

UNIVERSIDAD DE SALAMANCA

DEPARTAMENTO DE ESTADÍSTICA



**CONTRIBUCIONES AL ANÁLISIS
MULTIVARIANTE DE MATRICES DE DATOS
MULTIVÍA**

Autor:

MARIELA ALEXANDRA GONZÁLEZ NARVÁEZ

Directoras:

**MARÍA JOSÉ FERNÁNDEZ GÓMEZ
SUSANA LUÍSA DA CUSTÓDIA MACHADO
MENDES**

**DOCTORADO EN ESTADÍSTICA MULTIVARIANTE
APLICADA
TESIS DOCTORAL**

**Salamanca, España
Septiembre 2021**

UNIVERSITY OF SALAMANCA

DEPARTMENT OF STATISTICS



**CONTRIBUTIONS TO THE MULTIVARIATE
ANALYSIS OF MULTI-WAY DATA**

PhD Thesis

MARIELA ALEXANDRA GONZÁLEZ NARVÁEZ

2021

**CONTRIBUCIONES AL ANÁLISIS
MULTIVARIANTE DE MATRICES DE DATOS
MULTIVÍA**

*Memoria que, para optar al
grado de Doctor por la
Universidad de Salamanca,
presenta:*

Mariela Alexandra González Narváez

Salamanca

2021



Universidad de Salamanca
Departamento de Estadística

MARÍA JOSÉ FERNÁNDEZ GÓMEZ

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

Y

SUSANA LUÍSA DA CUSTÓDIA MACHADO MENDES

Profesor Adjunto del Instituto Politécnico de Leiria (Portugal)

CERTIFICAN:

Que Dña. Mariela Alexandra González Narváez, Ingeniera en Estadística e Informática, ha realizado en el Departamento de Estadística de la Universidad de Salamanca, bajo su dirección, el trabajo que para optar el grado de Doctor, presenta con el título: *Contribuciones al análisis multivariante de matrices de datos multivía*; y para que conste, firman el presente certificado en Salamanca, a Septiembre 2021.

Fdo.: M^a José Fernández Gómez. Fdo.: Susana Luísa da Custódia Machado Mendes

AGRADECIMIENTOS

Un infinito agradecimiento y reconocimiento a Dios, quien fue un pilar fundamental para poder lograr esta meta.

Mi más sincero agradecimiento a mis directoras las doctoras María José Fernández Gómez y Susana Mendes, por la confianza que depositaron en mí y por los invaluable conocimientos y consejos que me brindaron durante el transcurso de este programa de doctorado. ¡¡¡Gracias!!!

Un agradecimiento en especial a Puri, por creer en mí y por ser la promotora de esta hermosa y gratificante travesía alcanzada.

No puede faltar mi agradecimiento a Ana y a todos los compañeros del doctorado, con quienes nos reuníamos cada día en el Departamento de Estadística. Gracias a cada uno de ustedes por la amistad, el apoyo y el aprendizaje compartido.

A Omar y a mis hijos Melany y Andrés por el apoyo y la comprensión que me brindaron. A mis padres y hermanos por toda la ayuda que recibí de parte de ustedes.

Gracias

DEDICATORIA

A Dios,

A Omar, Melany y Andrés,

A mis padres y hermanos.

INDICE DEL CONTENIDO

GLOSARIO DE TÉRMINOS:	2
INTRODUCCIÓN	6
Métodos estadísticos multivariantes.	6
La aplicación de la Estadística en la Ecología	8
Los datos ecológicos.	8
Errores comunes al analizar datos ecológicos.	11
Estructura del Trabajo	14
Objetivo general de la investigación	15
Objetivo General	15
Objetivos Específicos	15
PARTE I	18
CAPÍTULO 1: INICIOS DEL ANÁLISIS MULTIVÍA	22
1.1 Matrices de datos de dos vías.	22
1.1.1 Análisis de Componentes Principales (ACP)	22
1.1.2 Biplot	23
1.1.3 Análisis de Correspondencias (AC)	25
1.2 Relación entre dos matrices de datos de dos vías.	26
1.2.1 Análisis Canónico de Correspondencias (CCA)	26
1.2.2. Análisis de la CO-Inercia (Co-IA)	27
CAPÍTULO 2: MÉTODOS MULTIVÍA STATIS	32
2.1 Clasificación de la familia STATIS para el análisis de matrices de datos de tres vías.	32
2.2 Matrices de datos de tres vías.	35
2.2.1 Structuration des Tableaux à Trois Indices de la Statistique (STATIS)	35
2.2.2. Análisis Triádico Parcial (PTA)	37
2.2.3 DISTATIS	38
2.3 Relación entre dos conjuntos de datos de tres vías.	39
2.3.1 STATICO	39
2.3.2 COSTATIS	41
PARTE II	43
CAPÍTULO 3: CUANTIFICAR LA MAGNITUD DE LA RELACIÓN ENTRE DOS MATRICES DE TRES VÍAS: MÉTODO HJ-STATICO	47
3.1 Descripción del método HJ-STATICO.	47
3.1.1 Ventajas que ofrece el HJ-STATICO.	63
3.2 Aplicación del HJ-STATICO a datos reales.	63
3.2.1 Importancia de la zona de estudio.	64
3.2.2 Los datos.	65
3.2.3 Análisis realizados con la estructura de tres vías.	67
3.2.4 Aplicación del HJ-STATICO.	69
CAPÍTULO 4: MÉTODO PARA ANALIZAR LAS RELACIONES DE DOS CONJUNTOS DE MATRICES DE DATOS MIXTOS: MixSTATICO	88
4.1 Descripción del método MixSTATICO.	89
4.2 Aplicación del MixSTATICO a datos ilustrativos.	98
4.3 Aplicación del MixSTATICO a datos reales.	107
CAPÍTULO 5: HERRAMIENTAS INFORMÁTICAS DESARROLLADAS EN EL SOFTWARE R.	114
5.1 Librería HJSTATICO utilizando el software R.	114
5.1.1 Algoritmo HJSTATICO	119
5.1.2 Código fuente HJSTATICO	120
5.2 Librería MixSTATICO utilizando el software R.	137

5.2.1 Algoritmo MixSTATICO	137
5.2.2 Código fuente MixSTATICO	139
PARTE III	153
RESUMEN Y CONCLUSIONES EN INGLÉS	157
SUMMARY	157
CONCLUSIONS	162
CONTRIBUCIONES CIENTÍFICAS	168
Artículo publicado	168
Manuscritos en proceso de ser sometidos	193
Presentaciones en eventos científicos internacionales	194
REFERENCIAS BIBLIOGRÁFICAS	198
ANEXOS	213

INDICE DE TABLAS

<i>Tabla 3.1 Lista de las 23 especies de fitoplancton.</i>	67
<i>Tabla 3.2 Correlación Vectorial (RV).</i>	71
<i>Tabla 3.3 Porcentaje de Varianza explicada. (1) "Inter-estructura" y (2) "Compromiso"</i>	71
<i>Tabla 3.4 Índices de las K-matrices.</i>	72
<i>Tabla 4.1. Datos para ejemplo: Variables ambientales y especies en diferentes momentos y lugares.</i>	99
<i>Tabla 4.2. Preprocesamiento de matrices: Datos ambientales y especies son transformados para el análisis. Consultar la tabla 3.1 para los códigos de las especies.</i>	100
<i>Tabla 4.3. Matrices de distancias de Gower entre variables ambientales y especies. Consultar la tabla 3.1 para los códigos de las especies.</i>	101
<i>Table 4.4. Matriz de varianzas y covarianzas vectoriales.</i>	101
<i>Table 4.5. Matriz RV.</i>	101
<i>Tabla 4.6. Matriz compromiso de las relaciones especies-medioambiente. Consultar la tabla 3.1 para los códigos de las especies.</i>	103
<i>Tabla 5.1 Ventajas de la herramienta HJSTATICO en relación con el STATICO.</i>	118

INDICE DE FIGURAS

<i>Figura 3.1 A) Esquema de la estructura de dos conjuntos de matrices $X_{[NPK]}$ e $Y_{[NQK]}$ y B) ordenación de las matrices A) para el análisis.</i>	48
<i>Figura 3.2. Opciones de normalizar una matriz de tres vías: A) representación de fibras y B) cortes. Tomado del libro de (Kroonenberg, 2008, p.29).</i>	50
<i>Figura 3.3 Esquema de los pasos para HJ-STATICO.</i>	62
<i>Figura 3.4 Mapa del perfil costero de Ecuador Continental con la localización geográfica de las estaciones de muestreo: 1. Esmeraldas (0.92° N, 80.11° W), 2. Manta (0.88° S, 80.83° W), 3. La Libertad (2.08° S, 81.12° W), 4. Puerto Bolívar (Pto. Bolívar, 3.11° S, 80.49° W) y la Región Niño 1+2: (0° - 10° S y 90° W - 80° W).</i>	64
<i>Figura 3.5. Estructura de los conjuntos de matrices de partida para ambiente y especies.</i>	66
<i>Figura 3.6 Estructura de las matrices de datos de tres vías para la aplicación del método HJ-STATICO, N=11 individuos, P=7 variables ambientales y Q=23 especies de fitoplancton.</i>	70
<i>Figura 3.8 Mapa bi-factorial del Compromiso A) variables ambientales y B) variables especies. Consultar la Tabla 3.1 para los códigos de las especies.</i>	74
<i>Figura 3.9 Mapa bi-factorial de la Intra-estructura para Esmeraldas (A- ambiente y B - especies), Manta (C - ambiente y D - especies), La Libertad (E - ambiente y F - especies) y Pto. Bolívar (G - ambiente y H - especies). Consultar la Tabla 3.1 para los códigos de las especies.</i>	78
<i>Figura 3.10 Mapa bi-factorial del Compromiso aplicando HJ-Biplot.</i>	80
<i>Figura 3.11 Mapa bi-factorial de la Intra-estructura aplicando HJ-Biplot para la estación de A) Esmeraldas, B) Manta, C) La Libertad y D) Pto. Bolívar. Consultar la Tabla 3.1 para los códigos de las especies.</i>	84
<i>En la figura siguiente se representan mediante un esquema, los pasos a seguir para realizar el método MixSTATICO (Fig. 4.1):</i>	96
<i>Figura 4.1. Esquema del método MixSTATICO.</i>	97
<i>Figura 4.2. Esquema de la estructura de los datos.</i>	98
<i>Figura 4.3. (A) Gráfico de la Inter-estructura y (B) valores propios (eigenvalues).</i>	102
<i>Figura 4.4. Compromiso: plano bi-factorial para (A) ambiente y (B) especies. (C) Valores propios. Consultar la Tabla 3.1 para los códigos de las especies.</i>	104
<i>Figura 4.5. Intra-estructura: Proyecciones sobre el plano bi-factorial del compromiso (A) ambiente y (B) especies correspondientes a cada k-matriz (tiempo). Consultar la Tabla 3.1 para los códigos de las especies.</i>	106
<i>Figura 4.6. Esquema de la estructura de los datos. Variables ambientales (P = 7 cuantitativas), variables especies (Q = 10, que están clasificadas en: 7 cuantitativas, 2 ordinales y 1 nominal)</i>	107
<i>Figura 4.7. (A) Gráfico de la Inter-estructura (época seca: color rojo; época lluviosa: color azul). (B) Valores propios y (C) pesos (α) y Cos^2 para cada K-matriz.</i>	108

Figura 4.8. Compromiso: plano bi-factorial para (A) ambiente y (B) especies. (C) Valores propios. Consultar la Tabla 3.1 para los códigos de las especies. _____ 109

GLOSARIO DE TÉRMINOS:

Matrices de tres vías: tablas de tres vías o cubos de datos.

Condición: Tercera vía o dimensión de una matriz de tres vías.

Diag(Θ): Calcula la diagonal de la matriz Θ .

HJ-STATICO: método de tres vías propuesto en esta tesis para la realización de un análisis STATICO incorporando las métricas de contribuciones relativas y calidad de representación del HJ-Biplot.

HJSTATICO: librería informática que permite realizar un análisis STATICO desde una perspectiva del HJ-Biplot.

K-matrices: sub-tablas o sub-matrices que conforman una matriz de tres vías.

Matrices cruzadas: el resultado de aplicar el método de CO-Inercia (Co-IA) a dos pares de matrices $\mathbf{X}_{[NP]}$ y $\mathbf{Y}_{[NQ]}$ que tienen los mismos N individuos.

MixSTATICO: método propuesto en esta tesis para el análisis de la parte estable de la co-estructura entre dos matrices de tres vías con variables cuantitativas y cualitativas (mixtas).

MixSTATICO: librería informática que permite realizar un análisis MixSTATICO con datos mixtos.

NO_3 : concentraciones de Nitrato.

NO_2 : concentraciones de Nitrito.

OD : concentraciones de Oxígeno Disuelto en el mar.

PO_4 : concentraciones de Fosfato.

S : concentraciones de salinidad.

SiO_4 : concentraciones de Silicato.

T : valores para la Temperatura del mar.

$Vec(\mathbf{X}_{[NPK]})$: Vectorizar una matriz

$\mathbf{X}_{[NP]}$: Matriz de dos vías con N individuos y P variables.

$\mathbf{X}_{[NP]}^T$: Transpuesta de una matriz de dos vías con N individuos y P variables.

$\mathbb{X}_{[NPK]}$: Matriz de tres vías con N individuos, P variables y K condición.

$\mathbf{X}_{[NP]}^{(k)}$: k-ésima matriz de dos vías contenida en un arreglo matricial total de dimensión $(NK \times P)$, resultado de yuxtaponer las k-matrices iniciales.

$\vec{\alpha}_{[K]}$: Vector de dimensión $(K \times 1)$.

*Hay dos maneras de vivir una vida:
la primera es pensar que nada es un milagro.
la segunda es pensar que todo es un milagro.
de lo que estoy seguro es que Dios existe.*

Albert Einstein

INTRODUCCIÓN

El desarrollo de una investigación promueve la generación de valores que expresan las características del fenómeno o evento de interés en el estudio. Estos valores también conocidos como datos son el resultado de la medición de diversos factores en un conjunto de unidades de observación, que posteriormente son organizados en una estructura matricial de filas (observaciones) y columnas (variables) que facilita el proceso del análisis estadístico. Dentro de este tipo de análisis prevalece el análisis multivariante, que puede ser definido como: “El análisis multivariante es parte de la estadística y del análisis de datos que estudia, analiza, representa e interpreta los datos que resultan de observar más de una variable estadística sobre una muestra de individuos” (Cuadras, 2019, p. 13).

Métodos estadísticos multivariantes.

A inicios del siglo pasado aparecen los métodos de ordenación y reducción de la dimensión a una matriz de dos dimensiones (filas y columnas). En 1901 Karl Pearson presentó el método multivariante conocido como Análisis de Componentes Principales (ACP), el cual ha sido aplicado hasta la actualidad en innumerables investigaciones de los diversos campos de la ciencia.

Con este aporte metodológico realizado por Pearson (1901), se dio inicio a un importante avance en el análisis multivariante y al desarrollo de nuevas técnicas orientadas a la proyección de los datos en un espacio de baja dimensión, tales como: Análisis Factorial (AF; Spearman 1904; Thurstone 1947), Métodos Biplot (Gabriel, 1971; Galindo-Villardón, 1986; Galindo-Villardón & Cuadras, 1986), Análisis de Correspondencias (AC; Benzécri, 1973; Hill, 1973), Análisis de Correspondencias Múltiples (ACM; Tenenhaus & Young, 1985) y el Análisis Canónico de Correspondencias (CCA; ter Braak, 1986), entre otros.

Por otra parte, el enfoque del análisis de los datos ha ido evolucionando, debido al interés que tienen los investigadores de realizar estudios que consideren otras condiciones que inciden en los valores que puedan tomar las variables. Fue así, como se empezó a estructurar el conjunto de datos, considerando tres vías o dimensiones, formando los llamados “cubos de datos”, es decir, se estudia el comportamiento de los individuos (dimensión 1) en base a ciertas características medidas a través de las variables (dimensión 2), observadas en diferentes situaciones y/o su evolución en el tiempo (dimensión 3).

Fue [Tucker \(1966\)](#) quien dio origen al análisis conjunto de una serie de matrices de datos o “cubo de datos” con el análisis factorial de tres-modos Tucker 2 y 3; continuando posteriormente con importantes aportaciones de Richard Harshman con el modelo de Análisis del Factor Paralelo (PARAFAC; [Harshman, 1970](#)) y de Douglas Carroll con el método de Descomposición Canónica (CANDECOMP; [Carroll & Chang, 1970](#)). Otro método desarrollado para analizar matrices de tres vías es el Análisis Factorial Múltiple (AFM; [Escofier & Pagès, 1984, 1990, 1994](#)).

Para dar respuesta a estas necesidades se han desarrollado métodos estadísticos multivariantes basados en un ACP que permiten estudiar la parte estable de las matrices de tres vías, tales como los métodos de la familia SATIS (Structuration des Tableaux A Trois Indices de la Statistique, [L’Hermier Des Plantes, 1976](#)) que es una evolución del método propuesto por Tucker, mientras que para estudiar la parte estable de la relación de dos cubos de datos se tiene el método STATICO ([Simier et al., 1999](#)) y el COSTATIS ([Thioulouse, 2011](#)).

