBOLEDA et al. (2007) desarrollan el CANOSTATIS.

Aborda el análisis de datos de tres vías, con la particularidad que las filas tienen estructura de grupo (tratamientos replicados, por ejemplo) con varias variables y diferentes ocasiones o tiempos.

El CANOSTATIS permite obtener un sistema de referencia común para representar los Análisis Canónicos de Poblaciones realizados en cada una de las ocasiones.

VALLEJO-ARBOLEDA y col, 2007 presentan el biplot inducido por el stasis canónico y demuestran que es posible no solo conocer la calidad de representación, en el primer plano principal, de los diferentes tiempos, sino también la calidad de representación de los diferentes tratamientos (grupos compromiso) y de las variables. También, tal como sucede en el stasis clásico, es posible representar, en el biplot stasis canónico, las trayectorias de las variables.

La representación biplot de las técnicas, stasis canónico y stasis dual canónico, proporciona ventajas relevantes frente a la forma clásica de presentar los Métodos stasis clásicos.

## **POWER- STATIS**

BENASSENI & DOSSE, (2012) llaman Power stasis a una versión de stasis en la cual el peso que se da a cada matriz difiere de la propuesta de L'Hermier en la cual el peso está relacionado con las componentes del primer vector propio de la matriz de correlaciones vectoriales.

### **DO-ACT, STATIS 4 y t+1 STATIS**

En 2004, Vivien y Sabatier presentan una extensión del statis para datos multibloque, a la que llaman DOUBLE-STATIS-ACT (DO-ACT) para dos tablas multibloque que no necesariamente tienen que tener la misma longitud.

El DO-ACT permite trabajar con distinto número de variables en cada matriz y tiene en cuenta la estructura de bloques de los datos durante los cálculos.

En 2008, Sabatier y Viven, extienden el Statis y el el DO-ACT y proponen el STATIS-4 que es presentado por sus autores como un nuevo método lineal para analizar tablas multibloque de 4 vías; es decir, más de dos conjuntos de tablas que tengan las filas comunes.

El método ha sido programado en S-Plus y puede ser solicitado a los autores.

SAUZAY y col. (2006), consideran las t-matrices que se integran en el statis y una matriz más con información externa sobre los individuos y presentan el t+1 STATIS.

### **STATICO y COSTATIS**

SIMIER et al. (1999) y THIOULOUSE et al. (2004) presentan lo que ellos llaman el método STATICO y que en realidad es un estudio combinado de COINERCIA y STATIS, y más recientemente THIOULOUSE et al. (2011), presentan el COSTATIS (STATIS y COINERCIA).

Ambos suponen interesantes alternativas que permiten estudiar las co-estructuras.

Comenzaremos analizando brevemente en qué consiste un análisis de

coinerencia.

El Análisis de Coinerencia (DOLEDEC & CHESSEL, 1994), permite encontrar la estructura común entre dos grupos de variables. Ver también DRAY et al. (2003) para un estudio detallado del método de coinerencia.

Este método pertenece a la familia de métodos para el estudio de pares de matrices (two- table coupling methods), que incluye el Canonical Correspondence Analysis (CCA), Análisis de Redundancia (RDA) y el análisis Canónico, pero difiere de estos análisis.

Tal como se ha señalado más arriba, (SIMIER et al., 1999; THIOULOUSE et al., 2004) presentan el STATICO como una técnica para analizar dos conjuntos de tablas de tres vías obtenidas en el estudio, **apareadas dos a dos.**

Cada tabla de dos vías se analiza con un análisis básico, componentes principales, correspondencias, etc.

Cada par de tablas se analiza mediante un análisis de Co-inerencia y las series de Tablas cruzadas resultantes de los análisis de coinerencia se analizan mediante un STATIS o un análisis triádico.

THIOULOUSE (2011), presenta el COSTATIS que es un análisis de coinerencia del compromiso de dos análisis de k-tablas.

COSTATIS es el análisis de co-inerencia de dos compromisos, busca las relaciones entre dos estructuras estables.

## **CAPÍTULO II**

# **APLICACIÓN A DATOS REALES**

## **2. APLICACIÓN A DATOS REALES**

---

### **2.1 LA CALIDAD DEL AGUA EN PANAMÁ**

En Panamá, hay un vínculo histórico con la administración del recurso hídrico desde que se concibió la idea de construir un canal interoceánico, y es en el territorio de la Cuenca Hidrográfica del Canal donde se almacena el agua que permite el abastecimiento de agua para las ciudades más pobladas del país, para la operación del Canal de Panamá y para muchas otras actividades económicas. El agua de esta zona facilita el movimiento transoceánico de un 6 por ciento del comercio mundial.

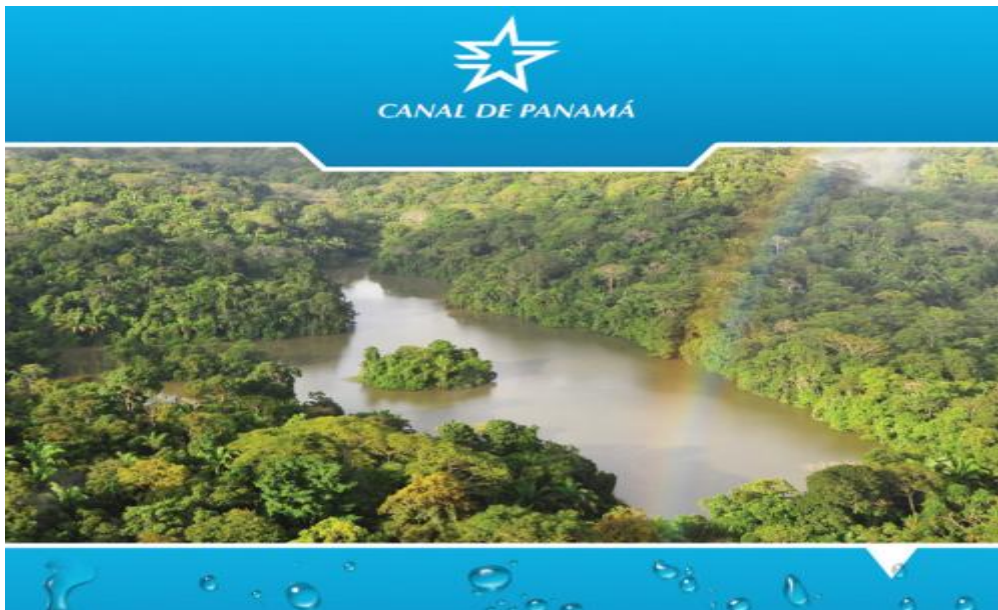
El recurso hídrico ha sido mencionado como uno de los factores más vinculado a las alteraciones del clima. Por ello, cada vez con mayor énfasis, aquellas organizaciones vinculadas al manejo de recursos hídricos invierten más tiempo y esfuerzo al estudio de los procesos naturales y ecosistemas con el fin de mejorar su gestión del agua y contribuir con los objetivos internacionales de construir una sociedad donde la humanidad y el resto de los seres vivos puedan existir en bienestar.

El agua es esencial para los seres vivos. La cantidad y calidad de este preciado líquido son condicionadas, en gran medida, por el estado de los bosques y el entorno natural y social en las cuencas hidrográficas.

El agua dulce es un recurso limitado y su calidad está bajo presión constante. Calidad de agua es el término utilizado para describir las condiciones físicas, químicas y biológicas que tiene un cuerpo de agua en un punto y tiempo dado (AUTORIDAD NACIONAL DEL AMBIENTE, 2006). Concentraciones de diferentes elementos, tales como, metales pesados, hidrocarburos, nutrientes, detergentes, temperatura, pH, turbiedad, oxígeno disuelto, demanda bioquímica de oxígeno, sólidos en suspensión, entre otros, son ejemplos de variables que determinan la calidad del agua.



Figura 1. Canal de Panamá



Ventajosamente, los conjuntos de variables de calidad de agua a menudo “se mueven juntos”; es decir, comparten información desde dimensiones diferentes. Por ejemplo, la turbiedad, la transparencia del agua y los sólidos en suspensión son variables disímiles con unidades diferentes, que sugieren acerca de la “claridad del agua”. El color y la turbiedad son dos conceptos que difieren entre sí, se tratan conjuntamente, ya que interfieren en la trasmisión de la luz en las aguas naturales y, por consiguiente, regulan los procesos biológicos que en ella se realizan (ROLDAN, 2008).

Las principales amenazas a las que se enfrenta la CHCP (Cuenca Hidrográfica del Canal de Panamá) están relacionadas con la creciente presión de las actividades económicas y el crecimiento de la población, que inciden en la deforestación, erosión, sedimentación, pérdida de la calidad del agua, disminución de la biodiversidad, etc. Los problemas más críticos en la CHCP no son diferentes a los que están presentes en otras cuencas del país: infraestructura básica insuficiente, mal manejo de desechos sólidos, descargas de aguas servidas a los cuerpos de agua, desnutrición y prevalencia de

enfermedades de origen hídrico y resultado de la falta de prácticas de saneamiento básico.

## **2.2 ANTECEDENTES**

Mantener la vitalidad natural de la Cuenca del Canal de Panamá es de suma importancia para Panamá. Es el agua que abastece la Cuenca del Canal de Panamá la que llena las esclusas del Canal de Panamá. Además, la Cuenca brinda agua potable a más de 1,5 millones de personas, suministra energía hidroeléctrica y mantiene una amplia extensión de biodiversidad tropical.

La Cuenca del Canal de Panamá es un área definida legalmente que cubre más de 500.000 hectáreas. Abarca las diferentes cuencas de captación de ríos que proporcionan el agua que necesita el Canal de Panamá ahora y en el futuro. La Cuenca del Canal de Panamá se extiende de la costa atlántica a la costa pacífica del país, conectando los centros metropolitanos más dinámicos del país, la ciudad de Panamá y Colón. La Cuenca del Canal de Panamá está ubicada en el epicentro económico nacional de Panamá, un sitio de un intenso desarrollo residencial y comercial.

Hasta 1999, la Cuenca del Canal de Panamá cubría un área de 326.000 hectáreas que corresponde a la cuenca que abastece de agua a los lagos Alhajuela y Gatún. En vista de la creciente demanda de agua, tanto para la operación del Canal como para uso industrial y residencial, la Ley 44 de agosto de 1997, extendió la Cuenca hacia el oeste para abarcar 225.000 hectáreas adicionales. En la actualidad, la Cuenca “extendida” cubre 552.761 hectáreas.

Desde principios de la década de los ochenta, las inquietudes acerca de la necesidad de proteger el recurso fundamental para la continua operación de la principal fuente de ingresos económicos de Panamá, el Canal, condujo a una serie de acciones para proteger y monitorear la Cuenca del Canal de Panamá. El Gobierno de Panamá empezó a aumentar la protección oficial de las áreas

boscosas en la Cuenca del Canal de Panamá. Como resultado, la mayor parte de las áreas boscosas de la Cuenca que quedaban se encuentran actualmente bajo protección oficial.

Desde septiembre de 1997, el Proyecto de Monitoreo de la Cuenca del Canal, bajo el patrocinio del Instituto Smithsonian de Investigaciones Tropicales (STRI), la Autoridad Nacional del Ambiente (ANAM) y la Agencia de Estados Unidos para el Desarrollo Internacional (USAID) han estado monitoreando el status ambiental del suelo, el agua y otros recursos de la Cuenca.

Los hallazgos hechos por el Proyecto de Monitoreo de la Cuenca del Canal, publicados en 1999, revelan que en los ríos primarios se han disminuido y los flujos de agua durante el año están regulados. Desafortunadamente, las condiciones de algunos ríos secundarios que fluyen en el embalse Gatún, que abastece la mitad del agua de la Cuenca, son alarmantes. Estos ríos tienen una protección forestal relativamente escasa; los proyectos residenciales han proliferado a lo largo del corredor transístmico entre Panamá y Colón. En consecuencia, las tasas de sedimentación han aumentado y los flujos de agua internacionales son irregulares.

Los embalses de la cuenca hidrográfica del Canal de Panamá permiten la operación del Canal de Panamá durante todo el año.

En ellos se almacena el agua necesaria para el tránsito de los buques, y para abastecer a la población. La calidad de sus aguas, por lo tanto, es de suma importancia para satisfacer una de las necesidades básicas de la población: agua potable para consumo humano.

El embalse Gatún es la principal reserva de agua para la operación del Canal. Se le considera la “caja fuerte” de la Autoridad del Canal DE Panamá (ACP), porque en él se almacena el agua necesaria para la operación del Canal, y una



contaminante. Sin embargo, se observa que existen sitios que requieren mayor atención, donde los valores más bajos en el ICA sugieren que el impacto por las actividades antropogénicas es una amenaza real y creciente.

### **2.3 OBJETIVOS**

El objetivo general de este trabajo, es poner de manifiesto la utilidad de métodos multivariantes de dos y tres vías (HJ-Biplot y Statis), como herramientas de inspección de matrices de datos físicos, químicos y biológicos que tiene una masa de agua en un punto y tiempo dado.

Los objetivos específicos son los siguientes:

- Hallar las asociaciones existentes entre las variables fisicoquímicas y biológicas que tiene una masa de agua.
- Comprobar las diferencias de los lugares donde se realizaron las colectas de las muestras de acuerdo a la calidad de agua.
- Encontrar las agrupaciones de variables fisicoquímicas y biológicas.
- Relacionar las agrupaciones de las variables fisicoquímicas y biológicas, con los meses de la estación climática de la región.
- Estudiar las estructuras multifactoriales en los diferentes puntos de muestreo y encontrar las componentes estables en todos ellos definiendo una configuración consenso.

## 2.4 MUESTRA

Los registros utilizados para el análisis, corresponden a las mediciones de los parámetros físico-químicos y biológicos que determinan la calidad de agua en los sitios de muestreo, Gamboa (643343E; 1007468N) y Paraíso (651144E; 997879N), ubicados en el embalse Gatún. Los datos del año 2009, fueron colectados conforme a los protocolos de la Unidad de Calidad de Agua de la ACP.