Entre los principales campos de estudio que más han incursionado en el análisis de los datos con tres o más vías se encuentran la Psicología, la Biología, la Ecología y el Medioambiente. Las investigaciones en estas áreas se caracterizan por generar conjuntos de datos que son el resultado de mediciones de múltiples variables en diferentes condiciones como lugares (espacio) y/o en diferentes

momentos (tiempo); es decir, los datos pueden ser organizados en varias matrices que representan las mediciones realizadas en diferentes condiciones o momentos.

Estos métodos multivía constituyen la base metodológica de esta tesis doctoral, que se orienta en realizar contribuciones a dichos métodos, con la finalidad de aportar nuevas herramientas al análisis de datos en general, y al estudio del área ecológica en particular.

La aplicación de la Estadística en la Ecología

En la Ecología, así como en otras áreas afines se generan datos complejos que requieren de métodos estadísticos apropiados para analizar las características naturales de los datos. La aplicación de las técnicas estadísticas adecuadas, exigen, en muchas ocasiones, un amplio conocimiento y experticia en el comportamiento de los datos ecológicos

Los datos ecológicos.

Son numerosas las investigaciones que se han venido desarrollando hasta la actualidad con respecto al estudio de la Ecología y de otras áreas afines en las que la Estadística ha tenido y tiene un papel importante para extraer de los datos información relevante que le permita a los investigadores conocer el comportamiento de los parámetros bióticos y abióticos y la relación entre ellos.

Al momento de analizar los datos ecológicos, la metodología que se va a utilizar debe estar relacionada con las condiciones propias de las observaciones y con los resultados esperados. Se pueden obtener diversos resultados, debido a que están asociados a la exploración de un conjunto particular de condiciones. Esto se debe a la gran cantidad de variables que están presentes en las bases de datos y que generalmente no todas están disponibles para el investigador. Además, se debe

considerar que la Ecología tiene la particularidad de que posee una importante fuente de variabilidad, que es la variabilidad genética y que está contenida en el material biológico (Legendre & Legendre, 1998).

Para analizar los datos ecológicos, se utilizan diversas herramientas, entre otras, herramientas computacionales y matemáticas, que han sido previamente desarrolladas en otras disciplinas. Por tanto, la mayoría de los ecólogos están familiarizados con el lenguaje matricial. Además, existen otras disciplinas que proporcionan a estos investigadores poderosas herramientas que se adaptan a la complejidad de los datos que manejan, por ejemplo de la física matemática, se adopta el análisis dimensional (Legendre & Legendre, 1998).

Así lo muestran Legendre & Legendre en (1998, p. 4) donde presentan un esquema de la aportación de las técnicas matemáticas (álgebra matricial y la física matemática) junto con la incorporación de las técnicas estadísticas (paramétricas, no paramétricas y la teoría de la información) para el análisis de los datos ecológicos complejos, y dependiendo del propósito del estudio que se tenga, clasifican a las técnicas en dos grupos: i) para el estudio de las **estructuras ecológicas** utilizando los métodos de clustering, de ordenación e interpretación de las estructuras ecológicas y ii) para el estudio de las **estructuras espacio-temporales** utilizando series temporales y el estudio de los datos espaciales.

Para facilitar el análisis numérico de los datos ecológicos, se recomienda que estén organizados en una base de datos adoptando la forma de una matriz, colocando en las filas lo que corresponde a los sitios o tiempos de muestreo, mientras que en las columnas se colocan a las variables que son las que describen las comunidades biológicas (como por ejemplo la presencia de especies, la abundancia o la biomasa, entre otros) o el entorno físico (medio ambiente). En el área de la Ecología, se necesitan, en general, numerosas variables para poder describir las comunidades y el medio ambiente, por tal motivo se asocia a los conjuntos de datos ecológicos con el término multidimensional o multivariante.

Estos datos multidimensionales pueden ser representados gráficamente en un hiperespacio con muchas dimensiones. Como por ejemplo el hiperespacio ecológico del nicho fundamental de [Hutchinson \(1965, 1957\)](#), el cual menciona que las variables ambientales que son fundamentales para la existencia de una especie, se pueden considerar como ejes ortogonales de un espacio multidimensional. En cada eje, existen condiciones limitantes y este conjunto de condiciones definen un hipervolumen denominado nicho fundamental de la especie.

Existen diversos trabajos que se han realizado en el área de la Ecología aplicando diferentes técnicas matemáticas y estadísticas para el análisis de estos datos. Los análisis han sido desarrollados utilizando desde técnicas descriptivas, como los realizados por: [Jiménez & Bonilla \(1980\)](#); [Gualancañay et al. \(2004\)](#); [Tapia & Naranjo \(2004\)](#); [Torres, Calderón et al. \(2004\)](#); [Torres, Mero, et al. \(2004\)](#); [Torres \(2005\)](#); [Cruz et al. \(2011\)](#) y [Jiménez \(2014\)](#), hasta las multivariantes (convencionales) que permiten tener una visión integral de la relación entre las variables, como por ejemplo los trabajos de: [ter Braak \(1986\)](#); [Vega et al. \(1998\)](#); [Yan et al. \(2000\)](#); [Zalocar De Domitrovic \(2002\)](#); [Cassis et al. \(2002\)](#); [Martínez et al. \(2003\)](#); [Diaz et al. \(2004\)](#); [De la Hoz Aristizábal \(2004\)](#); [Shrestha & Kazama \(2007\)](#); [Mendes et al. \(2009\)](#); [Noriega et al. \(2009\)](#); [Yi et al. \(2011\)](#); [Perugachi-Salamea et al. \(2014\)](#); [Torres et al. \(2017\)](#); [Ravindra et al. \(2008\)](#); [Borbor-Cordova et al. \(2019\)](#); [Kendal et al. \(2019\)](#); [Atique et al. \(2020\)](#), entre otros.

Sin embargo, las técnicas estadísticas multivariantes convencionales han demostrado ser herramientas óptimas para identificar patrones ambientales y biológicos, pero de manera independiente. [Carassou & Ponton \(2007\)](#) mencionaron que el principal inconveniente de estos métodos es que no estudian las variaciones espaciales o temporales de los factores ambientales y biológicos o la relación entre ellos.

Es de gran interés para los ecólogos predecir presencia de especies, determinar rasgos fisiológicos y/o morfológicos para predecir el rendimiento y/o la distribución geográfica de los organismos, identificar diferencias en la composición de las especies que podrían estar asociadas con gradientes ambientales, entre otras (Scott & Crone, 2021), así como también, estudiar el comportamiento, variabilidad, abundancia y composición de las especies en relación a factores como el espacio y el tiempo (Haury et al., 1978; Jiménez & Bonilla, 1980; Tapia & Naranjo, 2004).

Errores comunes al analizar datos ecológicos.

El rol de la Estadística en los análisis de datos ecológicos tiene un alto impacto e importancia entre los investigadores de esta área de la ciencia. Es así, que Fowler (1990) publicó en el boletín de la *Ecological Society of America* un artículo en el que mencionaba los 10 errores estadísticos más comunes que ocurren al momento de analizar los datos.

Sin embargo, tras esos 20 años, los errores citados por Fowler siguen aconteciendo en la actualidad, tal como lo mencionan Steel et al. (2013). Los autores se encargaron de hacer una comparación y una revisión actualizada de los principales desaciertos en los procedimientos que se cometen. A continuación, se enumeran los principales errores revelados:

Al momento de establecer un análisis.

1. No explorar los datos.
2. Fijar umbrales, métricas e indicadores de forma arbitraria.
3. Asumir que las observaciones son independientes.
4. Las poblaciones no se corresponden con el marco muestral (Muestras no representativas).

Al momento de realizar un diseño experimental.

5. El control difiere de los tratamientos antes de que ocurra el tratamiento.

6. Estrategias de medición que confunden los diseños experimentales.
7. No modelar las covariables en el nivel correcto (principalmente si ocurren interacciones).

Al momento de realizar una aplicación estadística.

8. Transformaciones innecesarias de los datos.
9. No tratar adecuadamente los ceros.
10. Desconocer estructuras de correlaciones subyacentes.
11. Cuando los residuales o modelos de diagnósticos no cumplen con los supuestos.
12. Realizar demasiadas pruebas estadísticas.
13. Aplicar una nueva técnica desconociendo su uso y los supuestos.

Al momento de interpretar modelos y pruebas estadísticas.

14. Incorrecta interpretación de un valor p no-significativo.
15. Comparaciones inapropiadas de valores p.
16. Indicar una importancia ecológica basándose en la significación estadística para muestras de tamaño muy grandes.
17. Incorrecta interpretación de los coeficientes en modelos de regresión múltiple.
18. Extrapolación.

De todos los errores mencionados, en este trabajo de investigación nos centraremos en considerar los errores del punto 3 y 10, por estar asociados con la metodología multivariante multivía que propondremos en el **Capítulo 3**.

Para el error detallado en el punto 3 (“asumir que las observaciones son independientes”), [Steel et al. \(2013\)](#) mencionan que hay diversas formas en las que dos o más observaciones pueden ser dependientes y que muchas de las veces no son reconocidas. Estas dependencias se deben a las relaciones del tipo espacial, temporal o filogenéticas ([Magnusson, 1997](#)). En Ecología las dos formas más comunes de dependencia son: por las medidas repetidas ([Gurevitch & Chester, 1986](#)) y por la pseudo-replicación ([Hurlbert, 1984](#)).

Mientras que, para el error detallado en el punto 10 (“desconocer estructuras de correlaciones subyacentes”), los autores indicaron que los datos ecológicos a menudo presentan estructuras de correlación, de desbalance o de dependencia y que deben ser analizados con modelos (métodos) que consideren dentro de sus análisis estas características. Hay ocasiones que estos datos complejos pueden ser analizados de forma significativa mediante análisis exploratorios y jerárquicos, pero se comete un error cuando se ignora la correlación o se analizan sub-muestras de los datos.

Los autores [Steel et al. \(2013\)](#) coinciden en el hecho de que mientras los errores en los conceptos básicos del pensamiento y del análisis estadístico sigan presentes en la formación ecológica, existe la necesidad de incorporar estadísticas de alta calidad en la educación e investigaciones ecológicas. La inferencia estadística es un componente integral de la investigación ecológica, y ha tenido una evolución rápida debido tanto a los aumentos masivos en la potencia de cálculo como a la disponibilidad de datos.

Estructura del Trabajo

El presente documento se encuentra estructurado de la siguiente forma:

PARTE I: METODOLOGÍA ESTADÍSTICA.

- Capítulo 1. Inicios del análisis multivía.
- Capítulo 2. Métodos multivía STATIS.

PARTE II: RESULTADOS OBTENIDOS.

- Capítulo 3. El HJ-STATICO para el análisis de co-estructuras a través de índices.
- Capítulo 4. MixSTATICO como una generalización del método STATICO.
- Capítulo 5. Código desarrollado en R para la aplicación de las dos metodologías propuestas en los capítulos 3 y 4.

PARTE III: RESUMEN Y CONCLUSIONES EN INGLÉS

Resumen

Conclusiones

Contribuciones Científicas

Referencias Bibliográficas

Anexos

Objetivo general de la investigación

En la presente investigación, se estudian los cambios en la abundancia de las especies asociados a las variaciones ambientales, considerando el dominio de la localización espacial, optando por aplicar métodos estadísticos multivariantes de tres vías que permitan incorporar en un solo análisis el efecto de los factores espacios-temporales sobre la relación especie-ambiente.

Como resultado de las aplicaciones metodológicas realizadas, surgieron necesidades tales como: i) cuantificar la cantidad de información expresada por las variables al momento de explicar la relación entre dos matrices de tres vías como, por ejemplo: el medio ambiente y las especies y ii) realizar análisis en los que se estudien de forma conjunta datos cuantitativos y cualitativos.

Objetivo General

Dados los antecedentes antes mencionados, esta investigación tiene como objetivo general:

Aportar alternativas al análisis de matrices de datos de tres vías desde una perspectiva multivariante, desarrollando metodologías innovadoras y utilizando diferentes métricas, que cuantifiquen y faciliten el análisis de las relaciones entre pares de matrices de tres vías que tengan características particulares.

Objetivos Específicos

El primero de ellos consiste en:

- contribuir al análisis de la parte estable de la estructura en común de dos matrices de datos de tres vías con la propuesta de un nuevo método, que

combine en un solo análisis las ventajas del método STATICO (Simier et al., 1999; Thioulouse et al., 2004) y del HJ-Biplot (Galindo-Villardón, 1986; Galindo-Villardón & Cuadras, 1986), permitiendo el cálculo de una serie de índices (contribuciones y calidades de representación) para que el investigador interprete de manera objetiva los resultados del análisis, además de una representación gráfica en el mismo sistema de referencia y con máxima calidad de representación tanto las filas como las columnas.

- Desarrollar un código fuente de acceso libre, utilizando el programa informático de código abierto RStudio (R Core Team and contributors worldwide, 2018), para la obtención de resultados del nuevo método propuesto (STATICO con HJ-Biplot), que facilite la interpretación de los resultados, presentando gráficos e índices que provean información sobre la relación entre pares de matrices de tres vías.
- Proponer una nueva metodología, como una extensión del análisis STATICO, analizando la parte estable de la co-estructura de dos matrices de datos de tres vías que están conformadas por una mezcla de variables medidas en diferentes escalas, compuestas por variables cuantitativas y cualitativas a la vez, y representándolas de la misma manera que ofrece el STATICO.

PARTE I

CAPÍTULO 1

*Cuando mi
mente se llenó
de dudas, tu
consuelo renovó
mi esperanza y
mi alegría.
Salmos 94:19*

CAPÍTULO 1: INICIOS DEL ANÁLISIS MULTIVÍA

En este apartado, se presenta el origen de los métodos estadísticos disponibles para el análisis de las matrices de datos multivariantes. Se inicia con una revisión de las técnicas que aparecen en la literatura para estudiar matrices de dos vías (dimensiones), para luego continuar en el siguiente capítulo con las matrices de tres vías. Se presentarán las ventajas y desventajas de cada método, así como también, diversas aplicaciones que se han realizado en el ámbito ecológico que es el contexto en el que se desarrolla la presente investigación.

1.1 Matrices de datos de dos vías.

Cuando se lleva a cabo el análisis multivariante a una matriz de datos \mathbf{X} estructurada por filas y columnas, se concibe la idea de la reducción de la dimensión de la matriz, con la finalidad de expresar la misma información en un espacio de dimensión reducida $R^P \rightarrow R^D, D < P$.

Para un mejor entendimiento, se inicia definiendo \mathbf{X} , tal que

$$\{\mathbf{X}_{[NP]} = x_{ij} | i = 1, \dots, N; j = 1, \dots, P\} \quad (1.1)$$

que contiene N individuos en sus filas y P variables en sus columnas.

1.1.1 Análisis de Componentes Principales (ACP).- Es el principal método de ordenación para una matriz de múltiples variables; fue desarrollado por [Pearson \(1901\)](#) con la finalidad de encontrar una matriz de menor dimensión que mejor represente la información de los datos originales basándose en el método de los mínimos cuadrados. Posteriormente [Hotelling \(1933, 1936\)](#) realizó una aportación

al obtener combinaciones lineales sucesivas (componentes principales) a partir de las variables originales.

Esta técnica, consiste en proyectar la matriz de datos original X en un espacio de menor dimensión (generalmente bidimensional) construido a partir de las nuevas componentes, que son combinaciones lineales de las variables de partida, que tienen la característica de ser ortogonales y no correlacionadas entre sí. La primera componente representa la mayor varianza retenida, la segunda componente contiene la siguiente mayor varianza posible y así sucesivamente para el resto.

Existen diversas contribuciones científicas de investigadores que han utilizado este método en el campo de la Ecología. A continuación, se citan algunas de estas investigaciones: evaluación de los efectos estacionales y contaminantes sobre la calidad del agua de río (Vega et al., 1998); análisis de la calidad del agua superficial en la cuenca del río Fuji, Japón (Shrestha & Kazama, 2007); evaluación del riesgo ecológico de los metales pesados en los sedimentos de la cuenca del río Yangtze (Yi et al., 2011); análisis de los impactos de vertederos en la química del agua fluvial y en los conjuntos de peces en el río Geum, Korea del Sur (Atique et al., 2020); entre otros.

Este método descriptivo, permite realizar una fácil interpretación de la ordenación de los datos en un espacio bi-factorial. Como restricción requiere que el número de las observaciones sea igual o mayor al número de variables.

1.1.2 Biplot.- Los métodos Biplot desarrollados por Gabriel (1971) consisten en representar de forma gráfica los elementos de una matriz de datos multivariantes. El método aproxima una matriz X de rango superior a una matriz de rango dos. Utiliza el método de Descomposición de Valores Singulares (DVS) y obtiene matrices de marcadores para filas y columnas que posteriormente son representadas gráficamente en un espacio bi-factorial.

[Gabriel \(1971\)](#) presenta dos alternativas de análisis: el método GH-Biplot, que se caracteriza por representar con una mayor calidad de representación las P variables (columnas) pero con una menor calidad a los N individuos; y el JK-Biplot que permite representar con una mayor calidad de representación a los N individuos (filas) pero con una menor calidad a las P variables de la matriz de partida. Una extensión de estos métodos es la propuesta por [Galindo-Villardón \(1986\)](#), denominada HJ-Biplot, que consiste en proyectar en un mismo sistema de referencia de ejes bi-factoriales los N individuos (filas) y las P variables (columnas) de la matriz de partida, ambos con la máxima calidad de representación.

Existen diversas contribuciones científicas de investigadores que han utilizado estos métodos, por ejemplo: el uso del Biplot desde la perspectiva del Análisis Directo del Gradiente para integrar la información de la distribución de especies y la información ambiental ([ter Braak, 1986](#)), el estudio de la relación entre bacterioplancton y el medioambiente en el Archipiélago de Berlangas ([Mendes, Fernández-Gómez, Galindo-Villardón, et al., 2009](#)), análisis de la interacción entre genotipos y el medioambiente ([Kendal et al., 2019](#)) utilizando una variante del método Biplot, el GGE Biplot ([Yan et al., 2000](#)), entre otros.

La técnica Biplot se puede utilizar para analizar matrices que tienen un gran volumen de datos, y además, tiene la particularidad de que permite representar dichos datos a través de la utilización de diferentes métricas, las cuales ponen de manifiesto diversas características relevantes de los datos y permite cuantificar, mediante indicadores, la representatividad de las proyecciones sobre los ejes del plano bi-factorial. Posee menos restricciones que el Análisis de Correspondencias (AC) y el Análisis Factorial (AF). No incorpora el comportamiento de mediciones repetidas en el tiempo o el espacio y requiere tener una matriz de datos con un número de individuos mayor o igual al número de variables.

1.1.3 Análisis de Correspondencias (AC).- Se caracteriza por estudiar las asociaciones entre las categorías de dos variables cualitativas. Es un método de ordenación que permite representar las filas y las columnas de una tabla de contingencia en un espacio vectorial de dimensión reducida (Benzécri, 1973; Hill, 1973). Su procedimiento se centra en analizar los perfiles de los elementos filas y columnas de la tabla y utiliza la distancia ji-cuadrado para evaluar las distancias entre los perfiles. Permite representar la variabilidad total de los perfiles en el espacio vectorial a través de la inercia.

El AC es la base metodológica para el Análisis de Correspondencias Múltiples (ACM) el cual permite analizar múltiples variables cualitativas.

Existen diversas contribuciones científicas de investigadores que han utilizado este método en el campo de la Ecología y medioambiente, por ejemplo: para el estudio de la distribución de Lycosidae en una área de dunas (ter Braak, 1985), encontrar especies indicadoras y conjuntos de especies que permitan caracterizar zonas (Dufrene & Legendre, 1997), el estudio de conjuntos de macro-invertebrados bentónicos y la conectividad de macrohábitats en arroyos de clima mediterráneo, norte de California (Bonada et al., 2006), comparación de las propiedades del AC para datos no relativizados, con el AC regular y con una variante del AC no simétrico, utilizando datos de abundancia de bentos (Greenacre, 2010), análisis de conjuntos de diatomeas en sedimentos de superficie y su reflexión en la vía de migración de agua desde la bahía de Bohai occidental, China (Li et al., 2020), para definir las estrategias de vida de especies nectónicas del mar Mediterráneo, analizar los rasgos relacionados con el uso de recursos para determinar el nivel de superposición de rasgos y nichos y su relación con las estrategias de vida (Koutsidi et al., 2020), entre otros.

Se caracteriza por analizar y cuantificar, como hemos dicho, dos variables nominales, mostrando resultados que son de fácil interpretación; estos resultados, representados de forma grafica, expresan una buena representación de las

categorías que presentan una fuerte contribución con la dimensión (eje). Este método no evalúa distancias u orden entre las categorías de una misma variable, solo permite analizar asociaciones.

1.2 Relación entre dos matrices de datos de dos vías.

Para un mejor entendimiento de los conceptos en este apartado, se define la matriz \mathbf{X} como en la Ec. (1.1) y a \mathbf{Y} como una matriz de dimensión $(N \times Q)$; tal que,

$$\{\mathbf{Y}_{[NQ]} = y_{ij'} \mid i = 1, \dots, N; j' = 1, \dots, Q\} \quad (1.2)$$

que contiene N individuos en sus filas y Q variables (por ejemplo, especies) en sus columnas.

De esta forma, se tiene un par de matrices de dos vías \mathbf{X} y \mathbf{Y} que contienen los mismos N individuos en sus filas y diferentes variables en sus columnas (P y Q respectivamente), a partir de este par de matrices se desea estudiar la relación existente entre ambas.

1.2.1 Análisis Canónico de Correspondencias (CCA). – El CCA ([ter Braak, 1986](#)) es una técnica derivada del AC que permite analizar la relación entre la composición de las comunidades de especies \mathbf{Y} (variables dependientes) y las variaciones ambientales \mathbf{X} (variables independientes). Calcula los ejes de ordenación del AC de la matriz \mathbf{Y} , de forma que sean combinaciones lineales de las variables de \mathbf{X} consiguiendo establecer, de esta forma, una relación directa entre las variaciones del ambiente \mathbf{X} y la distribución de las especies \mathbf{Y} que se conoce como análisis directo del gradiente.

Son muy numerosas las contribuciones científicas de investigadores que han utilizado este método, por ejemplo: se presentó el método CCA al relacionar variables ambientales y de especies en tres ejemplos diferentes: 1) con araña de

caza, 2) con vegetación y 3) con algas (ter Braak, 1986); análisis de la relación vegetación-ambiente (ter Braak, 1987); para identificar las variables ambientales que gobiernan la composición y estructura de los conjuntos de diatomeas en la Ria de Aveiro (Mendes et al., 2012); para determinar las respuestas de los ensamblajes de fitoplancton a los factores estresantes, además de evaluar el estado limno-ecológico de los lagos en el sistema del río Aras, Turquía (Çelekli et al., 2020); entre otros.

Este método se caracteriza por permitir que las variables independientes o predictoras puedan ser del tipo cuantitativo y nominal. Es una combinación del AC y del método de regresión. Está enfocado en inspeccionar el comportamiento de las variables respuestas en función de las predictoras, por lo que lo convierte en un método asimétrico; además, para un correcto funcionamiento, requiere que el número de individuos sea mayor al número de variables.

1.2.2. Análisis de la CO-Inercia (Co-IA). – El método de la CO-Inercia (Dolédec & Chessel, 1994) analiza de forma simultánea un par de matrices X y Y con la finalidad de encontrar la estructura en común (co-estructura) entre ambas matrices. Para este proceso, se calculan, para cada matriz, los ejes principales que representan la dirección de los vectores que maximizan la variabilidad, sobre estos ejes pueden estar proyectados los individuos. Posteriormente se obtienen nuevos ejes uno resultado de X y otro de Y tal que expliquen la máxima covarianza (correlación) entre los dos nuevos conjuntos de puntuaciones proyectadas. El procedimiento matemático de este método es presentado de forma exhaustiva en el **Capítulo 3**, por formar parte de la base metodológica del presente trabajo.

Existen también numerosas contribuciones científicas de investigadores que han utilizado este método, por ejemplo: para estudiar la relación entre macroinvertebrados y el medioambiente (Dolédec & Chessel, 1994); efectos espaciales y temporales en las relaciones especie-ambiente (Franquet et al., 1995); análisis de la Co-IA para estudiar la relación entre tablas de datos ecológicas, mostraron la

comparación de diversos métodos utilizando datos simulados para crear respuestas de especies a un espacio ambiental bidimensional (Dray et al., 2003); estudio de la concordancia entre los rasgos biológicos y ecológicos de los escarabajos acuáticos en los ecosistemas mediterráneos (Picazo et al., 2012); análisis de la correspondencia entre la variación de 3 rasgos claves de la hoja y 14 características (vegetativas y reproductivas) funcionales (Pierce et al., 2017); entre otros.

El Co-IA es apropiado para analizar variables cuantitativas y/o cualitativas o difusas, se caracteriza por ser un método muy flexible y por permitir diversas posibilidades de acoplamiento, es conocido como un método simétrico ya que está enfocado en encontrar ejes que maximizan la covariación de las variables de ambas matrices de datos; además, tiene la particularidad de que no requiere que el número de individuos sea mayor al número de variables como en el CCA. Es un método que solo funciona con matrices de dos vías.

CAPÍTULO 2

*Él da esfuerzo
al cansado,
y multiplica las fuerzas
al que no tiene ninguna*
Isaías 40:29

CAPÍTULO 2: MÉTODOS MULTIVÍA STATIS

En esta sección, se presenta un compendio de los métodos utilizados para el análisis de matrices de datos multivía. Se revisan varias aportaciones metodológicas desarrolladas dentro de este contexto, las cuales son clasificadas como pertenecientes a la escuela francesa, a la escuela anglosajona y a la escuela salmantina. Para propósitos de este trabajo de investigación, nos centraremos en los métodos desarrollados por la escuela francesa.

2.1 Clasificación de la familia STATIS para el análisis de matrices de datos de tres vías.

A continuación, se hace referencia a los métodos de la familia STATIS, que han tenido una gran acogida entre los investigadores de muy diversos campos para realizar análisis de matrices de tres vías.

Los métodos de la familia STATIS pueden ser clasificados de la siguiente manera:

- 1) *Según los datos de partida*: Cuando la matriz de tres vías contiene información de los mismos individuos sobre los que son medidas diferentes variables, los datos se analizan mediante el análisis STATIS. Este método se basa en trabajar con matrices de operadores (matrices de varianzas-covarianzas). Por el contrario, si se disponen de las mismas variables medidas en diferentes individuos se utiliza el STATIS-DUAL (L'Hermier Des Plantes, 1976), pero si las variables e individuos son los mismos, los datos se pueden analizar con la técnica X-STATIS o Análisis Triádico Parcial (Jaffrenou, 1978), que trabaja con las matrices originales y no con

operadores. [Gaertner et al. \(1998\)](#) propusieron una variante del STATIS combinándolo con un Análisis de Correspondencias (AC) llamado STATIS-CoA. Si la matriz de datos de tres vías inicial contiene operadores, como por ejemplo las covarianzas calculadas en los mismos individuos, se aplica COVSTATIS ([Thioulouse, 2011](#)). Si las matrices de datos no representan una función lineal, se aplica el método Kernel-STATIS ([Marcondes Filho et al., 2011](#)). Cuando se tienen los mismos individuos y el propósito es clasificarlos, entonces se utiliza STATIS-LDA ([Sabatier et al., 2013](#)). En cambio si las matrices representan distancias evaluadas en los mismos individuos se aplica el DISTATIS ([Abdi et al., 2007](#)). [Kriegsman \(2018\)](#) propuso el HiDiSTATIS y DiDISTATIS, métodos que están basados tanto en el método DISTATIS como en el análisis discriminante. Para datos de intervalo se tiene el INTERSTATIS ([Rodríguez & Corrales, 2014](#)). Si se desea analizar datos sensoriales, se dispone del método CATATIS ([Llobell et al., 2019](#)) y finalmente para crear agrupaciones a partir de una matriz de tres vías, se aplica CLUSTATIS ([Llobell et al., 2020](#)) que utiliza el método CLUSTER.

- 2) *Conforme con los pesos asignados al crear la matriz compromiso*: cuando los pesos asignados a cada matriz difieren de la propuesta original de L'Hermier, se tiene el POWER-STATIS ([Bénasséni & Bennani Dosse, 2012](#)); cuando difieren los pesos asignados a cada variable en la matriz de tres vías, se tiene el ANISOSTATIS ([Abdi et al., 2012](#)); cuando se incorporan las restricciones Elastic net a la matriz compromiso, se tiene Sparse STATIS-DUAL ([Rodríguez-Martínez, 2020](#)).
- 3) *Cuando se incorpora información externa*: El método $K + 1$ STATIS ([Sauzay et al., 2006](#)) considera una matriz de tres vías analizada con STATIS y una matriz adicional que contiene información de los mismos individuos; posteriormente a partir de esta técnica se deriva el STATIS-4 desarrollado por [Sabatier & Vivien \(2008\)](#) que consiste en una tablas de 4

vías. El CANOSTATIS se utiliza en el caso que los individuos se estructuren en grupos (Vallejo-Arboleda et al., 2007). Si se desea analizar una matriz de tres vías con respecto a una variable numérica Y de interés, se aplica el método Sir-STATIS propuesto por Sautron et al. (2016).

- 4) *En el caso de tener dos cubos de datos, y los pares de matrices tienen diferentes condiciones o momentos:* Cuando los pares de matrices tienen la misma condición o momentos se aplica el STATICO (Simier et al., 1999) o el COSTATIS (Thioulouse, 2011) que son técnicas que permiten estudiar la estructura en común entre dos pares de cubos de datos; en cambio, si se combinan los métodos STATICO y AC se tiene el STATICO-CoA (Mérigot et al., 2019). Si las condiciones o los momentos de las dos matrices de tres vías son diferentes se tiene el DOUBLE STATIS (DO-ACT) (Vivien & Sabatier, 2004), que calcula dos representaciones óptimas de las observaciones (consenso).

En resumen, estos métodos permiten analizar la estructura promedio de una matriz de tres vías (métodos STATIS) o la imagen promedio de la estructura en común entre dos matrices pareadas de tres vías (SATICO y COSTATIS) a lo largo del tiempo o en diferentes condiciones. Partiendo del análisis de matrices de datos con un solo tipo de variables que pueden ser cuantitativas o cualitativas (utiliza ACP o AC en el preprocesamiento inicial dependiendo de los datos). Ofrecen resultados como una representación gráfica de las proyecciones de los elementos de la matriz sobre un espacio bi-factorial, quedando al mejor criterio del investigador la interpretación del grado de relación y/o aportación de las variables y observaciones en el análisis.

A continuación, se presenta una descripción de los métodos STATIS que forman parte de la base metodológica de las nuevas técnicas desarrolladas en este trabajo de investigación. Además, de las ventajas y desventajas de cada método, así como también diversas aplicaciones que se han realizado en el ámbito ecológico.

2.2 Matrices de datos de tres vías.

Estas matrices poseen la característica de estar formadas por un conjunto de matrices de dos vías que tienen información en común. Su estructura está determinada por los N individuos (filas), P variables (columnas) y K el factor tiempo o lugares (condición). Al estar formado por estas tres vías se le da el nombre de cubo de datos.

Para este apartado se define a \mathbb{X} (ver Ec. (2.1)) como una matriz de datos multivariantes de tres vías, tal que:

$$\mathbb{X}_{[NPK]} = \{X_{[NP,1]} \mid X_{[NP,2]} \mid, \dots, \mid X_{[NP,K]}\} \quad (2.1)$$

donde $\{x_{ijk} \mid \text{para } i = 1, \dots, N; j = 1, \dots, Q \text{ y } k = 1, \dots, K\}$.

2.2.1 Structuration des Tableaux à Trois Indices de la Statistique (STATIS).–

Fue en 1976 cuando L'Hermier Des Plantes (1976) desarrolló el método STATIS y posteriormente Robert & Escoufier (1976), Lavit (1988) y Lavit et al. (1994) realizaron importantes contribuciones como la obtención del coeficiente de correlación vectorial (RV) y la descripción del escalado euclidiano entre las configuraciones de las mismas observaciones obtenidas en K circunstancias diferentes.

El STATIS se caracteriza por analizar una matriz \mathbb{X} de tres vías para obtener la parte estable entre las K matrices de datos, las que contienen en sus filas los mismos N individuos, pero con diferentes conjuntos de P variables en sus columnas, con la finalidad de resaltar la posición relativa de los individuos. Sin embargo, si las K matrices tienen las mismas P variables en sus columnas, pero con diferentes N individuos en sus filas, se tiene la alternativa conocida como STATIS DUAL que consiste en enfatizar las posiciones relativas de las variables.

El método STATIS realiza los análisis a partir de matrices de operadores que representan el producto vectorial $\mathbf{S}_{[NN,k]} = \mathbf{X}_{[NP,k]} \mathbf{X}_{[NP,k]}^T$ para cada K condición; mientras que en el STATIS DUAL las matrices de operadores se obtienen como $\mathbf{S}_{[PP,k]} = \mathbf{X}_{[NP,k]}^T \mathbf{X}_{[NP,k]}$, estas nuevas matrices \mathbf{S} poseen la característica de ser cuadradas definidas positivas.

La técnica se basa en tres etapas: i) *inter-estructura* que expresa la relación entre las K-matrices, ii) *compromiso* que obtiene la parte estable de las K-matrices e iii) *intra-estructura* que proyecta los individuos o variables (STATIS DUAL) de cada matriz inicial sobre el espacio bidimensional del compromiso, permitiendo analizar de forma independiente la información de cada una de las K-matrices.

Existen diversas contribuciones científicas de investigadores que han utilizado este método, por ejemplo: análisis de factores físico-químicos del agua del embalse Uribante (Márquez et al., 1992); estudio de los cambios de las comunidades de zooplancton en el Golfo de Tigullio y la influencia de factores hidro-climáticos (Licandro & Ibanez, 2000); estudio de la riqueza de especies y condiciones ambientales de los peces a lo largo de la costa noruega de Skagerrak (Lekve et al., 2002); análisis de concentraciones promedio anuales de componentes químicos en el agua monitoreados en espacio y tiempo (Stanimirova et al., 2004); para caracterizar los patrones estacionales de los ensamblajes de peces en el estuario de Gironde, Francia (Lobry et al., 2006); estudio de patrones espaciales de correlaciones cruzadas de procesos hidrobiológicos en Europa (Gudmundsson et al., 2011); para evaluar la biogeografía del fitoplancton: implicaciones ecológicas y de gestión (Feki-Sahnoun et al., 2018); entre otros.