Se colectaron tres muestras en cada sitio a un metro de la superficie. El muestreo se realizó de manera sistemática cada siete días hasta completar los catorce primeros días del mes; durante febrero a mayo y julio a diciembre de 2009. Se estableció un horario de muestreo entre las 11:00 am a 2:30 pm. En campo, se realizaron mediciones in situ con la sonda multiparamétrica YSI, modelo 556 MPS la cual fue previamente verificada y/o calibrada en laboratorio. Se emplearon envases apropiados para cada muestra (vidrio, plástico de polipropileno, etc.). Las muestras colectadas fueron puestas en neveras con hielo hasta su transporte al laboratorio. El acceso a los sitios se realizó vía terrestre.

Se colectaron muestras para realizar determinaciones analíticas en el laboratorio de la Unidad de Calidad de Agua, División de Agua del Canal de Panamá, además, muestras para análisis de nitrógeno y fósforo total fueron colectadas y preservadas (H<sub>2</sub>SO<sub>4</sub> al 20%), enviadas y analizadas en el laboratorio de la Universidad Tecnológica de Panamá, contratado por el canal de Panamá.

Las metodologías de análisis corresponden a las descritas en el Standard Method for the Examination of Water and Wastewater (APHA, AWWA, WEF, 21ª Edición) (ver Tabla 1). Un total de doce parámetros de calidad de agua, los cuales incluyen: temperatura (°C), pH, transparencia (m), turbiedad (NTU), nitratos (mg/l), ortofosfatos (mg/l), fósforo total (mg/l), nitrógeno total (mg/l),

clorofila a ( $\mu\text{g/l}$ ), radiación solar (Langleys), toxina (microcistina,  $\mu\text{g/l}$ ), y oxígeno disuelto ( $\text{mg/l}$ ), fueron medidos en campo y laboratorio.

Parámetro	Método	Unidades de medida	Límite detección
<sup>1</sup> Oxígeno disuelto	SM 4500-O C	mg/l	0,1 mg/l
N-Nitratos	SM 4500-NO <sub>3</sub> - E	mg/l	0,01 mg/l
P-Fosfatos	SM 4500-P E	mg/l	0,02 mg/l
Fósforo total	SM 4500-P D	mg/l	0,03 mg/l
Nitrógeno total	SM 4500-Norg-B	mg/l	0,03 mg/l
<sup>1</sup> pH	SM 4500-H+B	pH units	0,1 pHunits
<sup>1</sup> Temperatura	SM 2550-A	°C	0,1°C
Turbiedad	SM 2130-B	NTU	0,05 NTU
<sup>1</sup> Transparencia	Profundidad de Secchi	m	0,1m
Clorofila a	SM 10200-H, modificación USEPA 445.0	$\mu\text{g/l}$	0,05 $\mu\text{g/l}$

Tabla 1. Métodos analíticos empleados según parámetros de calidad de agua  
<sup>1</sup> Mediciones *in situ*.

Las muestras para análisis de toxina (microcistinas) fueron colectadas en envases de vidrio ámbar, y las mediciones se realizaron utilizando el método Quanti tube kit ET039 de Envirologix. Los registros de radiación solar fueron suministrados por la Sección de Recursos Hídricos de la División de Agua del Canal de Panamá, estos corresponden a la estación SRFAA, que se encuentra ubicada en Balboa, encima de la caseta próxima al mercado de abastos.

Contamos con ciento veinte observaciones como resultado de tres muestras (puntos de muestreo) colectadas en cada sitio de muestreo a saber Paraíso y Gamboa dos veces al mes febrero(feb), marzo(mar), abril(abr), mayo(may), julio(jul), agosto(ago), septiembre(sep), octubre(oct), noviembre(nov) y diciembre(dic)) del año 2009. Donde cada punto de muestreo constituye una muestra y que representa el mes, el número de veces que se muestrea durante el mes y el sitio del muestreo. Por ejemplo el punto de muestreo feb1P1 representa al mes de febrero, primer muestreo del mes en el sitio de

Paraíso; y de igual manera el punto de muestreo feb2G2, representa al mes de febrero, segundo muestreo del mes en el sitio de Gamboa.

El esquema que presenta es el siguiente:

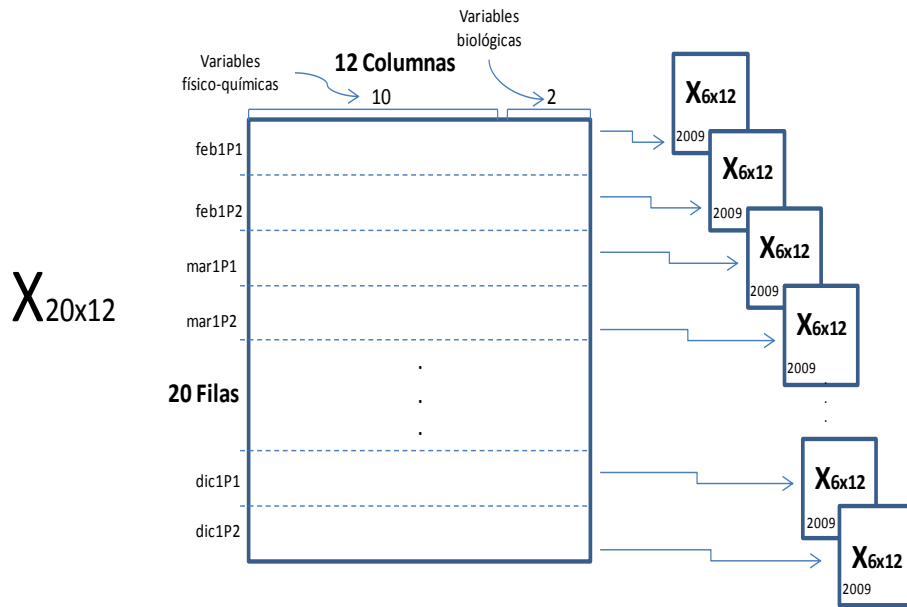
**I=120 observaciones:** seis muestras recogidas en dos puntos de muestreo (tres en Paraíso y tres en Gamboa), durante diez meses, dos veces cada mes, es decir, información mensual de las muestras en el período de estudio febrero de 2009 hasta diciembre de 2009 (a excepción del mes de junio, de manera que podemos fragmentar la base en 20 matrices). Cada punto de muestreo constituye una muestra y que representa el mes, el número de veces que se muestrea durante el mes y el sitio del muestreo.

**K=20 matrices,** que corresponden a 2 mediciones por 10 meses (feb1, feb2, mar1, mar2, abr1, abr2, may1, may2, jul1, jul2, ago1, ago2, sep1, sep2, oct1, oct2, nov1, nov2, dic1 y dic2) del año 2009.

**J=12 variables:** Temperatura (Temp), pH, Transparencia (Trasp), Turbiedad, Nitrato, Ortofosfatos, P total, N total, clorofila a, Radiación Solar (Rad. Solar), Oxígeno, Toxina que corresponden a parámetros físico-químicos y toxina y la clorofila a que corresponden a parámetros biológicos que determinan la calidad de agua en los lugares de muestreo, Gamboa y Paraíso, ubicados en el embalse Gatún.



Figura 3. Base de datos



Para los cálculos estadísticos, se utilizaron el Excel, IBM® SPSS® v.21, ADE-4 (DRAY & DUFOUR, 2007) y MultBiplot (VICENTE VILLARDON, 2010).

## 2.5 DESCRIPCIÓN DE LA MUESTRA

En el estudio se consideran diez variables físico-químicas y dos variables biológicas, cuyos valores medios y sus correspondientes desviaciones típicas, se presentan en la tabla siguiente:

Variables	GLOBAL	PARAÍSO	GAMBOA	p-valor
	Media±DS	Media±DS	Media±DS	
Temperatura (°C)	28,30±0,69	28,40±0,57	28,20±0,78	0.124
pH	7,12±0,46	7,27±0,46	6,97±0,41	<0.01
Transparencia (m)	0,81±0,48	0,52±0,31	1,11±0,44	<0.01
Turbiedad	19,59±21,81	29,83±26,33	9,35±7,28	<0.01
Nitatos (mg/l)	0,17±0,11	0,20±0,12	0,15±0,09	0.004
Ortofosfato (mg/l)	0,01±0,01	0,01±0,01	0,01±0,01	0.361
Fósforo total (mg/l)	0,07±0,11	0,08±0,13	0,06±0,08	0.894
Nitrógeno total (mg/l)	0,49±0,53	0,54±0,56	0,42±0,50	0.099
Clorofila a (ug/l)	3,98±5,23	3,84±3,01	4,12±6,75	0.764
Radiación solar	39,68±25,44	39,91±25,84	39,46±25,26	0.922
Oxígeno disuelto (mg/l)	5,65±1,09	5,94±0,84	5,37±1,23	0.366
Toxina (ug/l)	0,14±0,11	0,12±0,06	0,15±0,14	0.004

Tabla 2. Características de las variables físico-químicas y biológicas

En cuanto a los lugares de muestreo Gamboa y Paraíso ubicados en el embalse Gatún, el promedio de oxígeno disuelto en agua fue de 5,94mg/L±0,84 en Paraíso y de 5,37mg/L±1,23 en Gamboa (p=0.366). La temperatura promedio de ambos lugares fue de 28,30°C±0,69, siendo prácticamente igual en ambos puntos de muestreo y no dándose por tanto, diferencias estadísticamente significativas. La temperatura es importante porque es el reflejo de lo que es la fuente primaria de energía solar y de ella dependen las reacciones físico-químicas que se puedan dar en las aguas.

Las aguas sin contaminación normalmente presentan unos niveles de nitratos inferiores a 4 ppm (mg/L). Si la concentración de nitratos es superior a 40 ppm el agua se considera no potable, por lo que la calidad del agua es de excelente a buena en ambas zonas de muestreo, con concentraciones de oxígeno disuelto adecuadas para el soporte de la vida acuática. En lo relativo a los nitratos (NO<sub>3</sub>) el área de Gamboa presenta potencias espectrales menores

que las del área de Paraíso ( $0,15\pm 0,09$  vs.  $0,20\pm 0,12$ ;  $p=0,004$ ). Los nitratos son derivados del uso excesivo de fertilizantes, y pueden ocasionar una contaminación significativa en aguas superficiales y subterráneas.

Las concentraciones espectrales de potencia en cuanto a los ortofosfatos guardan la misma proporción en ambas áreas ( $p>0,05$ ).

El fósforo junto con el nitrógeno, son dos de los nutrientes fundamentales de todos los seres vivos, de forma que contenidos anormalmente altos de estos en las aguas pueden producir un crecimiento incontrolado de la biomasa acuática (eutrofización), con problemas de crecimiento de algas indeseables en embalses y lagos, con acumulación de sedimentos, etcétera. El fósforo generalmente está presente en las aguas naturales en forma de fosfatos. Los fosfatos se encuentran en los fertilizantes y los detergentes y pueden llegar al agua con el escurrimiento agrícola, los desechos industriales y las descargas de aguas negras. Los fosfatos, al igual que los nitratos, son nutrientes para las plantas. Estimulan el crecimiento de las algas y cianobacterias, que se reconocen con facilidad como capas de limo verde y pueden, eventualmente, cubrir la superficie del agua. Las concentraciones de nitrógeno y fósforo en el área de Paraíso fueron más elevadas que en Gamboa. La concentración media del fósforo total (FT) encontrada en el embalse fue de  $0,07$  mg/L.

En relación a la clorofila a, la concentración media fue de  $3,98$  mg/L, en un rango amplio de  $0,1$  a  $47,6$  mg/L. Entre los sitios de muestreo, se observó que las concentraciones más altas se presentaron en el área de Gamboa.

Con respecto al pH, es medido en escala con rango de 0 a 14, mostrando en la zona un punto aproximadamente neutro ( $\text{pH}=7$ ). Los valores por debajo de 7 corresponden a un pH ácido y por encima de 7 corresponden a un pH básico o alcalino.

## **2.6 CARACTERIZACIÓN MULTIVARIANTE DE LA INFLUENCIA DE VARIABLES FÍSICO-QUÍMICAS Y BIOLÓGICAS EN LA CALIDAD DEL AGUA: APLICACIÓN AL EMBALSE GATÚN (PANAMÁ)**

El estudio de la calidad del agua, y en especial de aquellas aguas destinadas al uso o consumo humano, resulta imprescindible para garantizar su buen estado y la seguridad de todas aquellas personas que vayan a hacer uso de ésta, así como las condiciones óptimas de la misma para mantener la biodiversidad de las especies que habitan su entorno. Una clara muestra de este interés se refleja en las múltiples investigaciones que tienen como objetivo principal analizar la calidad del agua (BEAMONTE et al., 2004, 2007; KARAVOLTSOs et al., 2007; SHRESTHA y KAZAMA, 2006). El componente agua no es más que una parte del ecosistema acuático en el que se desarrollan una serie de comunidades vivas que dependen de las características fisicoquímicas de sus aguas que pueden verse notablemente modificadas, con la consiguiente alteración de aquéllas. La mayoría de los parámetros utilizados para la evaluación de la calidad de aguas son de carácter fisicoquímico, que no reflejan las posibles alteraciones existentes que hayan podido suceder tiempo atrás. Por este motivo la vigilancia y control de la contaminación del agua se complementa con organismos como "bioindicadores" (ALBA-TERCEDOR 1996).

Se realiza un estudio de la calidad del agua en el embalse de Gatún en las áreas de Gamboa y Paraíso en los meses de febrero a diciembre del año 2009. Se propone para ello, el uso de las herramientas de análisis multivariante como forma de detectar anomalías en las propiedades del agua no visibles mediante técnicas univariantes, así como para poder discernir cuándo el agua cumple los criterios de calidad exigidos.