Este método permite encontrar la parte estable de las K-matrices considerando el efecto de la escala espacial y/o temporal, enfatizando la relación entre los individuos (o variables en el caso del STATIS DUAL); tiene la característica que no requiere que la cantidad de individuos sea mayor o igual al número de variables.

Solo permite enfatizar las posiciones relativas de los individuos (STATIS) o solo de las variables (STATIS DUAL) y no de ambos de forma conjunta; su matriz consenso recoge, fundamentalmente, la estructura de las matrices altamente correlacionadas con el primer vector propio de la matriz de correlaciones vectoriales y pondera a todas las variables de una matriz con un mismo peso (aunque no hayan participado en esa asociación).

2.2.2. Análisis Triádico Parcial (PTA). – Este método se conoce también con el nombre de X-STATIS y fue propuesto por [Jaffrenou \(1978\)](#) y posteriormente tuvo una importante incursión en el campo de la ecología de la mano de [Thioulouse & Chessel \(1987\)](#). Tiene como objetivo buscar estructuras que sean estables entre la secuencia de las matrices. El PTA requiere que las K-matrices tengan igual número de P variables e igual número de N individuos. El principio del PTA es que trabaja directamente con las matrices originales y no con operadores como lo hace el STATIS.

Al igual que el STATIS este análisis consta de tres fases: i) análisis de la interestructura; ii) construcción y análisis del compromiso; y iii) análisis de la intraestructura; la diferencia radica en que el PTA proyecta los individuos y las variables de cada matriz inicial en el espacio bidimensional del compromiso, compromiso que en este caso es una media ponderada de las matrices de partida, mientras que en el STATIS (y STATIS DUAL) es una media ponderada de los operadores. El procedimiento matemático de este método es presentado de forma exhaustiva en el **Capítulo 3**, por formar parte de la base metodológica del presente trabajo.

Existen diversas contribuciones científicas de investigadores que han utilizado este método, por ejemplo: análisis de la estructura estable de variables físico-químicas medidas en estaciones de muestreo ([Thioulouse & Chessel, 1987](#)); para estudiar la distribución del zooplancton en la Reserva Natural de Berlengas (Portugal) ([Mendes et al., 2011](#)); identificación de patrones espaciales y estacionales de la

calidad del agua de los ríos en una cuenca mediterránea agrícola de regadío semiárida (Darwiche-Criado et al., 2015); para determinar la variabilidad de los ensamblajes de ostrácodos a lo largo de un ciclo hidrológico y analizar la respuesta de estas comunidades a variables ambientales y espaciales (Castillo-Escrivà et al., 2017); análisis de la distribución longitudinal de macro-invertebrados en una cuenca norteafricana muy húmeda: Oued Melloulou, Marruecos (Taybi et al., 2020); evaluación de la calidad del agua superficial en una región mediterránea semiárida de Medjerda, norte de Túnez mediante análisis triádico parcial (Slimani et al., 2020); entre otros.

Es un método que permite calcular la estructura en común entre las variables y los individuos al mismo tiempo; no requiere que la cantidad de individuos sea mayor o igual al número de variables. Además, tiene la particularidad de trabajar con las matrices iniciales, permitiendo estudiar la parte estable de las matrices a lo largo del tiempo o del espacio. Sin embargo, al ser un método basado en STATIS presenta las mismas restricciones del STATIS.

2.2.3 DISTATIS. – El DISTATIS (Abdi et al., 2007) es un método multivía que analiza matrices de distancias como punto de partida. Este método es una generalización del método de Escalamiento Multidimensional clásico (EMD). Cuando es utilizado como técnica de clasificación, el método considera los datos de clasificación individuales. Utiliza un mapa bi-factorial para representar la proximidad entre los puntos reflejando su similitud. Estos mapas pueden ser interpretados de la misma manera que los obtenidos con el EMD o ACP.

Para este método, existen diversos estudios de aplicación en el área del análisis sensorial y de alimentos. A la fecha no se cuenta con aplicaciones para el campo de la Ecología. Se opta por presentar este método ya que forma parte de la base metodológica del método propuesto en el **Capítulo 4**.

Es un método que calcula la parte estable de las similitudes. No obstante, requiere que las matrices de distancias tengan los mismos individuos.

2.3 Relación entre dos conjuntos de datos de tres vías.

Para analizar la relación entre dos conjuntos de datos de tres vías (dos cubos de datos), se requiere contar con dos matrices de tres vías \mathbb{X} (ver Ec. (2.1)) y \mathbb{Y} (ver Ec. (2.2)) que contienen en su tercera vía los mismos K tiempos o condiciones; cada par de matrices $\mathbf{X}_{[NP,k]}$ y $\mathbf{Y}_{[NQ,k]}$ con $k=1, \dots, K$ deben contener en sus filas la información de los mismos N individuos, pero pueden ser diferentes entre pares de matrices, sin embargo, las variables deben ser las mismas en las K -matrices para cada cubo de datos.

$$\mathbb{Y}_{[NQK]} = \{\mathbf{Y}_{[NQ,1]} \mid \mathbf{Y}_{[NQ,2]} \mid , \dots, \mid \mathbf{Y}_{[NQ,K]}\} \quad (2.2)$$

donde $\{y_{ij'k} \mid \text{para } i = 1, \dots, N; j' = 1, \dots, Q \text{ y } k = 1, \dots, K\}$.

2.3.1 STATICO.- El método STATICO (Simier et al., 1999; Thioulouse et al., 2004) se caracteriza por analizar la parte estable de la relación entre dos matrices de tres vías $\mathbf{X}_{[NPK]}$ y $\mathbf{Y}_{[NPK]}$ a través del tiempo o condiciones. Su procedimiento se basa en la combinación del método STATIS (Lavit et al., 1994) y del análisis de la Co-IA (Dolédec & Chessel, 1994). El método comienza con el cálculo de la estructura común (relación) entre cada par de matrices (mediante análisis de Co-IA), para posteriormente analizar el cubo de datos así obtenidos (las K -matrices cruzadas), mediante un PTA, para así estudiar la parte estable de la relación entre ambos cubos de datos.

El método STATICO consta de tres etapas: i) cada matriz de datos es analizada de manera independiente (ACP, AC o ACM) dependiendo del tipo de datos; ii) cada par de matrices es analizado con un análisis de Co-IA en $k=1, \dots, K$; iii) se lleva a

cabo un PTA del conjunto de K-matrices cruzadas obtenidas de los K análisis de Co-IA en el paso (ii).

Por las bondades que ofrece este método, es utilizado como base metodológica para el desarrollo de un nuevo método propuesto en el presente trabajo. La explicación del procedimiento matemático se presenta en detalle junto con el PTA y la Co-IA en el **Capítulo 3**.

Existen diversas contribuciones científicas de investigadores que han utilizado este método, por ejemplo: para analizar la caracterización de la tipología de los sitios de cultivo de café según la evolución temporal del vínculo entre sus características fitosanitarias y ambientales (Simier et al., 1999); para analizar la relación de la composición de especies con el medioambiente evaluados en 6 lugares y en 4 tiempos en Francia (Thioulouse et al., 2004); análisis de la relación entre las especies Ephemeroptera y las variables ambientales durante las estaciones del clima (Thioulouse, 2011); el estudio de la estructura espacio-temporal entre el ensamblaje de diatomeas y el medioambiente en un estuario y en la Ría de Aveiro Mendes et al. (2009) y Mendes et al. (2012) respectivamente; estudio de las respuestas de las etapas del ciclo de vida de los copépodos a la variabilidad climática en un estuario templado del sur de Europa (Gonçalves et al., 2012b); para investigar la distribución vertical diel de una comunidad de copepodos a lo largo del ciclo estacional en el estuario templado de Mondego, Portugal (Gonçalves et al., 2012a); entre otros.

Es un método que permite encontrar la parte estable de la estructura entre dos matrices de tres vías $\mathbb{X}_{[NPK]}$ y $\mathbb{Y}_{[NPK]}$. Debido a que en su metodología se utiliza como base el análisis del PTA, permite, además, analizar la relación entre las variables y los individuos. Sin embargo, es un método restrictivo al requerir que se mantengan los mismos individuos entre cada par de matrices $\mathbb{X}_{[NP,k]}$ y $\mathbb{Y}_{[NP,k]}$ (para realizar la Co-IA), y los mismos individuos y variables en cada matriz de tres

vías (para realizar el PTA), además de que los dos conjuntos de matrices de tres vías tengan el mismo número de K matrices.

2.3.2 COSTATIS. - El COSTATIS (Thioulouse, 2011) es una alternativa al método STATICO, su diferencia radica en que primero realiza un PTA de cada conjunto de datos de tres vías $\mathbb{X}_{[NPK]}$ y $\mathbb{Y}_{[NPK]}$, para después aplicar un análisis de CoIA entre las dos matrices compromiso obtenidas en la etapa anterior: la de ambiente y la de especies.

Existen diversas contribuciones científicas de investigadores que han utilizado este método, por ejemplo: análisis de la relación entre las especies Ephemeroptera y las variables ambientales durante las estaciones del clima (Thioulouse, 2011); análisis de la ecología de heteroptera acuáticos en la cuenca hidrográfica de Medjerda, Tunisia (Slimani et al., 2017) en comparación con el STATICO; entre otros.

Es un método que enfatiza la proyección de las variables y las observaciones en el compromiso al momento de realizar el paso del PTA. Además, permite maximizar la información contenida en las matrices del compromiso. Este método permite que las K condiciones puedan variar entre ambas matrices de tres vías. Aunque, requiere que las matrices de tres vías $\mathbb{X}_{[NPK]}$ y $\mathbb{Y}_{[NPK]}$ contengan los mismos individuos y variables en cada matriz que compone cada conjunto de matrices de tres vías para realizar el análisis del PTA.

PARTE II

CAPÍTULO 3

*Porque yo Jehová
soy tu Dios,
quien te sostiene
de tu mano derecha,
y te dice: No temas,
yo te ayudo.
Isaías 41:13*

CAPÍTULO 3: CUANTIFICAR LA MAGNITUD DE LA RELACIÓN ENTRE DOS MATRICES DE TRES VÍAS: MÉTODO HJ-STATICO

En el presente capítulo, se hace una exposición de la propuesta de una nueva técnica estadística que hemos denominado HJ-STATICO, que combina los métodos multivariantes STATICO con una variante de la metodología propuesta por [Abdi et al. \(2012\)](#) y el HJ-Biplot. Su nombre se deriva de estos dos métodos.

Este método permite realizar el análisis de la parte estable de la co-estructura entre dos matrices de tres vías $\mathbb{X}_{[NPK]}$ y $\mathbb{Y}_{[NQK]}$, permitiendo a su vez cuantificar y representar gráficamente la contribución y la calidad de representación de las variables en la construcción de la matriz compromiso y la intra-estructura.

3.1 Descripción del método HJ-STATICO.

Se tienen dos conjuntos de matrices de tres vías $\mathbb{X}_{[NPK]}$ y $\mathbb{Y}_{[NQK]}$ (Fig. 3.1A) que tienen en las filas de los pares de matrices ($\mathbf{X}_{[NP,k]}$ y $\mathbf{Y}_{[NQ,k]}$; para $k=1,\dots,K$) los mismos N individuos, y en cada conjunto de matrices las mismas P y Q variables con igual K condición, respectivamente (Fig. 3.1A).

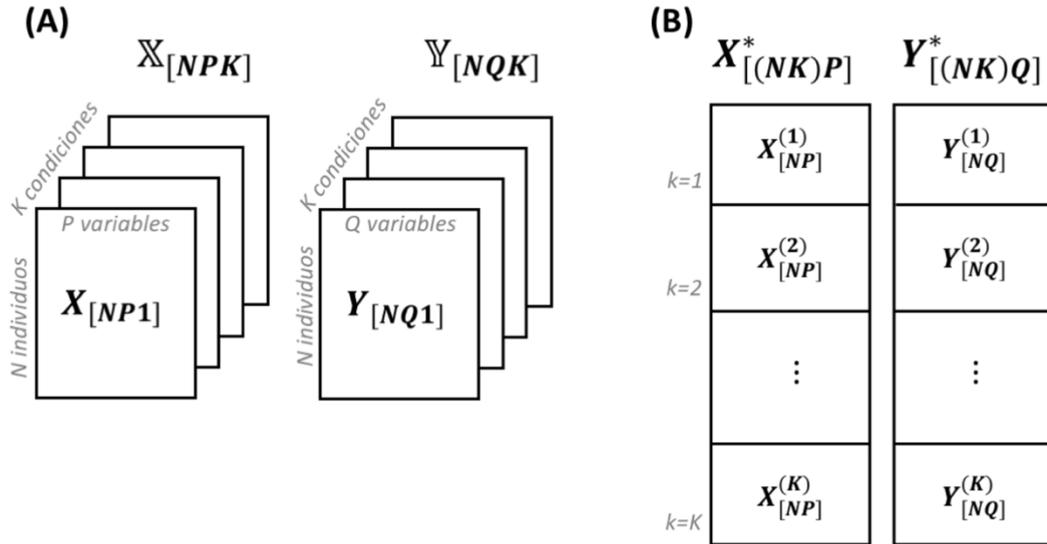


Figura 3.1 A) Esquema de la estructura de dos conjuntos de matrices $\mathbb{X}_{[NPK]}$ e $\mathbb{Y}_{[NQK]}$ y B) ordenación de las matrices A) para el análisis.

Previamente los conjuntos de matrices de datos (Fig. 3.1A) deben ser preprocesados para posteriormente, ordenarlos de forma yuxtapuesta vertical (Fig. 3.1B) manteniendo el orden establecido por la tercera vía (Fig. 3.1A), de tal forma que las nuevas matrices yuxtapuestas $\mathbf{X}^*_{[(NK)P]}$ y $\mathbf{Y}^*_{[(NK)Q]}$ conserven y permitan identificar por secciones las posiciones de las K-matrices iniciales.

Estas nuevas matrices totales (Fig. 3.1B) tienen como dimensión $(N * K)$ filas y P columnas para $\mathbb{X}_{[NPK]}$ y Q columnas para $\mathbb{Y}_{[NQK]}$. Sin embargo, para efectos de interpretación e identificación de posiciones de las K-matrices, en el presente trabajo se hace referencia a las matrices yuxtapuestas como matrices compuestas por índices que representan las N filas, P columnas y la ubicación de cada una de las k -matrices contenidas en la matriz total, por ejemplo: $\mathbf{X}^{(k)}_{[NP]}$

Los pasos para el desarrollo del HJ-STATICO se basan en la metodología presentada por: [Dolédéc & Chessel \(1994\)](#) y [Dray et al. \(2003\)](#) (Co-IA); [Jaffrenou \(1978\)](#) y [Thioulouse & Chessel \(1987\)](#) (PTA); [Simier et al. \(1999\)](#) y [Thioulouse](#)

et al. (2004) (STATICO); Abdi et al. (2012) (estandarización de matrices y re-escalamiento de los pesos); Galindo-Villardón (1986) y Galindo-Villardón & Cuadras (1986) (HJ-Biplot).

A continuación, se presenta una explicación detallada del proceso algebraico que comprende el método HJ-STATICO.

Paso 1. Realizar un preprocesamiento (estandarización) a cada conjunto de matrices de datos.

El preprocesamiento consiste en centrar, normalizar y estandarizar los datos. Se realiza con la finalidad de eliminar el efecto que causaría la presencia de diferentes promedios y escalas en la matriz de datos, tales como ocultar las relaciones entre las variables, además de eliminar las diferencias que pueden ocurrir entre las puntuaciones de los individuos con las variables. Con la estructura de los datos en tres vías se incrementa la complejidad de este proceso ya que hay más promedios y factores de escala a considerar, como tres conjuntos de promedios de una-vía, tres conjuntos de promedios de dos vías y un solo promedio general; así como también un factor de normalización total, tres tipos de normalizaciones por fibra y tres tipos de normalizaciones por cortes (Fig. 3.2) (Kroonenberg, 2008, p.109-110,112).

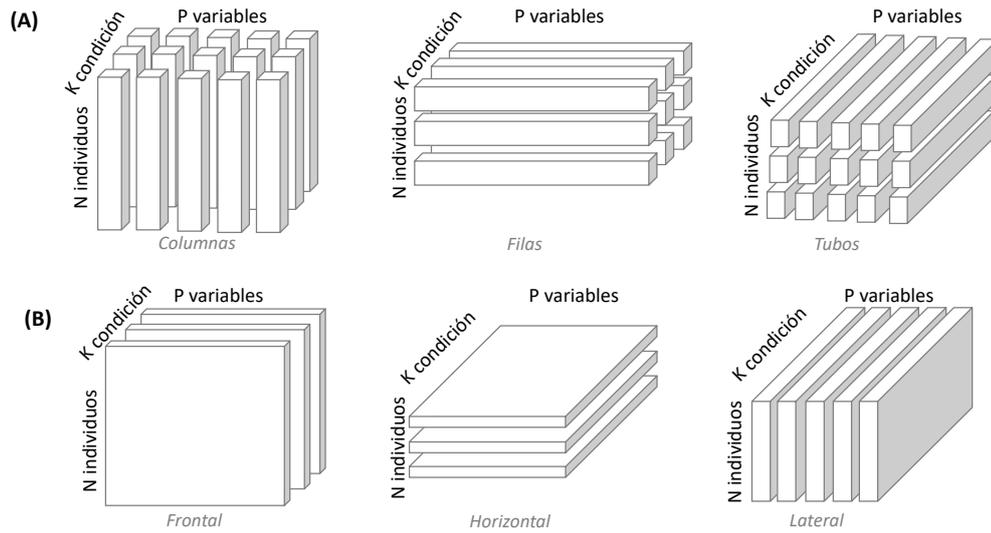


Figura 3.2. Opciones de normalizar una matriz de tres vías: A) representación de fibras y B) cortes. Tomado del libro de (Kroonenberg, 2008, p.29).