La disponibilidad de equipos de instrumentación sofisticados para medición y análisis de calidad de agua en los sitios de muestreo, Gamboa y Paraíso, ubicados en el embalse Gatún parte de la Cuenca Hidrográfica del Canal de

Panamá, nos da la oportunidad de recoger la medición de decenas de variables en ocasiones de forma simultánea, lo que nos permite la aplicación de métodos estadísticos multivariantes de dos y tres vías, tales como el HJ-Biplot y Statistica.

El desmejoramiento de la calidad del agua es un motivo de preocupación por el crecimiento desproporcionado de la población humana, por la expansión de la actividad industrial y agrícola y por las amenazas del cambio climático que pueden tener importantes alteraciones en el ciclo hidrológico. Por ello, es importante a través de estos métodos estadísticos multivariantes evaluar las diversas características fisicoquímicas y biológicas del agua de los lugares de Gamboa y Paraíso perteneciente al embalse Gatún.

En el trabajo que se presenta se pretende estudiar y conocer desde un punto de vista multivariante, las posibles relaciones existentes entre los distintos “parámetros de calidad de las aguas” establecidos según la Normativa de la Autoridad del Canal de Panamá, así como la relación de estos en los distintos puntos de muestreo. Dada la dificultad de analizar la gran cantidad de datos disponibles, el objetivo es sintetizar dicha información, lo que equivale a reducir la cantidad de estos datos minimizando a su vez la pérdida de información. Un modo de conseguir este objetivo consiste en analizar la estructura inicial de la nube de puntos del hiperespacio mediante una configuración simplificada en un espacio de menor dimensión. La inspección de matrices de datos multivariantes se ha llevado utilizando el HJ-Biplot (Galindo, 1986), estandarizando por columnas (variables). Este método nos permite una representación conjunta, en dimensión reducida, de los meses de muestreo y las variables físico-químicas y biológicas. En el HJ-Biplot esta reducción se lleva a cabo mediante la aproximación de matrices por otras de menor rango, mediante la descomposición en valores singulares. Para ello, se dispuso de una matriz de datos cuyas filas correspondía a los individuos bajo estudio (meses de muestreo) y las columnas a las variables físico-químicas y biológicas. Para llevar a cabo el análisis se utilizó el software Multbiplot (Vicente-Villardón, 2010).

El primer paso fue determinar el número de dimensiones necesarias que mejor representan o explican la matriz de datos. Aunque existen varios criterios, en este trabajo se optó por usar el menor número de dimensiones que permitan una interpretación correcta. Con respecto a la absorción de inercia por tanto, se han retenido dos ejes, pues con ello se consigue una inercia acumulada del 48,23%, siendo la tasa de inercia en el primer eje del 33,85%, suficiente para caracterizar las variables físico-químicas y biológicas en las áreas de Gamboa y Paraíso (ver Tabla 3).

<b>Ejes</b>	<b>Valores propios</b>	<b>Inercia %</b>	<b>Inercia acumulada %</b>
Axis 1	483.404	33.852	33.852
Axis 2	205.438	14.386	48.238
Axis 3	161.113	11.282	59.521
Axis 4	156.875	10.986	70.506
Axis 5	114.612	8.026	78.532
Axis 6	93.933	6.578	85.11

Tabla 3. Valores propios y varianza explicada

Atendiendo a la contribuciones del factor al elemento para las columnas (ver tabla 4), se observa que todas las variables se han de interpretar en el plano 1-2, a excepción del porcentaje de Ntotal que aunque presenta contribuciones ligeramente superiores en el en el plano 1-3, se analiza en el plano 1-2 por resultar de mayor interés su interpretación junto con el resto de variables físico-químicas.

El eje 1, queda determinado por las variables pH, transparencia, turbiedad, nitrato, ortofosfato, clorofila a y oxígeno, mientras que las variables radiación solar y toxina, son variables de eje 2.

	<b>Eje 1</b>	<b>Eje 2</b>	<b>Eje 3</b>
Temp	2	63	71
pH	<b>397</b>	68	5
Transp	<b>546</b>	37	48
Turbiedad	<b>712</b>	23	18
Nitrato	<b>764</b>	24	59
Ortofosfatos	<b>564</b>	3	2
Ptotal	242	139	177
Ntotal	1	15	<b>715</b>
Clorofila a	<b>289</b>	1	142
Rad.Solar	11	<b>643</b>	36
Oxígeno	<b>535</b>	8	26
Toxina	0	<b>704</b>	55

Tabla 4. Calidad de representación para las columnas

En cuanto a las filas de los 120 puntos de muestreo tan solo cuatro no han quedado bien recogidos en los tres primeros ejes: may1P1 (mes de mayo, primer muestreo del mes en el lugar de Paraíso), dic2P2 (2) (mes de diciembre, segundo muestreo del mes en el lugar de Paraíso) y dic2G2 (mes de diciembre, segundo muestreo del mes en el lugar de Gamboa). Ver tabla 5.

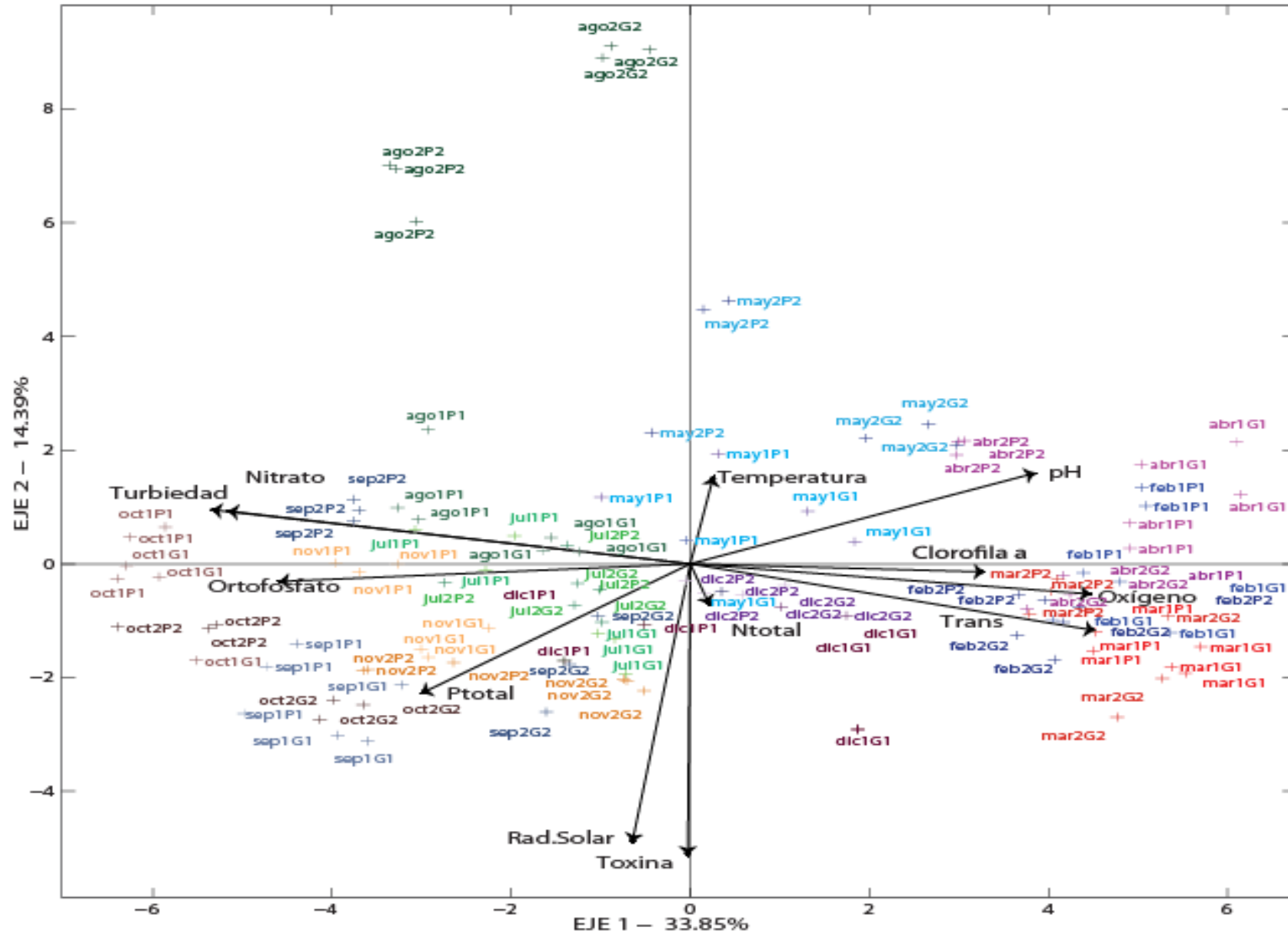
	Eje 1	Eje 2	Eje 3		Eje 1	Eje 2	Eje 3
feb1P1	609	610	674	ago1P1	315	520	570
feb1P1	729	758	787	ago1P1	279	304	427
feb1P1	708	758	779	ago1P1	242	258	374
feb1G1	668	708	724	ago1G1	118	120	378
feb1G1	694	729	741	ago1G1	80	87	455
feb1G1	758	767	777	ago1G1	54	56	440
feb2P2	677	697	730	ago2P2	154	841	848
feb2P2	553	565	693	ago2P2	174	849	849
feb2P2	697	715	767	ago2P2	159	853	853
feb2G2	524	586	587	ago2G2	2	655	697
feb2G2	502	588	591	ago2G2	9	731	731
feb2G2	619	657	659	ago2G2	7	705	738
mar1P1	651	726	742	sep1P1	367	420	623
mar1P1	697	746	747	sep1P1	273	301	538
mar1P1	727	753	755	sep1P1	294	376	552
mar1G1	693	777	845	sep1G1	266	422	646
mar1G1	787	839	846	sep1G1	258	371	489
mar1G1	738	822	826	sep1G1	261	457	645
mar2P2	622	625	655	sep2P2	556	578	579
mar2P2	675	685	692	sep2P2	527	575	576
mar2P2	687	724	727	sep2P2	558	594	595
mar2G2	499	659	711	sep2G2	41	112	171
mar2G2	546	626	919	sep2G2	70	253	277
mar2G2	503	518	946	sep2G2	39	70	86
abr1P1	510	512	560	oct1P1	598	605	640
abr1P1	618	620	625	oct1P1	667	671	736
abr1P1	645	659	722	oct1P1	704	705	707
abr1G1	688	715	864	oct1G1	610	668	668
abr1G1	606	679	817	oct1G1	531	531	585
abr1G1	558	627	897	oct1G1	564	565	736
abr2P2	368	520	666	oct2P2	581	607	609
abr2P2	382	581	604	oct2P2	520	541	546
abr2P2	376	564	694	oct2P2	504	519	636
abr2G2	406	424	622	oct2G2	301	440	477
abr2G2	324	329	715	oct2G2	334	480	542
abr2G2	520	421	572	oct2G2	384	523	560
may1P1	0	9	21	nov1P1	511	512	604
may1P1	64	153	218	nov1P1	597	597	611
may1P1	5	212	315	nov1P1	468	468	640
may1G1	104	156	209	nov1G1	116	145	296
may1G1	176	184	204	nov1G1	200	263	334
may1G1	2	6	443	nov1G1	214	267	267
may2P2	7	202	303	nov2P2	177	253	481
may2P2	1	539	640	nov2P2	468	593	651
may2P2	4	471	577	nov2P2	489	618	700
may2G2	240	545	591	nov2G2	25	238	260
may2G2	374	558	574	nov2G2	11	210	221
may2G2	217	401	430	nov2G2	42	352	290
jul1P1	183	194	797	dic1P1	20	109	303
jul1P1	307	319	838	dic1P1	171	182	542
jul1P1	225	226	826	dic1P1	185	456	554
jul1G1	26	213	362	dic1G1	57	195	357
jul1G1	33	115	402	dic1G1	106	159	318
jul1G1	54	130	378	dic1G1	103	150	328
jul2P2	85	90	690	dic2P2	15	28	101
jul2P2	389	395	638	dic2P2	0	4	47
jul2P2	48	58	775	dic2P2	1	12	151
jul2G2	62	67	265	dic2G2	113	144	223
jul2G2	78	103	328	dic2G2	34	53	97
jul2G2	42	88	399	dic2G2	37	42	65

Tabla 5. Calidad de representación para las filas

En la figura 4 se muestra el gráfico factorial del plano 1-2, donde la inercia acumulada asciende a 48.24%.



Figura 4. Representación HJ-BIPLLOT en el primer plano principal



Las variables físico-químicas y biológicas analizadas se representan mediante vectores, mientras que las áreas de muestreo se identifican mediante puntos, y su color varía en función del mes y lugar de muestreo (Gamboa y Paraíso). En el primer eje las variables con mayor variabilidad fueron la turbiedad, nitrato, ortofosfatos, oxígeno, transparencia y clorofila a.

El segundo eje queda determinado por las variables radiación solar y toxina, las cuales aparecen en la parte inferior del plano, donde los meses de muestreo de septiembre en el primer muestreo del mes en ambas zonas (Gamboa y Paraíso) y octubre y noviembre en el área de Gamboa en el segundo muestreo del mes en que se colectaron las muestras toman los valores más altos en estas variables. En contraposición a estos meses, aparecen los meses de agosto y mayo en Paraíso y Gamboa en la segunda toma de muestreo, los cuales se caracterizan principalmente por mayores valores en la variable Temp (Temperatura); esto es fácilmente justificado si consideramos que ambos meses se registraron altas temperaturas. Ambas variables aparecen muy correlacionadas. Cabe destacar que la toxina es la variable más importante para discriminar entre los puntos de muestreo en el eje 2, siendo además un indicador con una buena variabilidad en el estudio.