Para esta investigación, se establece a modo de ejemplo a $\mathbb{X}_{[NPK]}$ como el conjunto correspondiente a los datos ambientales y a $\mathbb{Y}_{[NPK]}$ como el conjunto de los datos de especies.

Para el conjunto de matrices de **datos ambientales** $\mathbb{X}_{[NPK]}$ se realiza una primera estandarización total por columna (variables) a la matriz ordenada en la Fig. 3.1B, que consiste en restarle a cada elemento x_{ij} la media \bar{x}_j (centrado) y posteriormente dividirlo por σ_j (es decir, normalizarlo con la desviación estándar de la población) para $j = 1, \dots, P$ variables, tal como aparece en la siguiente expresión:

$$\mathbf{Z}_{[(NK)P]} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} ; \text{ para } i = 1, \dots, (N * K); j = 1, \dots, P \quad (3.1)$$

Posteriormente, a la matriz $\mathbf{Z}_{[(NK)P]}$ estandarizada de dimensión $((N * K) \times P)$, se le aplica una segunda estandarización, teniendo como alternativas los siguientes tres métodos: Estandarización Parcial y

Estandarización Total (Bouroche, 1975) y Estandarización por Norma (Abdi et al., 2012). Es necesario indicar que en la matriz $\mathbf{Z}_{[(NK)P]}$ se debe identificar la posición de cada una de las k-matrices como se detalla a continuación:

- Estandarización Parcial: Centra y normaliza las columnas por cada k sección logrando que las K-matrices iniciales tengan media cero y varianza unitaria;

$$\mathcal{X}_{[NP]}^{(k)} = \frac{\mathbf{Z}_{ij}^{(k)} - \bar{x}_j^{(k)}}{\sigma_j^{(k)}} ; \text{para } i = 1, \dots, N ; j = 1, \dots, P \text{ y } k = 1, \dots, K \quad (3.2)$$

- Estandarización Total: Centra las columnas por cada k sección como se indica en la Ec. (3.3) y luego normaliza para la matriz total ver (Ec. 3.4), logrando que las k-matrices iniciales tengan media cero y la varianza de la matriz total sea uno.

$$\mathcal{X}'_{[NP]}^{(k)} = \mathbf{Z}_{ij}^{(k)} - \bar{x}_j^{(k)} ; \text{para } i = 1, \dots, N ; j = 1, \dots, P \text{ y } k = 1, \dots, K \quad (3.3)$$

$$\mathcal{X}_{[NP]} = \frac{\mathcal{X}'_{ij}}{\sigma_j} ; \text{para } i = 1, \dots, (N * K) \text{ y } j = 1, \dots, P \quad (3.4)$$

- Estandarización por Norma: Este método funciona de manera similar al método de dividir cada elemento de la matriz total para la raíz cuadrada de la suma total de los elementos al cuadrado, garantizando que la varianza total de cada k-matriz sea igual a uno y que todas las K-matrices participen equivalentemente en el análisis. Una forma similar a este método es utilizar la norma de la matriz (Abdi et al., 2012) como se presenta a continuación:

$$\mathcal{X}_{[NP]}^{(k)} = \mathbf{Z}_{[NP]}^{(k)} \times \left\| \mathbf{Z}_{[NP]}^{(k)} \times \mathbf{Z}_{[NP]}^{(k)T} \right\|^{-\frac{1}{2}} ; \text{para } k = 1, \dots, K \quad (3.5)$$

Para el conjunto de matrices de **datos de especies** $\mathbb{Y}_{[NQK]}$ se realiza un centrado por columna de la matriz total (Fig. 3.1B) para $j' = 1, \dots, Q$;

$$\mathbf{y}_{[NQ]} = \mathbf{y}_{i,j'} - \bar{y}_{j'}; \text{ para } i = 1, \dots, (N * K); j' = 1, \dots, Q \quad (3.6)$$

Paso 2. Realizar la Co-IA a cada par de matrices estandarizadas en el Paso 1 ($\mathbf{X}_{[NP]}^{(k)}$ y $\mathbf{Y}_{[NQ]}^{(k)}$; para $k = 1, \dots, K$):

Para este análisis se utilizó la descripción metodológica propuesta por [Dolédéc & Chessel \(1994\)](#) y [Dray et al. \(2003\)](#). Se inicia con dos matrices de dos vías estandarizadas (centradas y normalizadas) previamente en el Paso 1;

$$\mathbf{X}_{[NP]} \text{ y } \mathbf{Y}_{[NQ]} \quad (3.7)$$

que pueden ser expresadas en forma de triplas, (Ec. 3.8), tal que, las matrices diagonales D_N contiene los pesos ($1/N$) para los N individuos (filas), mientras que D_P y D_Q contienen los pesos (unitarios) asociados a las P y Q variables (columnas) respectivamente:

$$(\mathbf{X}, D_P, D_N) \text{ y } (\mathbf{Y}, D_Q, D_N) \quad (3.8)$$

además, permiten representar el esquema de dualidad ([Escoufier, 1987](#)). A continuación, cada tripleta es diagonalizada de la siguiente manera:

Para la primera matriz:

$$(\mathbf{X}, D_P, D_N) = R\Omega C; R \in \mathbb{R}^{N \times r}, C \in \mathbb{R}^{P \times r} \text{ y } \Omega \in \mathbb{R}^{r \times r} \quad (3.9)$$

siendo r el rango de la matriz \mathcal{X} , Ω la matriz que contiene los valores propios (v_1, \dots, v_r) almacenados en el vector \vec{v} , R contiene las puntuaciones de las N filas y C las puntuaciones de las P filas en los r ejes.

Para la segunda matriz:

$$(\mathcal{Y}, D_Q, D_N) = R\Gamma C ; R \in \mathbb{R}^{N \times s}, C \in \mathbb{R}^{Q \times s} \text{ y } \Gamma \in \mathbb{R}^{s \times s} \quad (3.10)$$

siendo s el rango de la matriz, Γ la matriz que contiene los valores propios (u_1, \dots, u_s) almacenados en el vector \vec{u} , R contiene las puntuaciones de las N filas y C las puntuaciones de las Q filas en los s ejes.

Por consiguiente, ξ y Ψ son matrices columnas que contienen N coordenadas generadas por las proyecciones del espacio multidimensional asociado con \mathcal{X} y \mathcal{Y} en los vectores \vec{u} (normalizado por D_P) y \vec{v} (normalizado por D_Q) respectivamente:

$$\xi = \mathcal{X}D_P u \quad \text{y} \quad \Psi = \mathcal{Y}D_Q v \quad (3.11)$$

A partir de aquí se obtiene la Co-IA como:

$$H(u, v) = \xi^T D_N \Psi \cong (\mathcal{Y}^T D_N \mathcal{X}, D_P, D_Q) \quad (3.12)$$

conocido como la tripleta estadística (Dray et al., 2003).

Su medida de covarianza es:

$$Cov(\xi, \Psi) = Corr(\xi, \Psi) \times VAR(\xi) \times VAR(\Psi) \quad (3.13)$$

$$Cov(\xi, \Psi) = Corr(\mathcal{X}D_P u, \mathcal{Y}D_Q v) \times (VAR(\mathcal{X}D_P u))^{1/2} \times (VAR(\mathcal{Y}D_Q v))^{1/2} \quad (3.14)$$

Por lo tanto, se denota a la Co-IA como una matriz cruzada, tal que:

$$\mathbf{Z}'_{[QP]}^{(k)} = \mathbf{y}_{[NQ]}^{(k)T} \mathbf{D}_N^{(k)} \mathbf{x}_{[NP]}^{(k)} \quad (3.15)$$

para $k = 1, \dots, K$. Según la definición de PTA que se detalla en el Paso 3, el compromiso es un promedio ponderado de las k-matrices cruzadas. Sus columnas son las columnas de \mathbf{X} y sus filas son las columnas de \mathbf{y} (Simier et al., 1999).

El proceso detallado en la Ec. (3.15) se repite para cada par de matrices, obteniéndose como resultado nuevas k-matrices cruzadas de la forma $\mathbf{Z}'_{[QP]}^{(k)}$, las cuales forman un nuevo conjunto de matrices $\mathbb{Z}'_{[QPK]}$ (cubo de datos) que contiene las mismas Q variables (filas) y P variables (columnas) en cada k-matriz.

Paso 3. Realizar el PTA clásico o el PTA incorporando el HJ-Biplot en el análisis del compromiso y de la intra-estructura.

A partir del resultado de la Co-IA obtenido en el Paso 2, se analiza la matriz de tres vías $\mathbb{Z}'_{[QPK]}$ aplicando la metodología propuesta por Jaffrenou (1978), es decir, el PTA.

Cada k-matriz puede ser expresada por las tripletas que contienen la matriz diagonal D_Q con los pesos para las filas y D_P los pesos para las columnas.

$$(\mathbf{Z}'_1, D_Q, D_P), (\mathbf{Z}'_2, D_Q, D_P), \dots, (\mathbf{Z}'_K, D_Q, D_P) \quad (3.16)$$

El procedimiento consta de tres etapas, a saber:

- 1) Inter-estructura

Se calcula el producto interno entre las k-matrices y se obtiene una matriz cuadrada C de dimensión $(K \times K)$ que es semidefinida positiva:

$$C_{[KK]} = COVV(\mathbf{Z}'_k, \mathbf{Z}'_l) = Trace(\mathbf{Z}'_k{}^T D_Q \mathbf{Z}'_l D_P) = Trace(\mathbf{Z}'_l{}^T D_Q \mathbf{Z}'_k D_P) \quad (3.17)$$

siendo k y $l = 1, \dots, K$.

Simier et al.(1999) mencionaron que en esta forma de cálculo de la varianza vectorial $VARV(\mathbf{Z}'_k)$, la matriz de covarianzas vectorial entre matrices $COVV(\mathbf{Z}'_k, \mathbf{Z}'_l)$ y la correlación vectorial entre matrices $RV(\mathbf{Z}'_k, \mathbf{Z}'_l)$ están basadas en las nociones definidas por Escoufier (1973).

Para obtener la varianza vectorial se aplica:

$$VARV(\mathbf{Z}'_k) = COVV(\mathbf{Z}'_k, \mathbf{Z}'_k) = Trace(\mathbf{Z}'_k{}^T D_Q \mathbf{Z}'_k D_P) \quad (3.18)$$

y para el coeficiente de correlación vectorial RV;

$$RV_{[KK]} = RV(\mathbf{Z}'_k, \mathbf{Z}'_l) = \frac{COVV(\mathbf{Z}'_k, \mathbf{Z}'_l)}{\sqrt{VARV(\mathbf{Z}'_k)} \sqrt{VARV(\mathbf{Z}'_l)}} \quad (3.19)$$

Como siguiente paso, se aplica una DVS a la matriz cuadrada RV para obtener el primer vector propio \vec{u}_1 de dimensión $(K \times 1)$ que contiene los pesos que se utilizarán para la construcción de la matriz compromiso, tal como sigue:

$$RV = U\Theta U^T; \text{ tal que } U^T U = I \text{ y } \Theta \in \mathbb{R}^{K \times K} \quad (3.20)$$

Se utilizan los dos primeros vectores propios $\{\vec{u}_1 \text{ y } \vec{u}_2\}$ de U para realizar la representación gráfica de la inter-estructura sobre un plano ACP de dos dimensiones:

$$G = U\text{Diag}(\Theta)^{1/2} \quad (3.21)$$

La descomposición de la matriz RV da como resultado que los elementos del primer vector propio \vec{u}_1 tienen el mismo signo (el cual debe ser considerado como positivo para el análisis). Este resultado es análogo al teorema de Frobenius en el que especifica que una matriz semi-definida positiva cuyos elementos son positivos, tiene un primer vector propio cuyos elementos conservan el mismo signo. Este vector otorga las comunalidades de cada k-matriz con respecto a las demás (Abdi et al., 2012).

2) Compromiso

Previo a la obtención de la matriz compromiso se aplica el método propuesto por Abdi et al. (2012), que consiste en el re-escalamiento de los pesos contenidos en el primer vector propio \vec{u}_1 , como se detalla a continuación:

$$\vec{\alpha}_{[k]} = \vec{u}_1 \times (\vec{u}_1^T \mathbf{1})^{-1}; \text{ tal que } \sum_{k=1}^K \alpha_{[k]} = 1 \quad (3.22)$$

La matriz compromiso definida en la Ec. (3.23) es calculada como la suma ponderada de las k-matrices cruzadas que son el resultado de la Co-IA, utilizando los pesos obtenidos de la Ec. (3.22).

$$\text{Compromiso} = \sum_{k=1}^K \alpha_{[k]} \mathbf{Z}'_{[QP,k]} \quad (3.23)$$

Posteriormente, al compromiso (Ec. 3.24) se le aplica una DVS, con la finalidad de obtener los vectores propios que se utilizan posteriormente para su representación gráfica en un mapa ACP,

$$\text{Compromiso} = U\Xi V^T; U \in \mathbb{R}^{Q \times h}, V \in \mathbb{R}^{P \times h} \text{ y } \Xi \in \mathbb{R}^{h \times h} \quad (3.24)$$

donde h es el rango de la matriz compromiso, U es una matriz de dimensión $(Q \times h)$ de los vectores singulares normalizados por la izquierda de la matriz y V es una matriz compromiso de dimensión $(P \times h)$ de los vectores singulares normalizados por la derecha de la matriz compromiso y Ξ es una matriz diagonal de dimensión $(h \times h)$ que contiene los valores singulares de la matriz compromiso.

A partir de los vectores propios obtenidos con la Ec. (3.24) se calculan las cargas para las variables P y Q , que permitirán representar, mediante un gráfico bi-dimensional, la parte estable de la co-estructura entre $\mathbb{X}_{[NPK]}$ y $\mathbb{Y}_{[NQK]}$, tal como se detalla a continuación,

$$F = \text{Compromiso} \times V \quad \text{y} \quad Q' = \text{Compromiso}^T \times U \quad (3.25)$$

siendo F las cargas para las Q variables en $\mathbb{Y}_{[NQK]}$ y Q' para las P variables en $\mathbb{X}_{[NPK]}$.

3) Intra-estructura

Se proyecta cada k -matriz cruzada $\mathbf{Z}'_{[QP,k]}$ sobre el mapa bi-factorial de la matriz compromiso, para lo cual se utilizan los vectores propios U y V obtenidos en la Ec. (3.24), como se detalla,

$$FK = \mathbf{Z}'_{[QP,k]} \times V \quad \text{y} \quad Q'K = \mathbf{Z}'^T_{[QP,k]} \times U; \text{ para } k = 1, \dots, K \quad (3.26)$$

Si el usuario desea analizar la matriz compromiso y las matrices de intra-estructura desde una perspectiva Biplot, se procede a emplear el método HJ-Biplot utilizando la base metodológica propuesta por [Galindo-Villardón \(1986\)](#) y [Galindo-Villardón & Cuadras \(1986\)](#).

Se escogió el método HJ para obtener la máxima calidad de representación para las P y Q variables (ambiente y especie) sobre un mismo sistema de ejes bifactoriales.

Para realizar un Biplot, siguiendo a [Gabriel \(1971\)](#), se parte de una matriz $\mathbf{X}_{[NP]}$ de rango r , con N filas (individuos) y P columnas (variables), que puede ser factorizada como se detalla en la ecuación Ec. (3.27).

Por tanto, en nuestra alternativa metodológica se tomará la matriz $\mathbf{X}_{[NP]}$ como la matriz **Compromiso** $_{[QP]}$ y las k -matrices cruzadas $\mathbf{Z}'_{[QP,k]}$, para $k = 1, \dots, K$,

$$\text{Compromiso} \cong GH^T \quad (3.27)$$

Tal que, G es una matriz de dimensión $(Q \times r)$ y H de dimensión $(P \times r)$, donde cada elemento $\text{Compromiso}_{ij} = g_i^T h_j$ puede ser calculado como el producto de la i -ésima fila de G y la j -ésima columna de H^T . Los vectores g_i son considerados los marcadores filas y los h_j los marcadores columna; para $i = 1, \dots, Q$ y $j = 1, \dots, P$.

Se aplica una DVS a la matriz **Compromiso** (Ec. 3.28) a partir de la cual podemos obtener los Biplots clásicos JK y GH, presentados en la Ec. (3.29).

$$X = U\Sigma V^T \quad (3.28)$$

$$\mathbf{JK - Biplot}: J = U\Sigma \text{ y } K = V \quad \text{y} \quad \mathbf{GH - Biplot}: G = U \text{ y } H = V\Sigma \quad (3.29)$$

[Galindo-Villardón \(1986\)](#) realiza una contribución a los métodos Biplot, al presentar el HJ-Biplot que consiste en una representación simultánea con máxima calidad de representación para los marcadores filas j_1, \dots, j_I y los marcadores

columnas h_1, \dots, h_p proyectados sobre un mismo sistema de referencia de ejes factoriales, eligiendo como matrices H y J:

$$HJ: J = U\Sigma \text{ y } H = V\Sigma \quad (3.30)$$

siendo σ el valor singular de la matriz Σ .