Observamos en el gráfico factorial que existe una relación fuerte y directa entre el oxígeno, transparencia y clorofila a, siendo esta última variable la más importante para discriminar entre los puntos de muestreo en el eje 1. Proyectando perpendicularmente cada uno de los puntos que representan los meses de muestreo sobre estas variables, observamos que toman valores especialmente altos en los meses de febrero, marzo y abril en ambas áreas (Paraíso y Gamboa), las dos veces del mes en que se recogieron las muestras, a excepción del segundo muestreo en el mes de abril en Paraíso que presentan altos valores de la variable pH.

Igualmente hay una cierta correlación entre las variables nitrato, turbiedad y ortofosfatos, las cuales toman valores especialmente altos en los meses de

septiembre y octubre en ambas áreas en el primer y segundo muestreo y el primero del mes de noviembre en Paraíso.

En el mes de noviembre tanto en Gamboa como Paraíso en el primer y segundo muestreo, el mes de septiembre en ambas áreas en el primer muestreo y en Gamboa en el segundo muestreo, mientras que en el mes octubre en el área de Gamboa en el segundo muestreo, se caracterizan por presentar valores elevados de fósforo total (Ptotal).

Además se observaron relaciones nulas entre la turbiedad y nitrato con la radiación solar. De igual forma, no se relaciona la toxina con el ortofosfato. Esto significa que los cambios que haya en estas variables no se reflejan en un aumento o disminución en las demás.

Destaca el hecho de que los puntos de muestreo correspondientes al mismo mes, tienden a situarse en la misma zona del gráfico HJ-Biplot, lo que indica que representan características similares.

### **2.6.1 CLUSTERS SEGÚN LOS PUNTOS DE MUESTREO**

El análisis de cluster se realizó a partir de las coordenadas que se obtuvieron del HJ-biplot (método K-means, coseno). Se formaron dos clusters con los distintos puntos de muestreo. En la representación gráfica de los mismos pueden observarse los clusters identificados con las líneas Convex-Hulls (figura 5). Este análisis nos permitió identificar las variables físico-químicas y biológicas que incidieron para las agrupaciones entre los distintos puntos de muestreo.

Las calidades de representación para cada conglomerado en el plano 1-2 se exponen en la tabla 6.

	<b>n</b>	<b>QLR - 1</b>	<b>QLR - 2</b>	<b>QLRA 1 - 2</b>
Cluster 1	64	96,83	2.00	98.83
Cluster 2	56	96,83	2.00	98.83

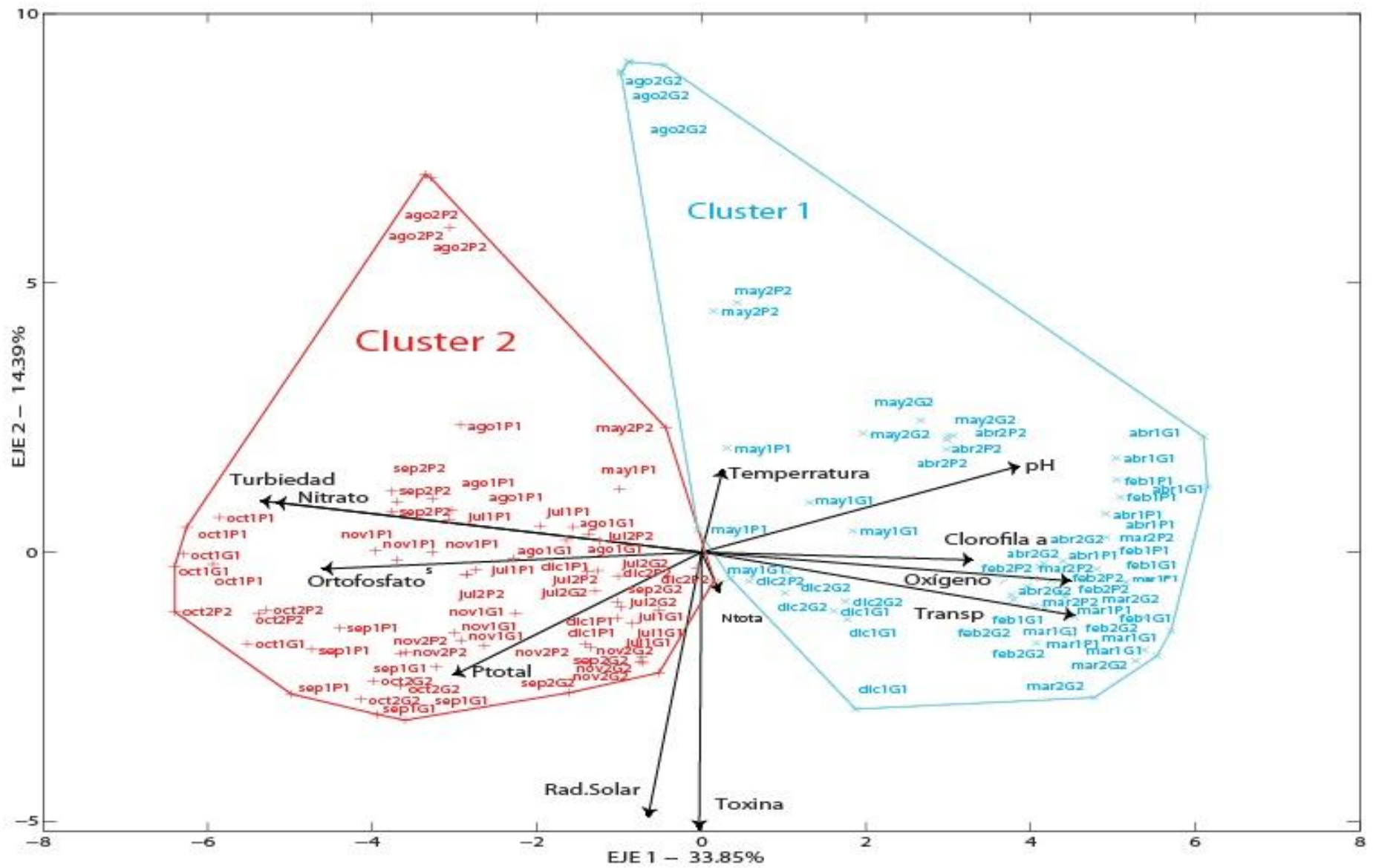
Tabla 6. Conglomerados y calidades de representación en el plano 1-2.

Los clusters encontrados pueden caracterizarse como:

Cluster 1: Está formado por puntos de muestreo de los meses de febrero, marzo, abril y mayo de Gamboa y Paraíso correspondientes a los meses de la estación seca y el mes de diciembre. Este cluster tiene la presencia de las variables pH, transparencia, clorofila a, oxígeno y Temperatura (Temp). Estas variables nos hablan de la productividad en un ecosistema acuático.

Cluster 2: Conformado por puntos de muestreo principalmente de los meses de mayo de Paraíso, julio, agosto, septiembre, octubre y noviembre. Este cluster está conformado por las variables, nitrato, ortofosfatos, turbiedad y P-total. Son parámetros que sufren variaciones en la estación lluviosa, y que a su vez pueden influir en la presencia de bacterias con potencial toxigénico.

Figura 5. Representación factorial del HJ-Biplot por clusters, plano 1-2



## **2.7 STATIS DUAL COMO HERRAMIENTA EVALUATIVA PARA LA VALORACIÓN DE CALIDAD DEL AGUA**

Como ya se ha dicho anteriormente, se realiza un estudio de la calidad del agua en el embalse de Gatún en las áreas de Gamboa y Paraíso en los meses de febrero a diciembre del año 2009, haciendo uso de técnicas de análisis multivariante. Este estudio es tanto longitudinal como transversal, es decir, la calidad del agua se evalúa tanto a lo largo del embalse en un área determinada, observando así las variaciones en sus características durante su trayectoria, como a través del tiempo, detectando de esta forma el cambio en sus propiedades y calidad durante el período considerado. Con el objetivo de analizar la estabilidad o la evolución de las estructuras de covariación entre los diferentes puntos de muestreo, vamos a llevar a cabo un STATIS-dual, el cual permite el análisis simultáneo de distintas matrices numéricas que recogen información sobre las mismas variables medidas en diferentes conjuntos de individuos y a grandes rasgos presenta las siguientes características:

- Analiza en términos generales las diferencias y semejanzas entre las matrices de correlación asociadas a las respectivas submatrices de datos.
- Proporciona medidas numéricas y planos factoriales que ponen de manifiesto las variables que son responsables de las diferencias entre las matrices analizadas.

Los datos de la Calidad del Agua longitudinales, pueden ser entendidos como datos de tres vías: el primer modo se refiere a los meses, el segundo a los puntos de muestreo utilizados para evaluar la calidad del agua y el tercero se refiere a las distintas ocasiones en que las propiedades de la calidad del agua se evalúan.

El Statis dual consiste en resumir la información proveniente de varias matrices, en nuestro caso 20 matrices. A esta matriz se denomina matriz compromiso, que es la media ponderada de las 20 matrices originales,

utilizando los pesos del primer vector propio resultante de la descomposición de la matriz de correlaciones vectoriales, utilizando el coeficiente de Hilbert-Schmidt. Este análisis de la información, desde el punto de vista de tabla múltiple, realiza un estudio comparativo de las tablas, en este caso se van a comparar las estructuras internas de los puntos de muestro de las dos zonas de recogida (Paraíso y Gamboa), poniendo de manifiesto no sólo cuáles tienen una estructura similar o diferente, sino también qué variables son responsables de estas diferencias y/o semejanzas.

### **2.7.1 LA INTERESTRUCTURA**

Consiste en la comparación de las estructuras de las 20 matrices de nuestro estudio. Esto es, se calcula la matriz de correlaciones entre estructuras y a partir de ahí se calculan las Componentes Principales correspondientes a esa matriz de correlaciones entre matrices. Así conseguimos reducir la dimensionalidad y representar cada matriz de datos como un punto en el subespacio euclídeo de baja dimensión en donde se reproducen las interdistancias del hiperespacio. Para ello, es necesario definir una distancia entre dos tablas y encontrar una representación gráfica de las mismas en la que la proximidad entre dos puntos corresponda a una semejanza entre tablas.

Se calcula la matriz de correlación vectorial, donde podemos observar que meses guardan mayor relación y cuáles presentan mayores diferencias. Por tanto, el análisis de la tabla formada por la yuxtaposición de las subtablas correspondientes a los 20 periodos de muestreo en el estudio, ha proporcionado los siguientes coeficientes de correlación RV entre los meses indicando los coeficientes próximos a 1 estructuras semejantes. Podemos apreciar que las relaciones entre todos los puntos en los meses de muestreo no es muy alta, destacando la relación entre los meses de octubre en la primera toma de la muestra y la segunda del mes de febrero, así como entre la primera del mes de agosto y la segunda del mes de septiembre (ver tabla 7).

	feb1	feb2	mar1	mar2	abr1	abr2	may1	may2	jul1	jul2	ago1	ago2	sep1	sep2	oct1	oct2	nov1	nov2	dic1	dic2
<b>feb1</b>	1																			
<b>feb2</b>	0,519	1																		
<b>mar1</b>	0,177	0,220	1																	
<b>mar2</b>	0,229	0,206	0,565	1																
<b>abr1</b>	0,124	0,167	0,776	0,516	1															
<b>abr2</b>	0,209	0,365	0,489	0,635	0,409	1														
<b>may1</b>	0,061	0,136	0,255	0,141	0,349	0,108	1													
<b>may2</b>	0,319	0,360	0,314	0,305	0,196	0,321	0,564	1												
<b>jul1</b>	0,272	0,569	0,379	0,420	0,252	0,393	0,273	0,526	1											
<b>jul2</b>	0,268	0,351	0,558	0,251	0,335	0,273	0,304	0,456	0,408	1										
<b>ago1</b>	0,165	0,301	0,376	0,463	0,381	0,472	0,286	0,424	0,250	0,369	1									
<b>ago2</b>	0,066	0,076	0,453	0,235	0,520	0,202	0,418	0,304	0,078	0,279	0,655	1								
<b>sep1</b>	0,435	0,188	0,300	0,153	0,206	0,113	0,192	0,224	0,167	0,278	0,174	0,134	1							
<b>sep2</b>	0,142	0,131	0,378	0,367	0,308	0,277	0,223	0,310	0,125	0,346	<b>0,680</b>	0,592	0,201	1						
<b>oct1</b>	0,324	<b>0,696</b>	0,209	0,152	0,134	0,405	0,119	0,391	0,464	0,489	0,306	0,059	0,171	0,261	1					
<b>oct2</b>	0,320	0,469	0,256	0,318	0,195	0,233	0,265	0,627	0,322	0,498	0,486	0,219	0,225	0,345	0,546	1				
<b>nov1</b>	0,227	0,392	0,165	0,111	0,094	0,327	0,171	0,540	0,322	0,443	0,459	0,141	0,152	0,313	0,685	0,590	1			
<b>nov2</b>	0,341	0,579	0,434	0,305	0,382	0,286	0,269	0,590	0,418	0,332	0,412	0,245	0,266	0,240	0,402	0,608	0,299	1		
<b>dic1</b>	0,123	0,073	0,277	0,235	0,330	0,147	0,476	0,433	0,198	0,156	0,356	0,443	0,111	0,522	0,130	0,106	0,107	0,211	1	
<b>dic2</b>	0,244	0,343	0,448	0,184	0,302	0,113	0,434	0,398	0,260	0,783	0,454	0,430	0,359	0,454	0,347	0,465	0,391	0,404	0,237	1

Tabla 7. Matriz de correlaciones vectoriales



La descomposición en valores y vectores propios de la matriz permite obtener una imagen euclídea en dos dimensiones de las veinte tablas analizadas. El análisis de los valores propios obtenidos tras el proceso de diagonalización de la matriz de correlaciones vectoriales indica como puede observarse en la tabla siguiente que el primer eje principal absorbe el 33,18% de la variabilidad y llegando a explicar los tres primeros ejes el 56,49% de varianza (Tabla 8).

<b>Eje</b>	<b>Valor Propio</b>	<b>%Inercia</b>	<b>%Acumulado</b>
1	1,46e+00	33,18	33,18
2	5,26e-01	11,97	45,15
3	4,99e-01	11,34	56,49
4	3,74e-01	8,51	65,00
5	3,06e-01	6,96	71,96
6	3,02e-01	6,87	78,83
7	2,74e-01	6,23	85,06
8	2,57e-01	5,84	9,90
9	2,07e-01	4,70	95,60
10	9,82e-02	2,23	97,83
11	7,10e-02	1,62	99,45
12	2,44e-02	0,55	1,00

Tabla 8. Valores Propios Interestructura

La representación gráfica (figura 6) de los distintos meses en las dos tomas de muestreo, muestra una similitud entre ellas. Así, como se dijo anteriormente los meses de octubre en la primera toma de la muestra y la segunda del mes de febrero muestran un ángulo muy pequeño y por tanto similitud entre ellos. Los puntos más alejados y por tanto menos similares, serían los meses de agosto y febrero en la segunda muestra.

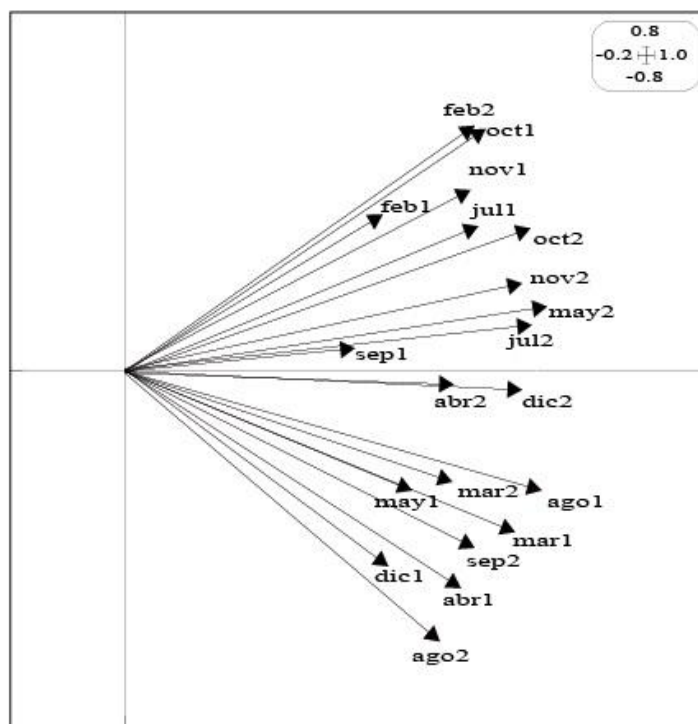


Figura 6. Interestructura de las 20 matrices del estudio

Podemos observar que los meses que están más alejados del operador WD (en longitud formando el coseno del ángulo con WD menor), son febrero y agosto en la segunda toma de la muestra, y octubre en la primera, lo que significa que estos meses son los que tienen mayor influencia en el comportamiento de las variables físico-químicas y por tanto son los que más se distancian de la matriz compromiso. Estos resultados no son concordantes con el estudio realizado por Márquez et al. (1992) en el embalse de Uribante (Venezuela) en el que señalaban que los meses con mayor influencia en el comportamiento de las variables físico-químicas eran marzo, septiembre y noviembre.

Podemos ver que la matriz compromiso es muy similar a las matrices de datos de los meses de abril, julio y diciembre en la segunda toma de la muestra.

Se observa además que la longitud de los vectores la mayoría de los meses es similar y los ángulos entre estos vectores son pequeños. Puede concluirse por tanto, que existe una estructura de covariación común entre las variables de los diez meses.

## 2.7.2 COMPROMISO

El siguiente paso del análisis es definir la matriz compromiso, resumiendo la información proveniente de todas las matrices. Sus elementos son un consenso de los elementos de cada matriz, ya que el compromiso debe ser un buen resumen de los mismos.

La descomposición de la matriz compromiso en valores y vectores singulares genera una imagen euclídea de espacio reducido, en el cual nuestras variables se proyectan y así podemos conocer las relaciones que manifiestan.

En la figura siguiente se muestra la inercia recogida por cada una de las componentes principales de la configuración compromiso, teniendo en cuenta todos los puntos de muestreo. Según la regla del codo, como se puede observar, sería suficiente seleccionar las tres primeras componentes pues recogen el 56,49% de la variabilidad.

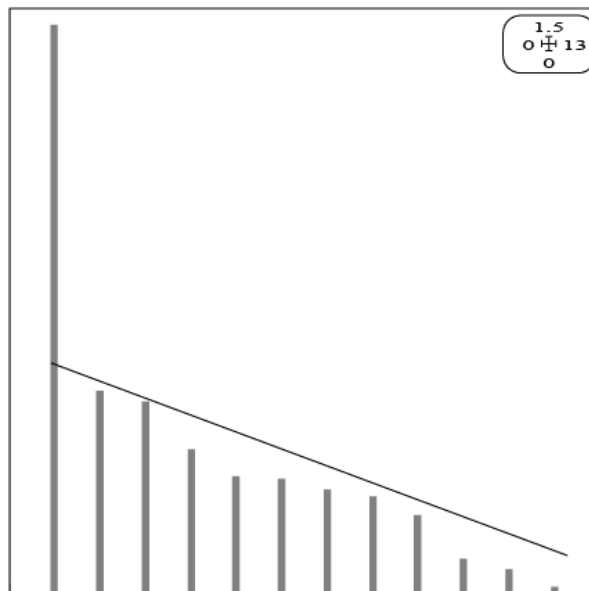


Figura 7. Inercia absorbida por las componentes principales de la configuración compromiso

Con 2 ejes, recogemos el 45,15% de la información total, asimismo podemos apreciar que es la primera componente la que lleva la mayoría de la información con un 33,18%. Los valores propios se muestran en la tabla siguiente:

<b>Eje</b>	<b>Valor Propio</b>	<b>%Inercia</b>	<b>%Acumulado</b>
1	1,46E+00	+0.3318	+0.3318
2	5,26E+03	+0.1197	+0.4515
3	4,99E-01	+0.1134	+0.5649
4	3,74E-01	+0.0851	+0.6500
5	3,06E-01	+0.0696	+0.7196
6	3,02E-01	+0.0687	+0.7883
7	2.74E-01	+0.0623	+0.8506
8	2.57E-01	+0.0584	+0.9090
9	2.07E-01	+0.0470	+0.9596
10	9.82E-02	+0.0223	+0.9783
11	7.10E-02	+0.0162	+0.9945
12	2.44E-02	+0.0055	+1.000

Tabla 9.Valores Propios del Compromiso

La tabla de información referente al proceso de construcción de la matriz compromiso es la siguiente:

Medición	Observaciones	Weights	NS norm <sup>2</sup>	Cos <sup>2</sup>
feb1	6	1,64e-01	4,33e-01	0.120
feb2	6	2,22e-01	4,80e-01	0.351
mar1	6	2,49e-01	3,25e-01	0.291
mar2	6	2,09e-01	5,04e-01	0.198
abr1	6	2,13e-01	3,99e-01	0.167
abr2	6	2,10e-01	5,66e-01	0.254
may1	6	1,83e-01	4,15e-01	0.116
may2	6	2,69e-01	3,44e-01	0.431
jul1	6	2,21e-01	4,51e-01	0.279
jul2	6	2,60e-01	3,45e-01	0.339
ago1	6	2,65e-01	4,92e-01	0.398
ago2	6	2,01e-01	5,99e-01	0.264
sep1	6	1,46e-01	4,92e-01	0.056
sep2	6	2,22e-01	4,45e-01	0.334
oct1	6	2,30e-01	4,47e-01	0.332
oct2	6	2,58e-01	4,82e-01	0.405
nov1	6	2,21e-01	6,34e-01	0.277
nov2	6	2,53e-01	5,42e-01	0.438
dic1	6	1,68e-01	4,24e-01	0.131
dic2	6	2,53e-01	3,38e-01	0.356

Tabla 10. Información de la matriz compromiso

Esta tabla presenta los valores de cada matriz sobre el compromiso. La primera columna *meses* hace referencia a cada matriz. La columna *Weights* nos indica el peso que adquiere cada matriz en la construcción del compromiso. La siguiente *NS Norm* es la norma al cuadrado, nos indica que a mayor valor, mayor diferencia entre la primera componente y el resto. Y por último *Cos<sup>2</sup>* nos muestra los cosenos al cuadrado, indicando la calidad de representación que tiene cada matriz en el compromiso (valores próximos a 1, tienen buena representación en el compromiso).

La representación compromiso de las relaciones entre las variables físico-químicas y biológicas se muestra en la figura 8, donde se pueden visualizar grupos de tres variables que presentan altas correlaciones positivas y que reflejan composiciones físico-químicas y biológicas:

- Nitrato, pH, turbiedad. Los nitratos son parte del componente inorgánico presente en el agua, estos iones pueden estar asociados a las partículas en suspensión que le dan turbiedad al agua. A su vez estos iones son usados como una fuente de alimento por el fitoplancton y las plantas acuáticas. Este parámetro tiene mucha influencia en una serie de reacciones que ocurren en el agua.
- P-total, oxígeno, N-total. Enriquecimiento de las aguas por nutrientes como nitrógeno y fósforo.
- Toxina, Clorofila a y Temperatura. Las variables clorofila a y toxina (microcistina) se consideran un potencial de riesgo sobre la base de un crecimiento excesivo de microorganismos llamados algas y cianobacterias, organismos que poseen clorofila a y que son capaces de producir toxinas. La Temperatura juega un papel importante en el desarrollo de la biomasa fitoplanctónica.

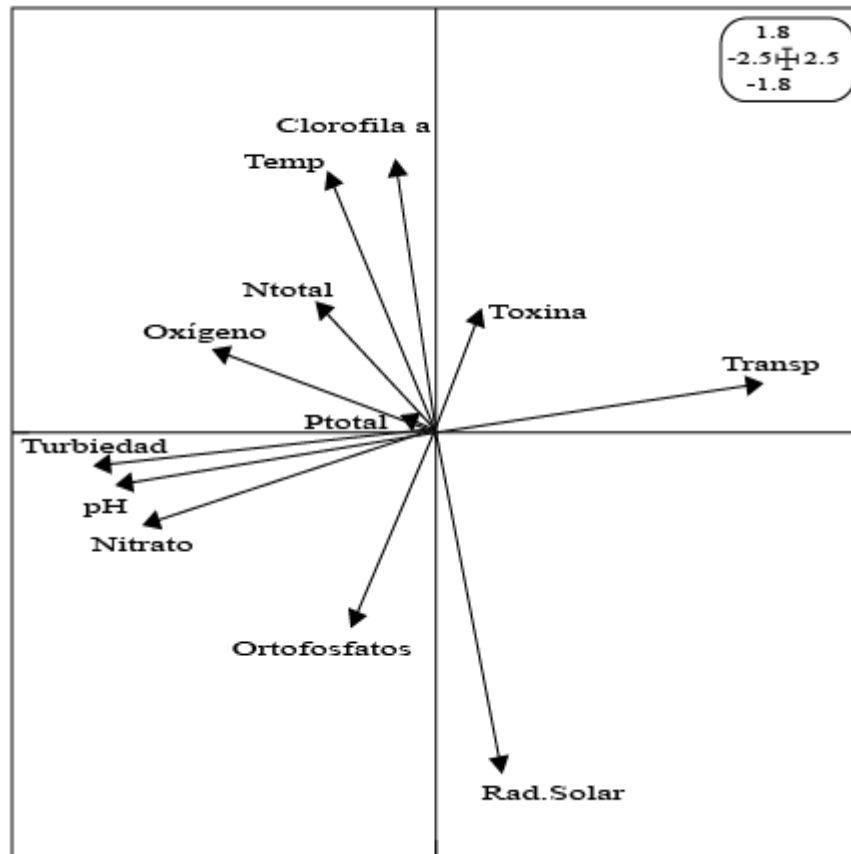


Figura 8. Representación de las variables compromiso (variables físico-químico y biológicas)

En la estructura del compromiso, las variables turbiedad, pH y nitrato, se muestran muy correlacionados, al igual que las variables clorofila a y temperatura. La variable radiación solar está correlacionada negativamente con las variables temperatura y clorofila a. La radiación tiene un efecto que puede ser letal para algunos microorganismos del agua por tal razón éstos se encuentran en capas donde la temperatura y la intensidad lumínica es óptima para que realicen sus funciones vitales (ROLDAN, 2008). De igual manera las variables ortofosfato y toxina, están correlacionadas negativamente. El ortofosfato es un nutriente limitante en el medio, las toxinas sólo están presentes cuando una condición ambiental influye sobre su producción y liberación.

No hay correlación entre las variables radiación solar y turbiedad, al formar un ángulo de 90°.

El eje 1 se observa que tiene alta correlación con las variables turbiedad, pH, nitrato y transparencia (Transp) y el eje 2 tiene correlaciones altas con la clorofila a, radiación solar, ortofosfatos y temperatura (Temp).

Asimismo, todas las variables presentan buena representación euclídea en el plano 1-2, al presentar mayor longitud del vector, con excepción de la variable P-total que presenta un vector corto, lo cual quiere decir que seguramente esté relacionado con el tercer eje.

Un primer análisis de la figura 9, nos muestra que en los diez meses de estudio puede haber una tendencia a la agrupación en los diferentes meses como ya sucedía en el HJ-biplot.