El HJ-Biplot permite cuantificar e interpretar a través de indicadores presentados en las Ecs. (3.31) a la (3.38) las relaciones entre los ejes factoriales con las variables y con los individuos; además de obtener las contribuciones relativas del factor al elemento y del elemento al factor; así como también la calidad de representación de las filas y las columnas de la matriz (Galindo-Villardón & Cuadras, 1986):

- Contribución relativa a la traza del elemento i -ésimo (fila), representa la variabilidad total explicada por el individuo.

$$CRT_i = \frac{\sum_{s=1}^r j_{is}^2}{\sum_{s=1}^r \sigma_s^2} \quad (3.31)$$

- Contribución relativa del elemento i -ésimo (fila) al factor l , representa la variabilidad del factor explicada por el individuo.

$$CRE_i F_l = \frac{j_{il}^2}{\sigma_l^2} \quad (3.32)$$

- Contribución del factor l al elemento i -ésimo (fila), representa la parte de la variabilidad de cada individuo explicada por el factor.

$$CRF_l E_j = \frac{j_{il}^2}{\sum_{s=1}^r j_{is}^2} \quad (3.33)$$

- Contribución relativa a la traza del elemento j -ésimo (columna), representa la variabilidad total explicada por la variable.

$$CRT_j = \frac{\sum_{s=1}^r h_{js}^2}{\sum_{s=1}^r \sigma_s^2} \quad (3.34)$$

- Contribución relativa del elemento j -ésimo (columna) al factor l , representa la variabilidad del factor explicada por la variable.

$$CRE_j F_l = \frac{h_{jl}^2}{\sigma_l^2} \quad (3.35)$$

- Contribución del factor l al elemento j -ésimo (columna), representa la parte de la variabilidad de cada variable explicada por el factor.

$$CRF_l E_j = \frac{h_{jl}^2}{\sum_{s=1}^r j_{js}^2} \quad (3.36)$$

- Calidad de representación de la i -ésima (fila).

$$QLRE_i P_{I-II} = CRF_I E_i + CRF_{II} E_i \quad (3.37)$$

- Calidad de representación de la j -ésima (columna).

$$QLRE_j P_{I-II} = CRF_I E_j + CRF_{II} E_j \quad (3.38)$$

Con el cálculo de los índices de contribución relativa del factor a los elementos fila (individuos) se logra interpretar la cantidad de variabilidad de cada individuo que es explicada por el factor. De igual forma ocurre con la contribución relativa del factor a los elementos columna (variables), que revela la cantidad de variabilidad explicada por el factor para cada variable. Mientras que, la calidad de representación para las filas expresa la cantidad total de contribución relativa de los individuos que consiguen revelar los factores al momento de proyectarlos sobre

el mapa factorial, un significado similar resulta con la calidad de representación para las variables.

A continuación, se representa mediante un esquema los pasos a seguir para realizar el análisis utilizando el HJ-STATICO (Fig. 3.3):

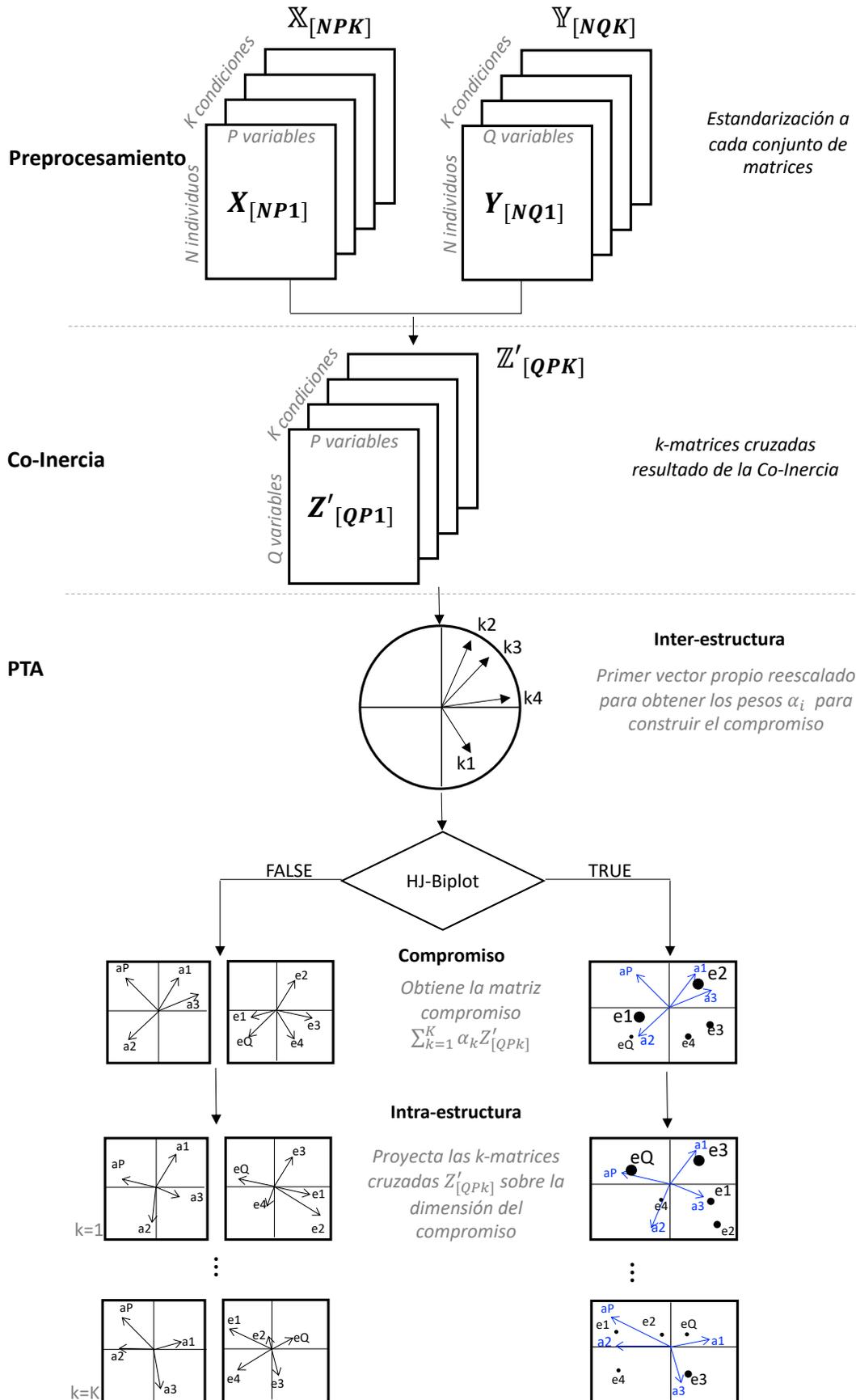


Figura 3.3 Esquema de los pasos para HJ-STATICO.

3.1.1 Ventajas que ofrece el HJ-STATICO.

Entre los principales beneficios que ofrece esta nueva propuesta de análisis, se destacan los siguientes:

- Mayor facilidad para realizar el análisis;
- Cálculo de indicadores del HJ-Biplot;
- Facilita el análisis de la estructura en común de los dos cubos de datos basándose en indicadores (para el compromiso y la intra-estructura);
- Representación gráfica aplicando HJ-Biplot;
- Diferentes ejes para presentar el mapa bi-factorial.

Para poder utilizar esta nueva propuesta en el análisis de datos reales, se desarrolló una librería en RStudio, denominada HJSTATICO la cual será presentada en detalle en el **Capítulo 5**.

3.2 Aplicación del HJ-STATICO a datos reales.

Para poner de manifiesto la utilidad de la metodología HJ-STATICO, propuesta como alternativa al STATICO, así como la utilización de la herramienta informática generada, se realizará el análisis con datos reales. Esta investigación está orientada al estudio del ecosistema marino de la región de Ecuador Continental, país que se encuentra ubicado en la costa Oeste de América del Sur (Fig. 3.4).

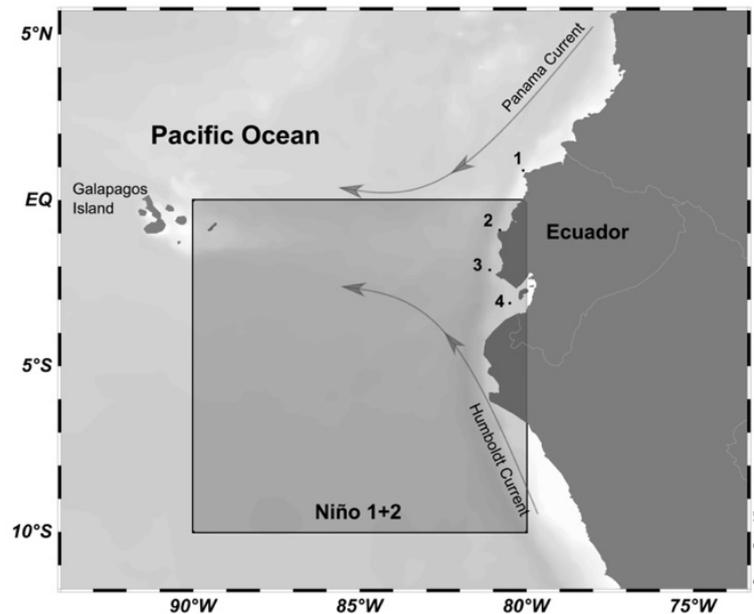


Figura 3.4 Mapa del perfil costero de Ecuador Continental con la localización geográfica de las estaciones de muestreo: 1. Esmeraldas (0.92° N, 80.11° W), 2. Manta (0.88° S, 80.83° W), 3. La Libertad (2.08° S, 81.12° W), 4. Puerto Bolívar (Pto. Bolívar, 3.11° S, 80.49° W) y la Región Niño 1+2: (0° - 10° S y 90° W - 80° W).

3.2.1 Importancia de la zona de estudio.

El ecosistema marino de la región de Ecuador Continental tiene un perfil costero (Continental e Insular) perteneciente a la zona del Pacífico Ecuatorial Oriental y está influenciado por un complejo sistema de corrientes marinas y por las aguas provenientes del Océano Pacífico. Se caracteriza por contar con una climatología variada ocasionada por la influencia del mar, la presencia de corrientes marinas frías (Humboldt) y cálidas (Panamá), así como también por la orientación perpendicular de los Andes con presencia de los vientos Alisios (INOCAR, 2012). Además, su perfil costero tiene la particularidad de estar localizado en la Región Niño 1+2 (Fig. 3.4), considerada como una zona de gran importancia para el estudio de los eventos El Niño Oscilación del Sur (ENOS: eventos cálidos El Niño y eventos fríos La Niña), por su sensibilidad para detectar variaciones en la temperatura superficial del mar provenientes del Pacífico Central. La costa oeste de América del Sur se ha visto afectada de manera continua por la presencia de los

eventos ENOS, causando alteraciones en los sistemas océano-atmosféricos, en los ecosistemas naturales (marinos y terrestres) y en los sectores socio-económicos (Santos, 2006).

Dada su importancia, se han realizado investigaciones orientadas al comportamiento ambiental, biológico y el efecto del evento ENOS en las costas ecuatorianas. Se han aplicado técnicas estadísticas descriptivas y multivariantes. Entre las principales investigaciones realizadas con técnicas descriptivas se encuentran las realizadas por Jiménez & Bonilla (1980); Gualancañay et al. (2004); Tapia & Naranjo (2004); Torres, Calderón, et al. (2004); Torres, Mero, et al. (2004); Torres (2005); Cruz et al. (2011) y Jiménez (2014). En cuanto a los estudios desarrollados con métodos multivariantes podemos detallar el trabajo de (Perugachi-Salamea et al., 2014) que hicieron una clasificación de las estaciones costeras del Ecuador aplicando ACP y Curvas de Andrews; la investigación de Torres et al. (2017) que estudiaron la variabilidad espacio-temporal del fitoplancton y variables ambientales en el Golfo de Guayaquil, utilizando el CCA y el trabajo de Borbor-Cordova et al. (2019), que emplearon ACM para estudiar la estacionalidad de la ocurrencia de HAB (floración de algas nocivas) en relación con variables oceanográficas en la costa central ecuatoriana y en el Golfo de Guayaquil.

En vista de lo anterior y dada la importancia y complejidad que representa estudiar el fenómeno descrito, esta investigación pretende proponer una alternativa, considerando la estructura de tres vías, que permita complementar los estudios existentes, así como llegar a nuevos descubrimientos y conclusiones.

3.2.2 Los datos.

Para el análisis se cuenta con un conjunto de datos oceanográficos físico-químicos y biológicos proporcionados por el Instituto Oceanográfico de la Armada de Ecuador (INOCAR). Se realizaron muestreos durante las horas de la mañana en cuatro estaciones ubicadas a 10 millas del perfil costero de Ecuador Continental,

localizadas de norte a sur: Esmeraldas (norte - estuario), La Libertad (centro norte - oceánica), Manta (centro sur - oceánica) y Pto. Bolívar (sur - estuario) (Fig. 3.4).

Las mediciones se realizaron de forma mensual (excepto en enero) durante los años 2013, 2014 y 2015 (presencia de condiciones oceanográficas normales y cálidas extremas) en el mismo punto georreferenciado de cada estación, en el que se tomaban muestras en siete profundidades estándar (0m, 10m, 20m, 30m, 40m, 50m y 75m). Los meses son clasificados según la época climática a la que pertenecen, es decir, época lluviosa-invierno (diciembre - mayo) y época seca-verano (junio - noviembre) (INOCAR, 2012).

Los datos de partida están estructurados en dos conjuntos de matrices $\mathbb{X}_{[NPK]}$ (ambiente) y $\mathbb{Y}_{[NPK]}$ (especies), tal que en la primera-vía contienen los datos de las mismas $N=42$ mediciones realizadas en diferentes momentos (épocas climáticas) y profundidades, en la segunda-vía las $P=7$ y $Q=23$ variables (respectivamente para las matrices \mathbb{X} y \mathbb{Y}) y en la tercera-vía las $K=4$ estaciones de muestreo (factor espacial) (Fig. 3.5). Cabe resaltar que los dos conjuntos de matrices mantienen los mismos individuos y variables en sus K -matrices.

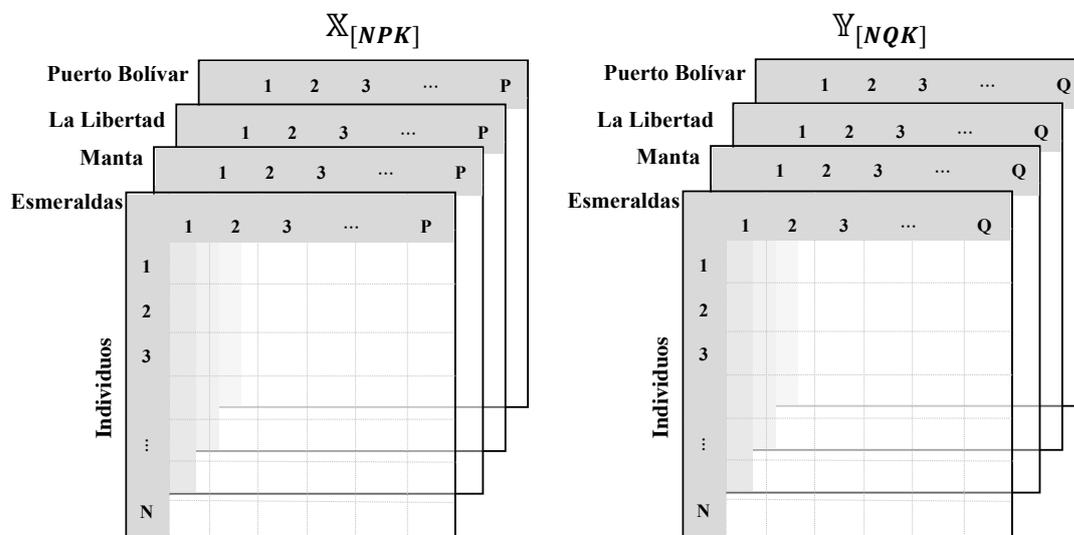


Figura 3.5. Estructura de los conjuntos de matrices de partida para ambiente y especies.

Como variables ambientales se estudiaron la temperatura (T en °C), oxígeno disuelto (OD en mg.L⁻¹), salinidad (S en ups), nitritos (NO₂ en µg-at.L⁻¹), nitratos (NO₃ en µg-at.L⁻¹), fosfatos (PO₄ en µg-at.L⁻¹) y silicatos (SiO₄ en µg-at.L⁻¹). Como variables biológicas se ha tenido en cuenta la abundancia (cel L⁻¹) de 23 especies del fitoplancton (Tabla 3.1) que han sido previamente transformadas aplicando log₁₀(x+1), con la finalidad de minimizar el efecto dominante de las capturas excepcionales (Legendre & Legendre, 1979).

Tabla 3.1 Lista de las 23 especies de fitoplancton.