Observamos como en la primera obtención de las réplicas en los meses de marzo, abril, mayo y diciembre y segunda de los meses de agosto y diciembre, el área de Gamboa presenta valores elevados en la toxina y transparencia y bajos valores en turbiedad y pH. Algo similar ocurrió en el año 2010 en la mayoría de las estaciones del embalse Alhajuela donde se detectaron aumentos en las concentraciones de sólidos totales disueltos (STD) y los valores de conductividad, así como una disminución de la turbiedad (AUTORIDAD DEL CANAL DE PANAMÁ, 2011).

Sin embargo, el área de Paraíso se caracteriza en estos mismos meses por altas concentraciones de turbiedad, pH, nitratos y ortofosfatos. Esto está unido a una escasa transparencia del cuerpo de agua en estos meses dado que los niveles del embalse son más bajos en verano. Los niveles altos de turbiedad pudieran ser causados por las diferentes partículas suspendidas observadas en el agua (tierra, sedimentos y aguas residuales) cuando entra el invierno.

En la misma línea, el mes de septiembre en la segunda recogida de la muestra presenta valores altos de turbiedad, pH, nitratos y ortofosfatos en Paraíso, y en Gamboa altos valores de transparencia y toxinas pero a diferencia de los meses anteriormente mencionados ahora se da una alta radiación solar. En esta misma línea ARIEL Y COL. (2014) en su trabajo sobre la determinación de

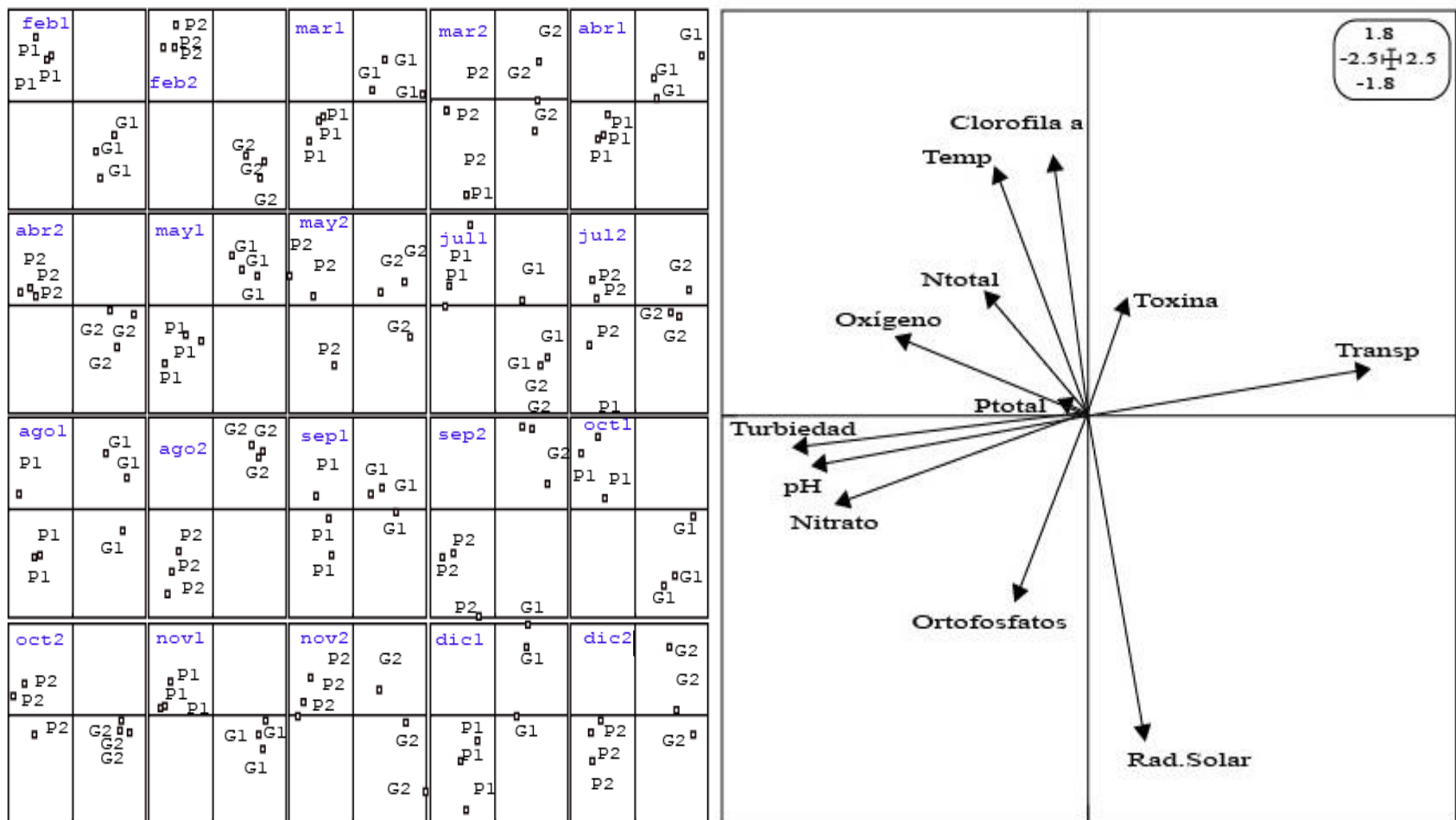


Indicadores fisicoquímicos y microbiológicos de calidad del agua superficial en la Bahía de Manzanillo los valores más bajos de pH y turbiedad se registraron en la estación predominante en el periodo lluvioso.

En los meses de febrero, julio, octubre y noviembre (primer muestreo en Paraíso y Gamboa) se pueden apreciar concentraciones elevadas de clorofila a, nitrógeno total, oxígeno y altas temperaturas en la zona de Paraíso y una alta radiación solar en el área de Gamboa. Comportamiento similar ocurre en los meses de febrero, abril, octubre y noviembre en el segundo muestreo de estos meses en Paraíso. Sin embargo, en el informe presentado de calidad de agua en la Cuenca Hidrográfica del Canal de Panamá 2003-2005, la temperatura en todas las estaciones del lago se encuentra por encima de 29°C, con excepción de las estaciones Gamboa y Toma de Agua de Paraíso en donde las temperaturas son menores. Las áreas de Gamboa y Paraíso son las que presentaron mayores concentraciones de nitratos, fosfatos y menores concentraciones de oxígeno disuelto sin que la mediana sobrepase los valores guía recomendados para evitar eutrofización y para uso recreativo con contacto directo (AUTORIDAD DEL CANAL DE PANAMÁ, 2006).

A diferencia de los anteriores, los meses de agosto y septiembre en el primer muestreo de estos meses, presentan en el área de Paraíso altos valores de ortofosfatos y nitratos, elevadas temperaturas y clorofila a, y el área de Gamboa altos niveles de transparencia y toxina. En concordancia con estos meses están marzo, mayo y julio en la segunda toma de muestra en ambas áreas.

Figura 9. Representación conjunta



## **2.8 COMPARACIÓN DE LAS HERRAMIENTAS ESTADÍSTICAS UTILIZADAS: HJ-BIPLLOT Y STATIS DUAL**

En este trabajo se ha realizado el estudio sobre un mismo conjunto de datos, es decir, variables físico-químicas y biológicas medidas en los diferentes meses del año 2009 para la determinación de la calidad del agua, desde dos perspectivas distintas aunque complementarias: tabla de dos vías (HJ-Biplot) y tabla múltiple (STATIS-dual).

El primer análisis que realizamos fue un HJ-Biplot, que nos facilitó la interpretación en las relaciones de similitud/disimilitud entre los individuos (puntos de muestreo), la covariación entre las variables físico-químicas y biológicas y sobretodo nos determinó qué variables marcan las diferencias entre los diferentes puntos de muestreo (Galindo, 1986). Aunque el HJ-Biplot tiene un carácter descriptivo nos ha permitido ver en un plano de forma sencilla y clara, la representación conjunta entre las variables físico-químicas y biológicas en los distintos meses en ambas areas de estudio. Esta técnica nos ha permitido por tanto, obtener relaciones entre las variables físico-químicas y biológicas, entre los distintos puntos de muestreo en los distintos meses y entre ambos conjuntamente.

Al igual que el HJ-Biplot el método STATIS-dual representa la estructura de covariación de las variables pero la diferencia se encuentra en el espacio en el que se proyecta, que en el caso del STATIS-dual se hace en un espacio promedio que recoge las características medias de las distintas situaciones de estudio, por eso se aprecian diferencias entre las dos representaciones. Comparando la representación de las variables obtenidas por ambos métodos se observa que en ambas imágenes la clorofila a mantiene una asociación importante con el oxígeno y la turbiedad con los nitratos y ortofosfatos.

No obstante, una vez analizados con detalle los resultados obtenidos en ambos estudios, se puede deducir que ante el estudio de un conjunto de información, referida al mismo colectivo de individuos, susceptible de ser analizada como una sola matriz o como un grupo de subtablas, los objetivos planteados en

dicho trabajo deben ser los que determinen en primer lugar, cuál de las dos posibles estructuras seleccionar. Así si se desea obtener una visión global de las asociaciones entre las variables físico-químicas y los distintos puntos de muestreo, el análisis del HJ-Biplot es suficiente. Si además se desea, obtener una visión parcial realizando una comparación global de las subtablas, es decir, poner de manifiesto los grupos con estructura interna similar/distinta, será necesario optar por el análisis de la tabla múltiple, y realizar un STATIS-dual, ya que el HJ-biplot no reproduce la matriz de datos original, por lo que es necesario aplicar el STATIS-dual para describir la estructura longitudinal de los datos.

## 2.9 CONCLUSIONES

1. Los clusters, resultantes del análisis HJ-Biplot de la matriz de las variables fisicoquímicas y biológicas, evidencian dos grupos de variables de acuerdo a los meses de la estación climática de la región.
2. Se pone de manifiesto, como el HJ-Biplot (GALINDO, 1986), se configura como una herramienta eficiente para los estudios concernientes a calidad de agua, mostrando que los meses de febrero y agosto en la segunda toma de la muestra, y octubre en la primera, son los que tienen mayor influencia en el comportamiento de las variables físico-químicas.
3. Los resultados obtenidos con la aplicación del Statis, permiten explicar la existencia de asociaciones entre variables y las diferencias en cuanto a calidad de agua entre los lugares donde se efectuaron las colectas de las muestras.
4. Gamboa y Paraíso son diferentes en relación a la calidad de agua, en algunos meses del año.
5. La correlación positiva en las variables Ptotal, oxígeno y Ntotal, expresa un enriquecimiento de las aguas por nutrientes como nitrógeno y fósforo.
6. La correlación positiva entre las variables temperatura, toxina (microcistina) y clorofila *a*, sugiere un afloramiento de cianobacterias.
7. El pH, la transparencia y el oxígeno, se asocian principalmente en los meses de febrero, marzo, abril y mayo (meses de la estación seca).
8. Se reflejan mayores cambios de la calidad de agua durante los meses de la estación lluviosa, nitrato, ortofosfatos, turbiedad y Ptotal.
9. El estudio de la Calidad del Agua y en especial de aquellas aguas destinadas al uso o consumo humano, resulta imprescindible para garantizar su buen estado y la seguridad de todas aquellas personas que vayan a hacer uso de ésta, así como las condiciones óptimas de la misma para mantener la biodiversidad de las especies que habitan en su torno.

### 3. BIBLIOGRAFÍA

---

Abdi, H., Valentin, D., Chollet, S., & Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality Prefer*, 18, 627–640.

Abdi, H., & Valentin, D. (2007). How to analyze multiple distance matrices. En N. Slakind(Ed.), *Encyclopedia of Measurement and Statistics*, pp. 1-15. Thousand Oaks, CA: SAGE.

Abdi, H., Williams, L.J., Connolly, A.C., Gobbin, M.I., Dunlop, J.P., & Haxby, J.V. (2012). Multiple Subject Barycentric Discriminant Analysis (MUSUBADA): How to Assign Scans to Categories without Using Spatial Normalization. *Computational and Mathematical Methods in Medicine*, 2012(ID 634165),15 pp. doi:10.1155/2012/634165.

Abdi, H., Williams, L.J., Valentin, D., & Bennani-Dosse, M. (2012). STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling. *WIREs Comput Stat*, 4, 124-167. doi: 10.1002/wics.198.

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2), 139-177.

Aitchison, J. (1986). *The statistical analysis of compositional data*. London: Chapman & Hall.

Aitchison, J. (1997). The one-hour course in compositional data analysis or compositional data analysis is simple. En V. Pawlowsky-Glain (Ed). *Proceedings of IAMG'97* (pp.3-35). Barcelona: International Center for Numerical Methods in Engineering (CIMNE).

Aitchison, J., & Greenacre, M. (2002). Biplots of compositional data. *Journal of the Royal Statistical Society. Series C ( Applied Statistics)*, 51(4), 375-392.

Alba-Tercedor, J. (1996). Macroinvertebrados acuáticos y calidad de las aguas de los ríos. IV Simposio del Agua en Andalucía (SIAGA). *Ameria*. 2: 203-213.

Alperin, M., Echeveste, H., Fernández, R., & Bellieni, G. (2007). Análisis estadístico de datos geoquímicos de volcanitas jurásicas del Macizo del Deseado, provincia de Santa Cruz. *Revista de la Asociación Geológica Argentina*, 62(2), 200-209.

Amaro, I., Vicente-Villardón, J., & Galindo-Villardón, M.P. (2004). MANOVA biplot for treatment arrays with two factors based on multivariate general linear models. *Interciencia*, 29(1), 26-32.

Araya, C.C., (2010). Modelo de clases latentes en tablas poco ocupadas: Una contribución basada en bootstrap. [Tesis Doctoral]. Universidad de Salamanca, España.

Arcidiacono, C., Sarnacchiaro, P., & Velleman, R. (2008). Testing fidelity to a new psychological intervention for family members of substance misusers during implementation in Italy. *Journal of Substance Use*, 13(6), 361–381. doi:10.1080/14659890802211291.

Ariel A.; Grey G.; Viccelda de Franco; Castillero M (2014). Determinación de Indicadores Fisicoquímicos y Microbiológicos de calidad del agua superficial en la Bahía de Manzanillo. *RIDTEC*, Vol10, nº 1.

Autoridad del Canal de Panamá (2006). Informe de Calidad de Agua 2003-2005. Departamento de Ambiente, Agua y Energía.

Autoridad del Canal de Panamá. (2008). Agua y Bosques en la Cuenca del Canal: Tendencia a Largo Plazo.

Autoridad del Canal de Panamá. (2008). Informe de Calidad de Agua de la Cuenca del Canal 2007.

Autoridad del Canal de Panamá. Departamento de Ambiente, Agua y Energía, División de Agua, Unidad de Calidad de Agua. (2011). Informe de Calidad de Agua de la Cuenca del Canal 2010.

Autoridad Nacional del Ambiente. (2006). Manual de curso de capacitación, Elaboración de las normas de calidad ambiental para aguas naturales. Calidad de aguas naturales características principales, su gestión y control. Contrato BID N° PAN61 – 2005, CSI Ingenieros. Panamá.

Baccala, N. (2004). Contribuciones al análisis de datos multivía: Tipología de las variables. [Tesis Doctoral]. Universidad de Salamanca, España.

Becker, H.C. (1981). Correlations among some statistical measures of phenotypic stability. *Euphytica*. 30, 835-840.

Beamonte, E.; Casino, A.; Veres, E. J. (2004). La calidad del agua en ciertas estaciones de control del canal Júcar-Turia (período 1994-2001), en *Revista Española de estudios Agrosociales y Pesqueros*, número 201; pp. 105-126.

Beamonte, E.; Bermúdez, J.; Casino, A.; Veres, E. (2007). A statistical study of the quality of surface water intended for human consumption near Valencia (Spain), en *Journal of Environmental Management*, número 83; pp. 307-314.

Benasseni, J., & Dosse, M.B. (2012). Analyzing multiset data by the Power STATIS-ACT method. *Advances in Data Analysis and Classification*, 6(1), 49–65. doi:10.1007/s11634-011-0085-8.

Benzécri, J. P. (1973). *L'Analyse des Données*. Paris: Dunod. Vol. 2. *L'analyse des correspondances*. Paris: Dunod.

Billheimer, D., Guttorp, P., & Fagan, W. (2001). Statistical interpretation of species composition. *J. Am. Stat. Assoc.*, 96 (456), 1205-1214.

Bradu, D., & Gabriel, K.R. (1974). Simultaneous statistical inference on interactions in two-way analysis of variance. *Journal of the American Statistical Association*, 69(346), 428-436.

Bradu, D., & Gabriel, K.R. (1978). The biplot as a diagnostic tool for models of two-way tables. *Technometrics*, 20(1), 47-68.



Caballero-Julia, D. (2011). El HJ-Biplot como herramienta en el análisis de grupos de discusión. Calidad de vida en la ludopatía: una propuesta sociológica.[Trabajo de Fin de Master]. Universidad de Salamanca, España.

Camargo-Buitrago, I., Intire, E.M., & Gordón Mendoza, R. (2011). Identificación de mega-ambientes para potenciar el uso de genotipos superiores de arroz en Panamá. *Pesquisa Agropecuaria Brasileira* 46(9), 1061-1069.

Cárdenas, O., Noguera, C., Galindo, M. P., & Vicente-Villardón, J. (2003). El uso de información externa en aproximaciones Biplot. *Revista Venezolana de Análisis de Coyuntura*, 9(2), 257-276.

Carlier, A., & Kroonenberg, P.M. (1996). Descompositions and biplots in three-way correspondence analysis. *Psychometrika*, 61(2), 355-373.

Chessel, D., & Mercier, P. (1993). Couplage de triplets statistiques et liaisons espèces- environnement. En J. Lebreton & B. Asselain (Eds.), *Biométrie et Environnement*, pp.15-44. Paris: Masson.

Chessel, D., & Hanafi, M. (1996). Analyses de la Co-Inertie de K nuages de points. *Revue de Statistique Appliquée*, 44(2), 35- 60.

Cox, C., & Gabriel, K.R. (1982). Some comparisons of Biplot display and pencil-and-paper E.D.A. methods. En R.L. Launer & A. F. Siegel (Eds). *Modern data analysis* (pp. 45-82). London: Academic Press.

Crossa, J., Gauch, H.G. Jr., & Zobel, R.W. (1990). Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. *Crop Sci*, 30 (3Cubilla-Montilla, M.I. (2014). Sinopsis y avances del biplot. [Trabajo Fin de Máster]. Universidad de Salamanca, España.

Daszykowski, M., & Walczak, B. (2011). Methods for the exploratory analysis of two- dimensional chromatographic signals. *Talanta*, 83(4),1088–1097. doi:10.1016/j.talanta.2010.08.032.

Demey, J.R., Vicente Villardón, J.L., Galindo Villardón, M.P., & Zambrano, A.Y. (2008). Identifying molecular markers associated with classifications of genotypes by external logistic biplot. *Bionformatics*, 24(24), 2832-2838.

- Díaz-Faes, A.A., González-Albo, B., Galindo-Villardón, M.P., & Bordons, M. (2013). HJ-Biplot como herramienta de inspección de matrices de datos bibliométricos. *Revista Española de Documentación Científica*, 36, 1-16.
- Dolédec, S., & Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater Biology*, 31(3), 277–294.
- Dray, S., & Dufour, A. B. (2007): The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 1-20.
- Eckart, C., & Young, G. (1939). A principal axis transformation for non-Hermitian matrices. *Bulletin of the American Mathematical Society*, 45(2), 118-121.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.
- Efron, B., & Tibshirani, R.J. (1993). *An introduction into the bootstrap*. New York: Chapman and Hall.
- Egido, J.F. (2015). *Biplot Dinámico*. Universidad de Salamanca. <http://gredos.usal.es/jspui/handle/10366/125246>
- Egozcue, J.J., & Pawlowsky Glahn, V. (2011). Análisis composicional de datos en Ciencias Geoambientales. *Boletín Geológico y Minero*, 122(4), 439-452.
- Ersan, Y., Karatas, A., Carkman, S., Cicek, Y., & Erguney, S. (2009). Late Results of Patients Undergoing Remedial Operations for Alkaline Reflux Gastritis Syndrome. *Acta Chirurgica Belgica*, 109(3), 364–370.
- Escobar, M., Berti, M., Matus, I., Tapia, M., & Johnson, B. (2011). Genotype Environment Interaction in Canola (*Brassica napus* L.) Seed Yield in Chile. *Chilean Journal of Agricultural Research*, 71(2), 175-186.
- Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, 750-760.
- Escoufier, Y. (1976). Opérateur associé à un tableau de données. *Annales de l'Insee*, 22-23, 165-178.

Escofier, B. and Pagès, J. (1990). Multiple factor analysis. *Computational Statistics & Data Analysis*, 18, 121–140.

Escoffier B., Pagès J. (1994). Multi factor analysis (AFMULT pack-age). *Comput. Stat. Data Anal.* 18, 121–140.

Ferraro, S., Ardoino, I., Bassani, N., Santagostino, M., Rossi, L., Biganzoli, E., & Panteghini, M. (2013). Multi-marker network in ST-elevation myocardial infarction patients undergoing primary percutaneous coronary intervention: When and what to measure. *Clinica Chimica Acta*, 417, 1–7. doi:10.1016/j.cca.2012.12.001.

Fogliatto, F.S., & Niang, N. (2009). *Multivariate Statistical Control of Batch Processes with Variable Duration*. New York: IEEE.

Fournier, M., Motelay-Massei, A., Massei, N., Aubert, M., Bakalowicz, M., & Dupont, J. P. (2009). Investigation of Transport Processes inside Karst Aquifer by Means of STATIS. *Ground Water*, 47(3), 391-400. doi:10.1111/j.1745-6584.2008.00532.x.

Frutos, E., & Galindo, M.P. (2013). GGEBiplotGUI: Interactive GGE Biplots in R. Versión 1.0-6. Recuperado de <http://cran.rproject.org/web/packages/GGEBiplotGUI/index.html>

Gabriel, K.R. (1971). The biplot-graphic display of matrices with application to principal component analysis. *Biometrika* Vol 58, 453-467.

Gabriel, K.R. (1972). Análisis of meteorological data by means of canonical decomposition and Biplots. *Journal of Applied Meteorology*, 11, 1071-1077.

Gabriel, K.R. (1972). Análisis of meteorological data by means of canonical decomposition and Biplots. *Journal of Applied Meteorology*, 11, 1071-1077.

Gabriel, K.R., & Odoroff, C.L. (1990). Biplots in biomedical research. *Statistics in Medicine*, vol. 9, 469-485.

Gallego-Álvarez, I., & Vicente-Villardón, J.L. (2012). Analysis of environmental indicators in international companies by applying the logistic biplot. *Ecological Indicators*, 23, 250-261.

Gallego-Álvarez, I., Galindo-Villardón, M.P., & Rodríguez-Rosa, M. (2014). Analysis of the sustainable society index worldwide: A study from the Biplot perspective. *Social indicators Research* (in press). \*dx.doi.org/10.1007/s11205-014-0579-9.

Galindo-Villardón M.P. (1986). Una alternativa de representación simultánea: HJ-Biplot. *Qüestió*, vol. 10 (1),13-23.

Galindo-Villardón, M.P., & Cuadras, C. M. (1986). Una extensión del método Biplot y su relación con otras técnicas. *Publicaciones de Bioestadística y Biomatemática*. n° 17. Universidad de Barcelona.

García-González, J.J., Correa-Londoño, G.A., & Pardo-Carrasco, S.C. (2012). Phytoplankton and periphyton in ponds with Nile tilapia (*Oreochromis niloticus*) and bocachico (*Prochilodus magdalenae*). *Revista Colombiana de Ciencias Pecuarias*, 25(4), 630-614.

Gauch, H.G. (1988). Model selection and validation for yield trials with interaction. *Biometrics*, 44, 705-715.

Gauch, H.G. (2006). Statistical analysis of yield trials by AMMI and GGE. *Crop Sci*, 46, 1488-1500.

Ghezzi, I. (2011). El análisis composicional en el estudio de la producción y distribución de la cerámica prehispánica. *Bulletin de l'Institut Francais d'Études Andines*, 40(1), 1-29.

Gollob, H. (1968), "A statistical model wich combines features of factor analytic and analysis of variance techniques", *Psichometrika*, 33: 73-115.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.

Gower, J. C. (1990). Fisher's optimal scores and multiple correspondence analysis. *Biometrics*, 46, 947-961.

Gross, R.R., Izaguirre, Y.P., Isaac, M.T., Liranza, N.B., Suárez, H.J., & Pérez, H.G. (2012). Métodos estadísticos multivariados en el estudio de la interacción genotipo ambiente en caña de azúcar. *Ciencia en su PC*, (1), 47-60.

Guebel, D.V., Canovas, M., & Torres, N.V. (2009). Model Identification in Presence of Incomplete Information by Generalized Principal Component Analysis: Application to the Common and Differential Responses of *Escherichia coli* to Multiple Pulse Perturbations in Continuous, High-Biomass Density Culture. *Biotechnology and Bioengineering*, 104(4), 785–795. doi:10.1002/bit.22438.

Hernández, M. (2011). Estudio de la variación estacional del sabor del tomate.[Trabajo Fin de Master]. Universidad de Salamanca, España.

Hernández, S. (2005). Biplots Robustos. [Tesis Doctoral]. Universidad de Salamanca, España.

Hernández, S., & Galindo-Villardón, M.P. (2006). BIPROB: Un método para obtener Biplot robusto. *Investigación Operacional*, 27(3), 34-46.

Iachine, I., Petersen, H.C., & Kyvik, K.O. (2010). Robust tests for the equality of variances for clustered data. *Journal of Statistical Computation and Simulation*, 8 (4), 365–37. doi:10.1080/00949650802641841.

Ibañez, M.A., Cavanagh, M. M., Bonamico, N.C., & Di Renzo, M. A. (2006). Análisis gráfico mediante biplot del comportamiento de híbridos de maíz. *Revista de Investigaciones Agropecuarias*, 35, 83-93.

International, Resources Group, LTD. (2000). Manejo Integral de la Cuenca.

Iñigo, A. C., García-Talegón, J., Vicente-Tavera, S., Martín-González, S., Casado-Marín, S., Vargas-Muñoz, M., & Pérez-Rodríguez, J. L. (2013). Colour and ultrasound propagation speed changes by different ageing of freezing/thawing and cooling/heating in granitic materials. *Cold Regions Science and Technology*, 85, 71-78.

Jaffrenou, P.A. (1978). Sur L'Analyse des Familles Finies de Variables Vectorielles: Bases Algébriques et Applications à la Description Statistique. [Thèse de Troisième Cycle]. Université de Lyon, France.

Karavoltzos, S.; Sakellari,A.; Mihopoulos, N.; Dassenakis, M.; Scoullou, M.J. (2008). Evaluation of the quality of drinking water in regions of Greece, en *Desalination*, número 224: pp. 317-329.

Kikuchi, M., & Guillén, K. (2010). Método bootstrap de factores estratificado. *Revista Mexicana de Investigación*, 3(5), 37-50.

Krishnan, A., Williams, L.J., McIntosh, A.R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review.

*Neuroimage*, 56(2), 455–475. doi:10.1016/j.neuroimage.2010.07.034.

Kroonenberg, P.M. (1983). Three-mode principal component analysis: Theory and applications. Leiden: DSWO Press.

Krzanowski, W.J. (1979). Between-groups comparison of principal components. *Journal of the American Statistical Association*, 74(367), 703-707.

Lavit, C. (1988). *Analyse Conjointe de Tableaux Quantitatifs*. Massons. Paris.