Grupo	Orden	Familia	Género	Especies	ID		
Diatomeas Céntricas	Chaetocerotales	Chaetocerotaceae	<i>Chaetoceros</i>	<i>affinis</i>	e1		
			<i>Chaetoceros</i>	<i>curvisetus</i>	e2		
			<i>Chaetoceros</i>	<i>didymus</i>	e3		
	Hemiaulales	Hemiaulaceae	<i>Hemiaulus</i>	<i>sinensis</i>	e4		
	Leptocylindrales	Leptocylindraceae	<i>Leptocylindrus</i>	<i>danicus</i>	e5		
	Lithodesmiales	Lithodesmiaceae	<i>Ditylum</i>	<i>brightwellii</i>	e6		
	Rhizosoleniales	Rhizosoleniaceae	<i>Dactyliosolen</i>	<i>fragilissimus</i>	e7		
			<i>Guinardia</i>	<i>striata</i>	e8		
			<i>Proboscia</i>	<i>alata</i>	e10		
			<i>Rhizosolenia</i>	<i>imbricata</i>	e9		
			<i>Dactyliosolen</i>	<i>antarcticus</i>	e11		
			Thalassiosirales	Lauderiaceae	<i>Lauderia</i>	<i>borealis</i>	e13
				Skeletonemaceae	<i>Skeletonema</i>	<i>costatum</i>	e12
	Thalassiosiranae	Thalassiosirales	<i>Thalassiosira</i>	sp.	e14		
Diatomeas Pennadas	Bacillariales	Bacillariaceae	<i>Nitzschia</i>	<i>longissima</i>	e16		
			<i>Nitzschia</i>	sp.	e15		
			<i>Pseudo-nitzschia</i>	<i>pungens</i>	e17		
	Naviculales	Naviculaceae	<i>Navicula</i>	sp.	e18		
		Stauroneidaceae	<i>Stauroneis</i>	<i>membranacea</i>	e19		
	Thalassionematales	Thalassionemataceae	<i>Thalassionema</i>	<i>nitzschioides</i>	e20		
	Dinoflagelados	Gymnodiniales	Gymnodiniaceae	<i>Gyrodinium</i>	sp.	e22	
<i>Gymnodinium</i>				sp.	e21		
Ciliados	Cyclotrichiida	Mesodiniidae	<i>Mesodinium</i>	<i>rubrum</i>	e23		

3.2.3 Análisis realizados con la estructura de tres vías.

Con los datos descritos en la sección 3.2.2. se realizaron varios análisis, y a continuación se detallarán los más relevantes, que definen la estructura de las matrices que posteriormente se utilizará en la sección 3.2.4, para la aplicación de la propuesta HJ-STATICO:

- *Análisis 1:* Aplicación del método STATICO para estudiar el comportamiento espacio-temporal de las especies del fitoplancton ante las variaciones

ambientales que ocurrieron durante la presencia de condiciones oceanográficas normales y cálidas extremas en todo el perfil costero. Para este análisis se consideró la estructura definida en la Fig. 3.5 estableciendo $N=42$ filas (individuos) como los valores promedios de las variables para cada una de las siete profundidades estándar, según la época climática (seca y lluviosa). Como resultado se identificaron dos patrones de comportamiento para la relación especie-ambiente, uno conformado por las estaciones del tipo estuario (Esmeralda al norte y Pto. Bolívar al sur) y otro por las estaciones con influencia oceánica (Manta y La Libertad ubicados en el centro del perfil). La co-estructura más fuerte se encontró en los primeros 20m de profundidad principalmente en Pto. Bolívar (considerada como zona de alta riqueza biológica). Dependiendo de la localización geográfica de las estaciones (espacio) se observaron co-estructuras más fuertes asociadas a la época climática (tiempo). Durante condiciones cálidas extremas (2015) la co-estructura se debilita en comparación con los años anteriores. El soporte de estos resultados se encuentra en un manuscrito sometido en una revista de alto impacto, ver Anexo A1.

- *Análisis 2:* Para este análisis se utilizaron los programas RStudio para realizar el método STATICO y el Multbliplot (Vicente-Villardón, 2016) para el HJ-Biplot. Combinamos los dos métodos con la finalidad de cuantificar mediante indicadores la variabilidad que es aportada por los factores que representan el mapa bi-factorial sobre el cual, se proyectó la relación especie-ambiente durante los eventos normales y cálidos. Los datos fueron clasificados según el año de muestreo (2013, 2014 y 2015). Para cada año de estudio se tenían los dos conjuntos de matrices: ambiente ($P=7$) y especies ($Q=23$), considerando como tercera dimensión las localidades de muestreo ($K=4$), mientras que en sus filas ($N=14$) los valores promedio de las variables para cada una de las siete profundidades estándar según la época climática (seca y lluviosa). La estructura de los datos para cada año fue similar al esquema presentado en la Fig. 3.5. Tras calcular el cubo de matrices de Co-IA entre cada par de matrices especie-

ambiente, se construyó la matriz compromiso que recogía la parte estable de la relación especie-ambiente y mediante un análisis con el HJ-Biplot se estudió esta relación calculando índices de calidad de representación. Se halló que durante las condiciones cálidas extremas (2015) las correlaciones entre las variables ambientales eran más fuertes. Además, los índices de calidad de representación fueron más altos que en los años anteriores y se logró identificar de forma más notoria las preferencias ambientales de las especies. El soporte de este resultado se encuentra en un manuscrito en correcciones para ser sometido en una revista, ver Anexo A2.

Se desarrollaron varios análisis que fueron presentados en eventos científicos. Para estos propósitos se consideraron diferentes perspectivas para la estructura de los conjuntos de matrices.

3.2.4 Aplicación del HJ-STATICO.

Basado en los resultados de los análisis descritos en la sección 3.2.3, se opta por investigar los datos del evento cálido extremo (2015) por manifestar una mayor correlación entre los parámetros ambientales y una relación especie-ambiente más definida.

En la Fig. 3.6 se presenta la estructura de los datos que se utilizó para este análisis, fueron organizados de forma similar a la estructura expuesta en la Fig. 3.5, registrándose en las filas (N=11) el promedio mensual de los primeros 20 m de profundidad donde, como hemos apuntado con anterioridad, se registró la mayor co-estructura (resultados de la aplicación STATICO) (ver Anexos A3, A4 y A5).

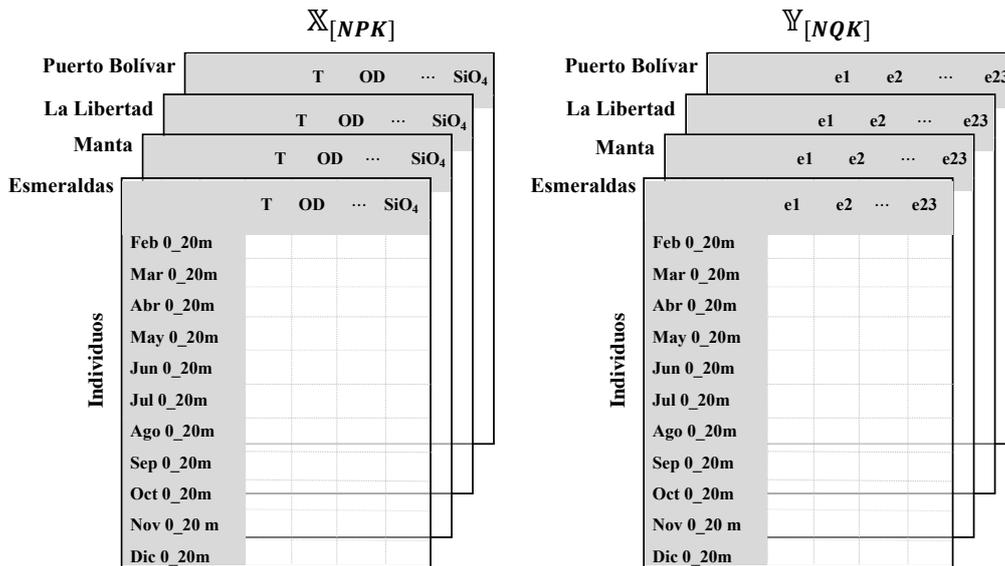


Figura 3.6 Estructura de las matrices de datos de tres vías para la aplicación del método HJ-STATICO, N=11 individuos, P=7 variables ambientales y Q=23 especies de fitoplancton.

A continuación, se presentan los resultados obtenidos con la herramienta informática que se ha elaborado en este trabajo para realizar un HJ-STATICO:

Las estaciones de muestreo que presentan una mayor similaridad ($RV=0.47$) en el patrón de comportamiento especie-ambiente son las dos estaciones oceánicas ubicadas en el centro del perfil costero (La Libertad y Manta), mientras que las dos estaciones estuarinas localizadas en los extremos del perfil costero presentaron las más notables diferencias (similaridad nula) (Tabla 3.2). Los resultados obtenidos se pueden explicar por el hecho de que en este año 2015, que fue un año cálido extremo, la estación de Esmeraldas (norte) presentó las temperaturas promedio más altas ($26.85\text{ }^{\circ}\text{C}$) y los nutrientes más bajos en todo el perfil costero, mientras que en Pto. Bolívar (sur) se evidenció un comportamiento opuesto (ver Anexo A6).

Tabla 3.2 Correlación Vectorial (RV).

	Esmeraldas	La Libertad	Manta	Pto. Bolívar
Esmeraldas	1	0.08	0.33	-0.01
La Libertad		1	0.47	0.41
Manta			1	0.47
Pto. Bolivar				1

En la Tabla 3.3 se presenta el porcentaje de inercia explicada y de la inercia explicada acumulada para la representación gráfica del plano bi-factorial de la inter-estructura (1) y el compromiso (2). Se observa que con los dos primeros factores se consigue explicar más del 75% de la variabilidad para la inter-estructura y más del 90% para el compromiso.

Tabla 3.3 Porcentaje de Varianza explicada. (1) "Inter-estructura" y (2) "Compromiso"

(1)

No.	Valores Propios	% Inercia	% Acum Inercia
1 F1	1.96	49.08	49.08
2 F2	1.05	26.16	75.23
3 F3	0.59	14.71	89.94
4 F4	0.40	10.06	100.00

(2)

No.	Valores Propios	% Inercia	% Acum Inercia
1 F1	0.29	88.42	88.42
2 F2	0.02	4.96	93.39
3 F3	0.01	3.58	96.97
4 F4	0.01	1.61	98.57
5 F5	0.00	0.80	99.38
6 F6	0.00	0.39	99.77
7 F7	0.00	0.24	100.00

En la tabla (1) se presenta la información de los valores propios (eigenvalues), % de Inercia y % Acumulado de la Inercia de los factores, que son el resultado de aplicar la DVS en la matriz RV formada por las $K = 4$ matrices de la inter-estructura. Mientras que, en la tabla (2) se obtiene la misma información para las $P = 7$ variables columna que conforman la matriz compromiso.

En la Tabla 3.4 se observa que la estación de Manta contribuye más a la construcción del compromiso por presentar el peso más alto (0.32) y además, es el lugar de muestreo mejor representado por el compromiso al tener el valor del \cos^2 más alto (0.88).

Tabla 3.4 Índices de las K-matrices.

Tablas	Pesos	Cos²
1 Esmeraldas	0.13	0.35
2 La Libertad	0.28	0.75
3 Manta	0.32	0.88
4 Pto. Bolívar	0.27	0.72

En la Fig. 3.7A se observa que existe una mayor similaridad en el comportamiento de la relación especie-ambiente en las dos estaciones localizadas al sur (La Libertad y Pto. Bolívar), mientras que Esmeraldas no se correlaciona con ninguna otra estación de muestreo, manifestando una relación especie-ambiente característica del calentamiento y de la escasez de nutrientes de sus aguas. Además, las condiciones físico-químicas del agua en la estación de Esmeraldas, presentaron una co-estructura más fuerte con las especies (vector más largo) debido a la mayor abundancia de especies indicadoras de aguas cálidas como los dinoflagelados *Gymnodinium* sp. (e21) y *Gyrodinium* sp. (e22) y la diatomea *Proboscia alata* (e10) que destacó por su gran abundancia en esta estación. Además, han sido características de esta región, otras especies del grupo de diatomeas como, *Guinardia striata* (e8), *Thalassiosira* sp. (e14), *Nitzschia* sp. (e15) y *Nitzschia longissima* (e16) (ver Anexo A6).

En la Fig. 3.7B se presentan de manera gráfica los resultados detallados en la Tabla 3.4, en el que Manta (estación oceánica) presenta los valores más altos para los índices del peso (0.32) y del \cos^2 (0.88), seguido de La Libertad (estación oceánica) y posteriormente Pto. Bolívar, quedando Esmeraldas como la estación con los índices más bajos en el análisis. El valor del peso indica que Manta es la estación que más aporta con su información a la construcción del compromiso,

mientras que, el valor del \cos^2 más alto para esta estación indica que la relación especie-ambiente que ocurrió en Manta durante la presencia del evento cálido es la mejor representada por el compromiso. Manta junto con La Libertad tienen una directa influencia oceánica, y además, reciben el efecto de la presencia del Frente Ecuatorial, que es una banda que separa las aguas frías y más salinas provenientes de la corriente del Perú, de las aguas cálidas y menos salinas del norte. El frente Ecuatorial se caracteriza por tener una alta productividad planctónica.

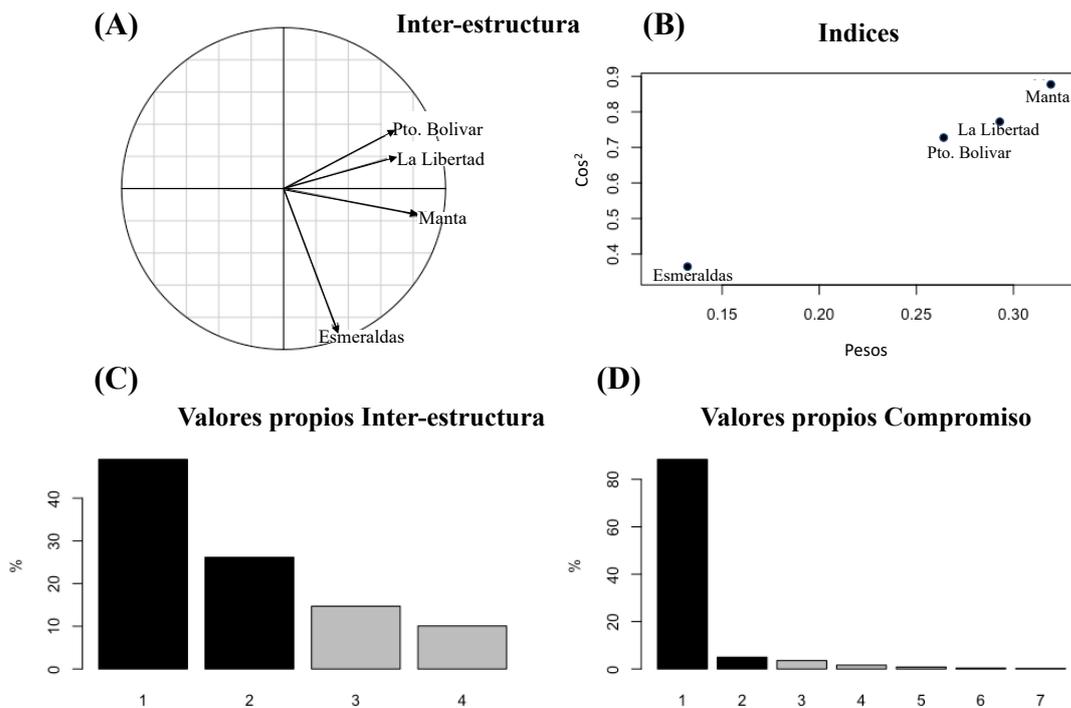


Figura 3.7 A) Representación gráfica de la Inter-estructura, B) índices de pesos y \cos^2 para el compromiso, C) explicación de la varianza para el mapa factorial de la inter-estructura y D) para el compromiso.

A continuación, se presentan los resultados de los análisis del compromiso y la intra-estructura desde dos perspectivas: la primera opción que utiliza la interpretación de un ACP y la segunda opción opta por utilizar la interpretación del HJ-Biplot con el cálculo de sus índices.

Opción 1: Gráficos del Compromiso e Intra-estructura utilizando ACP.

La Fig. 3.8A muestra que el eje 1 discrimina las variables de los nutrientes de las variables de T y OD, manifestando que las variables temperatura y OD tienen una correlación positiva entre ellas, e inversa con los nutrientes y la salinidad; mientras que para el caso de las especies se observa que la alta abundancia de los dinoflagelados *Gymnodinium* sp. (e21) y *Gyrodinium* sp. (e22) y el ciliado *Mesodinium rubrum* (e23) se asocia con valores por encima del valor promedio de temperatura y por debajo del valor promedio en silicatos y nitratos principalmente (Fig. 3.8B).