Lavit, C., Escoufier, Y., Sabatier, R., & Traissac, P. (1994). The ACT (Statis method). *Computational Statistics and Data Analysis*, 18, 97–119.

L'Hermier des Plantes, H. (1976). Structuration des Tableaux à Trois Indices de la Statistique. [Thèse de Troisième Cycle]. University of Montpellier, France.

L'Hermier des Plantes, H., & Thiébaud, B. (1977) Étude de la pluviosité au moyen de la méthode STATIS. *Revue de Statistique Appliquée*, 25, 57–81.

Luengo, A. (2010). Cianobacterias tóxicas y mortandad de aves en las Marismas de Txingudi. Informe de situación junio 2010. Gobierno Vasco. Departamento de Medio Ambiente, Planificación Territorial, Agricultura y Pesca

Mandel, J. (1971). A new analysis of variance model for non-additive data. *Technometrics*, 13, 1-18.

Manuel- González, V., Ávila, E., & Olaya, J. (2009). Analysis of the AFM and STATIS representation spaces in the Bogota's 1993 biographical mobility survey. *Revista Colombiana De Estadística*, 32(1), 1–15.

Martín-Rodríguez, J., Galindo-Villardón, M.P., & Vicente-Villardón, J.L. (2002). Comparison and integration of subspaces from a biplot perspective. *Journal of Statistical Planning and Inference*, 102(2), 411–423.

Martin-Rodríguez, J., & Vicente-Villardón, J.L. (2004). Metabiplot: Application to the Study of a Lake. En Salamanca Statistics Seminar IV. Advances in Multivariate Analysis (pp. 93-96). Salamanca: Universidad de Salamanca.

Morgado, S., & Galindo-Villardón, M.P. (2006). Statische Metabiplot. Un estudio comparativo. En XIII congreso Anual da SPE. Portugal.

Oksuz, S., Ulkur, E., Oncul, O., Kose, G.T., Kucukodaci, Z., & Urhan, M. (2013). The Effect of Subcutaneous Mesenchymal Stem Cell Injection on Stasis Zone and Apoptosis in an Experimental Burn Model. *Plastic and Reconstructive Surgery*, 131(3), 463–471. doi:10.1097/PRS.0b013e31827c6d6f.

Pawlowsky-Glahn, V., & Egozcue, J.J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, 15(5), 384-398.

Perrin, L., Symoneaux, R., Maitre, I., Asselin, C., Jourjon, F., & Pagès, J. (2008). Comparison of three sensory Mapping (R) procedure: Case of methods for use with the ten wines from Loire valley. *Food Quality and Preference*, 19(1), 1–11. doi:10.1016/j.foodqual.2007.06.005.

Pizarro, C., Esteban-Diez, I., Rodríguez-Tecedor, S., & María González-Saiz, J. (2013). A sensory approach for the monitoring of accelerated red wine aging processes using multi-block methods. *Food Quality and Preference*, 28(2), 519–530. doi:10.1016/j.foodqual.2013.01.002.

Pouliot, D., Latifovic, R., Fernandes, R., & Olthof, I. (2009). Evaluation of annual forest disturbance monitoring using a static decision tree approach and 250 m MODIS data. *Remote Sensing of Environment*, 113(8), 1749–1759. doi:10.1016/j.rse.2009.04.008.

Ramirez, I. C., Barrera, C.J., & Correa, J.C.(2013). Efecto del tamaño de muestra y el número de réplicas bootstrap. *Ingeniería y Competitividad*, 15(1), 93-101.

Regnault, A., Marfatia, S., Louie, M., Mear, I., Meunier, J., & Viala-Danten, M. (2009). Satisfactory cross-cultural validity of the ACTG symptom distress module in HIV-1-infected antiretroviral-naive patients. *Clinical Trials*, 6(6), 574–584. doi:10.1177/1740774509352515.

Rolland, A., Bertrand, F., Maumy, M., & Jacquet, S. (2009). Assessing phytoplankton structure and spatio-temporal dynamics in a freshwater ecosystem using a powerful multiway statistical analysis. *Water Research*, 43(13), 3155–3168. doi:10.1016/j.watres.2009.03.049.

Roldan, G. (2008). *Fundamentos de Limnología Neotropical*. Segunda Edición. Editorial Universidad de Antioquia. Colombia.

Rodríguez, M., & Sánchez, M. (2013). *Datos Composicionales: Una alternativa al estudio de test psicológicos*. [Trabajo Fin de Master]. Universidad de Salamanca, España.

Sabatier, R., & Vivien, M. (2008). A new linear method for analyzing four-way multiblocks tables: STATIS-4. *Journal of Chemometrics*, 22(5-6), 399–407. doi:10.1002/cem.1150.

Sauzay, L., Hanafi, M., Qannari, E.M., & Schlich, P. (2006). Analyse de K + 1 tableaux a l'aide de la méthode STATIS: application en evaluation sensorielle. En *Agro-Industrie et Méthodes Statistiques – 9èmes Journées Européennes*, Janvier 2006, pp.1-23, Montpellier, France.

Serghini, M., Boutayeb, A., Boumaaz, A., Srairi, A., Mesfioui, A., Zoubi, A., & Dridi, A. (2008). Stability of the spatial structures of demersal assemblage in the Moroccan southern Atlantic zone. *Applied Ecology and Environmental Research*, 6(1), 117–127.

Shrestha, S.; Kazama, F. (2007). Assesment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan, en *Enviroment Modelling & Software*, número 22; pp. 464-475

Simier, M., Blanc, L., Pellegrin, F., & Nandris, D. (1999). Approche simultanée de K couples de tableaux: application à l'étude des relations pathologie végétale—environnement. *Revue de Statistique Appliquée*, 47:31-46.

Stanimirova, I., Boucon, C., & Walczak, B. (2011). Relating gas chromatographic profiles to sensory measurements describing the end products of the Maillard reaction. *Talanta*, 83(4), 1239-1246. doi:10.1016/j.talanta.2010.09.018.



Stefanescu, V., Serban, F., Busu, M., & Ferrara, M. (2010). Portfolio Optimization Using Classification and Functional Data Analysis Techniques. *Economic Computation and Economic Cybernetics Studies and Research*, 44(3), 93–107.

Strauss, J. S., Gabriel, K. R., Kokes, R. F., Ritzer, B.A., VanOrd, A., & Tarana, E. (1979). Do psychiatric patients fit their diagnosis? Patterns of symptomatology as described with the Biplot, *Journal of Nervous and Mental Disease*, 167, 105-113.

Thioulouse, J., & Chessel, D. (1987). Les Analyses Multitableaux en Écologie Factorielle. 1. – De la Typologie d'État à la Typologie de Fonctionnement par l'Analyse Triadique. *Acta Ecologica, Ecologia Generalis*, 8, 463-480.

Thioulouse, J., Simier, M., & Chessel D. (2004). Simultaneous analysis of a sequence of paired ecological tables. *Ecology*, 85, 272–283.

Thioulouse, J. (2011). Simultaneous Analysis of a Sequence of Paired Ecological Tables: A Comparison of Several Methods. *Annals of Applied Statistics*, 5(4), 2300–2325. doi:10.1214/10-AOAS372.

Thomas, C. W., & Aitchison, J. (1998). The use of logratios in subcompositional analysis geochemical discrimination of metamorphosed limestones from the northeast and central Scottish Highlands. En Proceedings of IAMG98, The Fourth Annual Conference of the International Association for Mathematical Geology, De Frede, Naples (pp.549-554).

Tipán, A., & Adrian, J. (2014). Estudio comparativo del contenido nutrimental de los suelos bajo cecropia y economía de sus hojas en bosques de diferentes edades dentro del ecosistema húmedo tropical del Ecuador. [Tesis Doctoral]. Escuela de Ingeniería Agronómica, Riobamba, Ecuador.

Tsianco, M. C., & Gabriel, K. R. (1981). Modeling temperatura data: An ilustration of the use of Biplot and Bimodels in nom-linear modeling. University of Richester, Statistical Techical Report. 81/15.

Tsianco, M. C., & Gabriel, K. R. (1984). Modeling temperatura data: An ilustration of the use of Biplot and Bimodels in nom-linear modeling. *Journal of Climate and Applied Meteorology*, 23, 787-799.

Tucker, L.R. (1964). The extension of factor analysis to three-dimensional matrices. En: N. Frederiksen & H. Gulliken (Ed). Contributions to Mathematical Psychology (pp. 110-182. New York: Holt.

Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 279-31.

Tucker, L.R. (1972). Relation between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 37, 3-27.

Vallejo-Arboleda, A., Vicente-Villardón, J.L., & Galindo-Villardón, M.P. (2007). Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology. *Comp Stat Data Anal*, 51, 4193–4205.

Van Deun, K., Smilde, A.K., van der Werf, M.J., Kiers, H.A.L., & Van Mechelen, I. (2009). A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, 10, 246. doi:10.1186/1471-2105-10-246.

Van Eeuwijk, F.A., & Kroonenberg, P.M. (1998). Multiplicative models for interaction in three-way ANOVA, with applications to plant breeding. *Biometrics*, 54(4), 1315-1333.

Varas, M.J., Vicente Tavera, S., Molina, E., & Vicente Villardón, J.L. (2005). Role of canonical biplot method in the study of building stones: an example from Spanish monumental heritage. *Environmetrics*, 16(4), 405-419.

Varela, M., & Castillo, J.G. (2005). Modelos con término multiplicativo aplicación en el análisis de la interacción genotipo ambiente. *Cultivos Tropicales*, 26(3), 71-75.

Veiga de Cabo, J., & Martín-Rodero, H. (2011). Open Access: new models of scientific publishing in web 2.0 environments. *Salud colectiva*, 7, S19-S27.

Vicente-Villardón, J.L. (1992). Una alternativa a las técnicas factoriales basada en una generalización de los métodos Biplot. [Tesis Doctoral]. Universidad de Salamanca, España.

Vicente-Villardón, J.L. (2001). Biplot for Binary Data Base don Logistic. Response Surfaces. En Salamanca Statistics Seminar IV. Advances in Multivariate Analysis.

Salamanca: Universidad de Salamanca.

Vicente-Villardón, J.L., Galindo-Villardón, M.P., & Blázquez, A. (2003). Logistic Biplots. Correspondence Analysis and Related Methods. En International Conference on Correspondence Analysis and Related Methods, Barcelona 29 junio-2 julio 2003. Barcelona: Universidad Pompeu Fabra.

Vicente-Villardón, J.L., Galindo-Villardón, M.P., & Blázquez, A. (2006). Logistic Biplots. En M. & Greenacre & J. Blasius (Eds). Multiple Correspondence Analysis and Related Methods. Boca Raton: Chapman and Hall.

Vicente-Galindo, M.P. (2013). Análisis tablas de tres vías: recientes desarrollos del STATIS.[Trabajo Fin de Máster]. Universidad de Salamanca, España.

Vicente-Villardón, J.L., Galindo-Villardón, M.P., Barriuso, J., Endress, V., Martín, L., Villalobos, N., & Blázquez, A. (2007). MANOVA Biplot para la composición de azúcares neutros en *Medicago arborea* L. En XXX Congreso Nacional de Estadística e Investigación Operativa y IV Jornadas de Estadística Pública, Valladolid, España.

Vicente-Villardón, J.L. (2010). MULTBILOT: A package for Multivariate Analysis using Biplots. Departamento de Estadística. Universidad de Salamanca. <http://biplot.usal.es/ClassicalBiplot/index.html>

Vivien, M., & Sabatier, R. (2004). A Generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables. *Comp Stat Data Anal*, 46, 155-171.

Vivien, M., & Sune, F. (2009). Two four-way multiblock methods used for comparing two consumer panels of children. *Food Quality and Preference*, 20(7),472–481. doi:10.1016/j.foodqual.2009.04.004.

Wilcox, R.R. (2010). Comparing non-parametric regression lines via regression depth. *Journal of Statistical Computation and Simulation*, 80(4), 379–387. doi:10.1080/00949650802662805

Xavier, A., Socorro, M. D., & Belém, M. D. (2012). The recent changes in the Portuguese Farm's Land Uses: An application of an Approach based on HJ-Biplot. En recent researches in Environment, Energy Systems and Sustainability: 8th WSEAS

International Conference on Energy, Environment, Ecosystems & Sustainable Development. University of Algarve, Portugal, pp. 185-190.

Yan, W., Hunt, L.A., Sheng, Q., & Szlavics, Z. (2000). Cultivar evaluation and megaenvironment investigation based on GGE Biplot. *Crop Science*, 40, 597-605.

Yan, W. (2001). GGEbiplot. A Windows application for graphical analysis of multienvironment trial data and other types of two-way data. *Agronomy Journal*, 93(5), 1111-1118.

Yan, W., & Kang, M.S. (2002). *GGE Biplot: a graphical tool breeders, geneticists, and agronomists*. Boca Raton: CRC Press.

Yan, W., & Hunt, L.A. (2002). Biplot analysis of diallel data. *Crop Sci*, 42, 21-30.

Yan, W., & Rajcan, I. (2002). Biplot evaluation of test sites and trait relations of soybean in Ontario. *Crop Sci*, 42, 11-20.

Yan, W., & Kang, M.S. (2003). *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists*. Boca Raton: CRC Press.

Yan, W., Tinker, N.A. (2006). Biplot analysis of multi-environment trial data: principles and applications. *Can J Plant Sci*, 86, 623-645.

Yan, W., Kang, M.S., Ma, B., Woods, S., & Cornelius, P.L. (2007). GGE Biplot. vs. AMMI. Analysis of genotype-by-environment data. *Crop Sci*, 47, 643-655.

Yan, W. (2011). GGE biplot vs. AMMI graphs for genotype-by-environment data analysis. *Journal of the Indian Society of Agricultural Statistics*, 65(2), 181-193.

young, g. y Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances . *Psychometrika*, 3, 19-22.