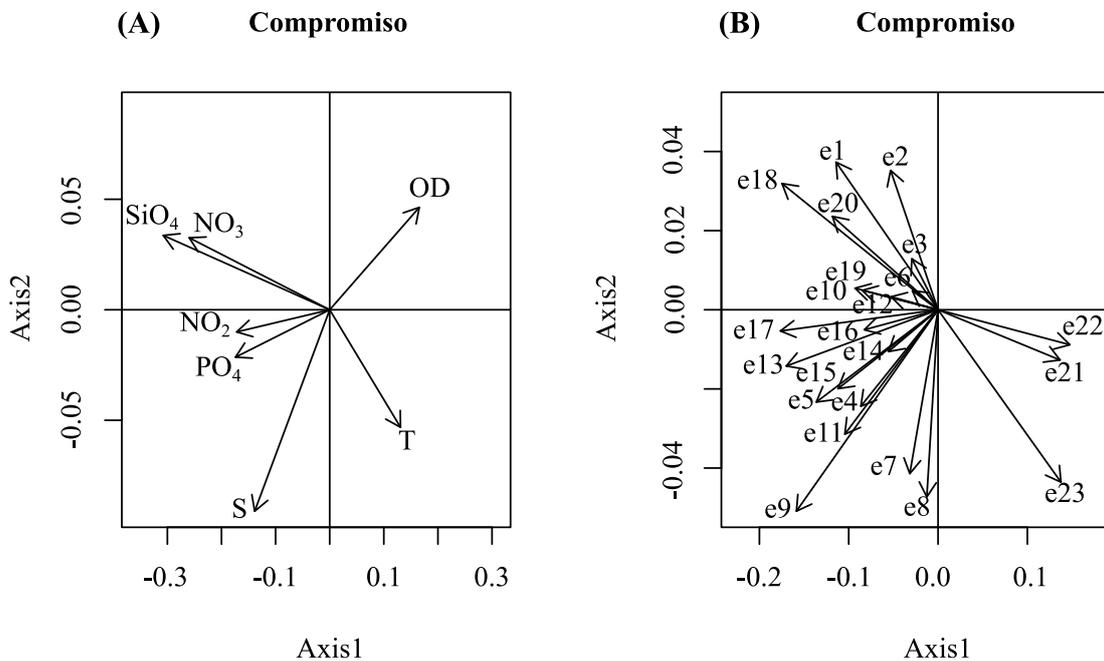


Figura 3.8 Mapa bi-factorial del Compromiso A) variables ambientales y B) variables especies. Consultar la Tabla 3.1 para los códigos de las especies.

En la Fig. 3.9 se observa la información de cada estación proyectada sobre el plano bi-factorial que se obtuvo para el compromiso. En términos generales para cada estación, la temperatura muestra una correlación negativa (vector con dirección opuesta) con alguno de los nutrientes (varía según la estación). Además, en cada

estación se pueden distinguir las especies que destacan en abundancia con las concentraciones ambientales propias del lugar.

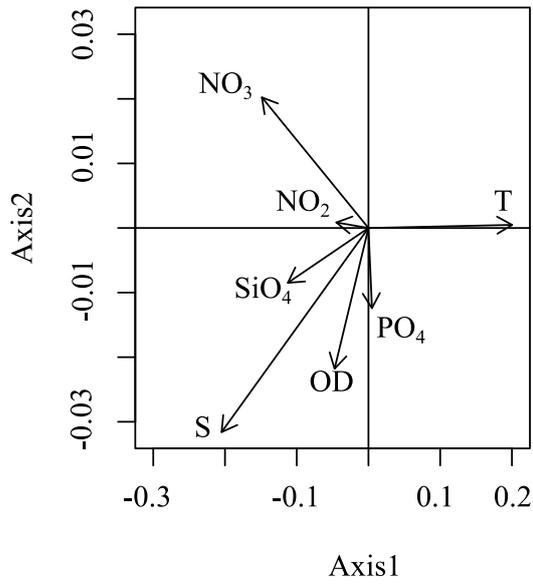
En Esmeraldas, el gradiente de temperatura tiene una correlación inversa con el nitrito, nitrato y silicato, mientras que los nutrientes muestran una correlación directa con la salinidad y el oxígeno (Fig. 3.9A). Las especies (Fig. 3.9B) mostraron dos agrupaciones claramente identificadas, el primer grupo de especies presentan valores más altos de abundancia asociados a valores por encima del valor promedio de temperatura y por debajo del valor promedio en salinidad, oxígeno y silicato principalmente. En este grupo se destacan las especies *Chaetoceros affinis* (e1), *Hemiaulus sinensis* (e4), *P. alata* (e10), *Skeletonema costatum* (e12) y *G. striata* (e8); mientras que el segundo grupo presentó mayor abundancia para valores por encima del valor promedio en silicato, salinidad, oxígeno y fósforo (eventos ambientales opuestos al primer grupo). En este grupo se encuentran las especies *N. longissima* (e16), *Pseudo-nitzschia pungens* (e17), *Lauderia borealis* (e13), *Nitzschia* sp. (e15) y *Rhizosolenia imbricata* (e9). Se observa además que los dinoflagelados *Gymnodinium* sp. (e21) y *Gyrodinium* sp. (e22) mostraron su preferencia por valores por encima del promedio en temperatura.

En Manta, se observó un comportamiento similar a lo ocurrido en Esmeraldas, en cuanto a las variables ambientales (Fig. 3.9C), con la excepción del oxígeno disuelto que presentó una correlación inversa con la salinidad y el fósforo. En cuanto a las especies (Fig. 3.9D) se evidenció que las únicas que mostraron preferencias por los valores por encima del valor promedio en temperatura fueron los dinoflagelados *Gymnodinium* sp. (e21) y *Gyrodinium* sp. (e22) y el ciliado *M. rubrum* (e23). El resto de las especies presentaron su mayor abundancia para valores por debajo del valor promedio en temperatura como, por ejemplo, *C. affinis* (e1), *Chaetoceros curvisetus* (e2), *P. pungens* (e17) y *Navicula* sp. (e18) y otras como *R. imbricata* (e9) y *Dactyliosolen fragilissimus* (e7), destacaron por ser más abundantes cuando los valores de fósforo y salinidad estaban por encima de la media y el oxígeno por debajo del valor promedio.

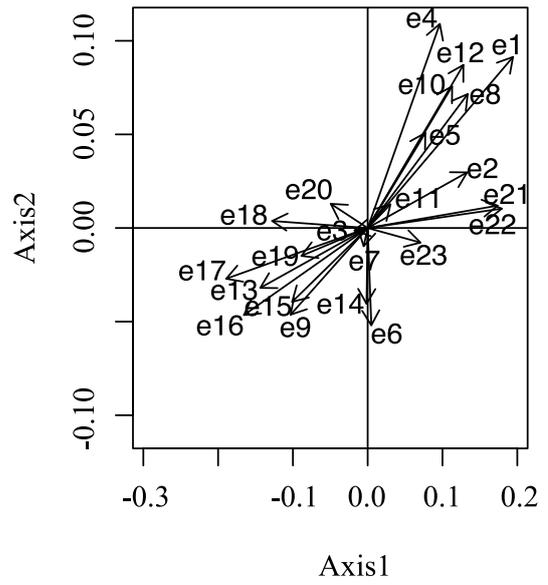
En La Libertad (Fig. 3.9E), la temperatura tuvo una correlación inversa con el nitrato, y el oxígeno mostró una correlación inversa con la salinidad. En esta estación oceánica, las especies (Fig. 3.9F) que destacaron por su alta abundancia fueron *H. sinensis* (e4), *Leptocylindrus danicus* (e5), *R. imbricata* (e9), *Dactyliosolen antarcticus* (e11) y *P. pungens* (e17) que mostraron preferencias por valores por encima del valor promedio en salinidad y temperatura, pero bajo en cuanto a oxígeno.

En Pto. Bolívar (Fig. 3.9G) la temperatura presentó una alta correlación directa con la salinidad, el oxígeno y el nitrito, pero inversa con el fosfato, nitrato y silicato. En esta estación estuarina localizada al sur y bajo la influencia de la corriente fría de Humboldt, el grado de correlación entre los parámetros ambientales es mayor, al ponerse de manifiesto dos evidentes agrupaciones que se proyectan en direcciones opuestas. Para el caso de las especies (Fig. 3.9H), se observa que la gran mayoría presentan mayores valores de abundancia para valores por encima del valor promedio en silicato, nitrato y fosfato, pero con valores por debajo del promedio en temperatura, salinidad y oxígeno. En este grupo se destacaron las especies *Navicula* sp. (e18), *Ditylum brightwellii* (e6), *P. pungens* (e17), *S. costatum* (e12) y *C. affinis* (e1). Además, se observa una evidente asociación entre las especies de dinoflagelados *Gymnodinium* sp. (e21), *Gyrodinium* sp. (e22) y el ciliado *M. rubrum* (e23) que mostraron su preferencia hacia valores por encima del promedio en temperatura, salinidad y oxígeno.

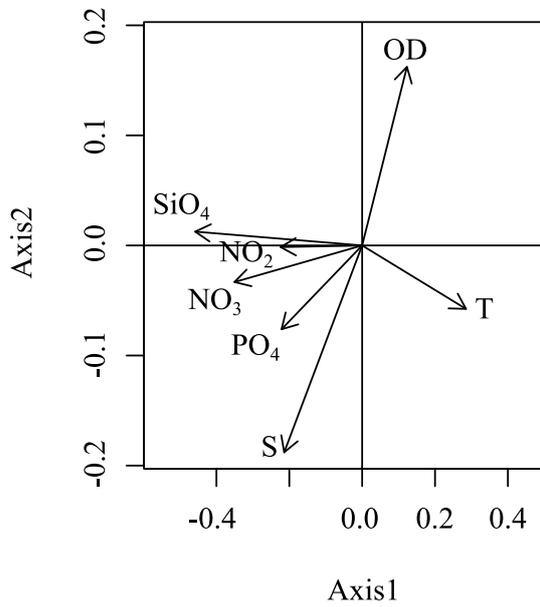
(A) Esmeraldas



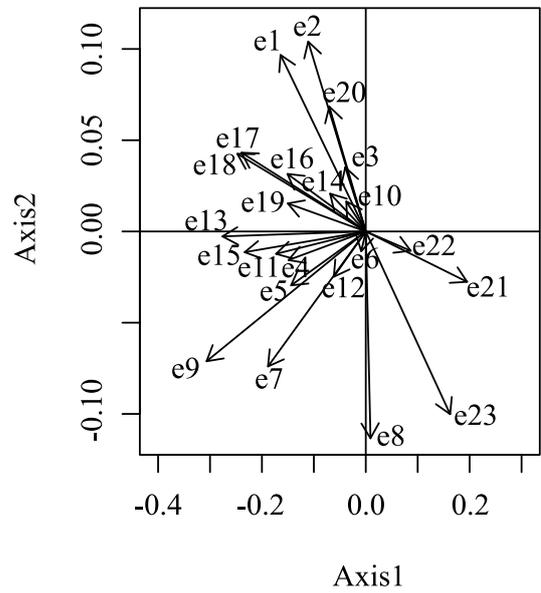
(B) Esmeraldas



(C) Manta



(D) Manta



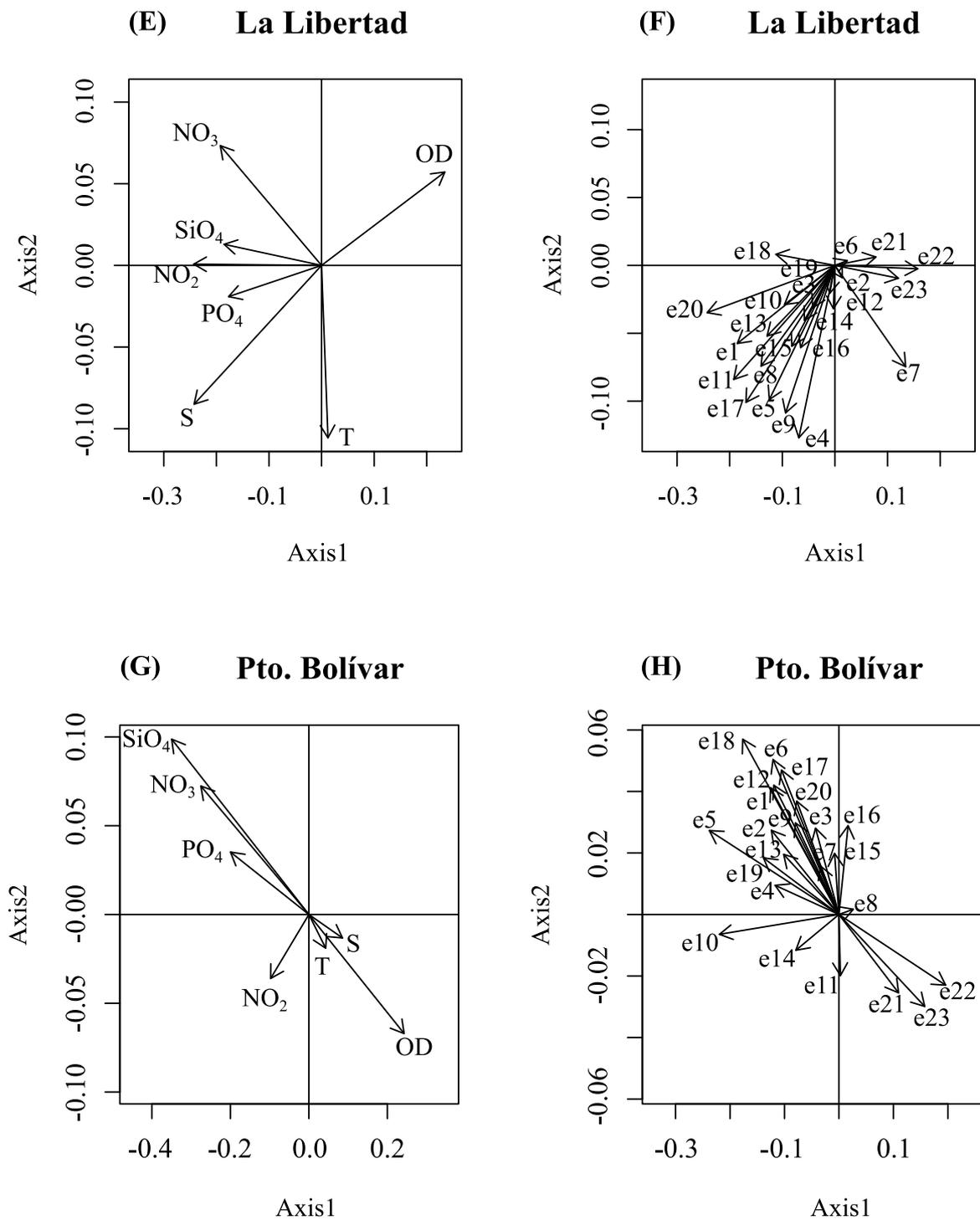


Figura 3.9 Mapa bi-factorial de la Intra-estructura para Esmeraldas (A- ambiente y B - especies), Manta (C - ambiente y D - especies), La Libertad (E - ambiente y F - especies) y Pto. Bolívar (G - ambiente y H - especies). Consultar la Tabla 3.1 para los códigos de las especies.

Opción 2: Gráficos del Compromiso e Intra-estructura aplicando el HJ-Biplot.

Los resultados del análisis anterior (Fig. 3.8) pusieron de manifiesto que el eje 1 discriminaba las variables en dos grupos. Sin embargo, en la Fig. 3.10 se puede determinar que el eje 1 por si solo, está recogiendo casi la totalidad de la variabilidad para el nitrato, fosfato y silicato ($CRFLE_j \geq 0.926$), mientras que, para el caso de las especies, mostró una alta explicación de la variación natural de la abundancia de *L. borealis* (e13), *P. pungens* (e17), *Gymnodinium* sp. (e21) y *Gyrodinium* sp. (e22) ($CRFLE_i \geq 0.962$). El eje 2 aportó en menor cantidad a la explicación de la variabilidad, destacándose principalmente en representar a la salinidad ($CRFLE_j = 0.269$) y a las especies *D. fragilissimus* (e7) y *G. striata* (e8) ($CRFLE_i \geq 0.403$). Las variables que más aportaron a la explicación de la variabilidad total de las variaciones ambientales, que ocurrieron en todo el perfil costero, fueron el silicato ($CRT_j = 0.3$) y el nitrato ($CRT_j = 0.214$), entre las dos explicaron más del 0.514 de la variabilidad total. Las especies que más aportaron a la explicación de la variabilidad de la abundancia fueron *P. pungens* (e17) ($CRT_i = 0.1$) y *Navicula* sp. (e18) ($CRT_i = 0.1$). Los dos ejes logran explicar una alta proporción de la variabilidad para cada una de las variables ambientales ($QLRE_j \geq 0.836$). Los indicadores del HJ-Biplot se pueden revisar en el Anexo A7.

En la Fig. 3.10 se observa que las especies que reflejan una mayor calidad de representación ($QLRE_i \geq 0.976$) y que son representadas en el gráfico con etiquetas de mayor tamaño, son aquellas que presentaron una co-estructura más fuerte con las variaciones ambientales como, por ejemplo, las diatomeas céntricas *L. borealis* (e13) y *R. imbricata* (e9), la pennada *Navicula* sp. (e18), el dinoflagelado *Gymnodinium* sp. (e21) y el ciliado *M. rubrum* (e23). Durante este evento cálido extremo destacaron, por su mayor abundancia, las especies *Gymnodinium* sp. (e21), *Gyrodinium* sp. (e22) y *M. rubrum* (e23) cuando ocurrieron valores por encima del valor promedio de temperatura y oxígeno, pero con valores por debajo del valor promedio en nitrato y silicato. Del mismo modo

la mayor abundancia de las diatomeas pennadas *Navicula* sp. (e18), *Thalassionema nitzschioides* (e20) y *P. pungens* (e17) exhibieron una fuerte co-estructura con valores por encima del valor promedio de nitrato y silicato, pero por debajo del valor promedio en temperatura y oxígeno. Las especies *D. brightwellii* (e6) y *S. costatum* (e12) con baja calidad de representación ($QLRE_i \leq 0.402$) son aquellas que mostraron una co-estructura débil y se representan en el gráfico con una etiqueta de menor tamaño. Los indicadores del HJ-Biplot se pueden consultar en el Anexo A7.

Compromiso: HJBiplot

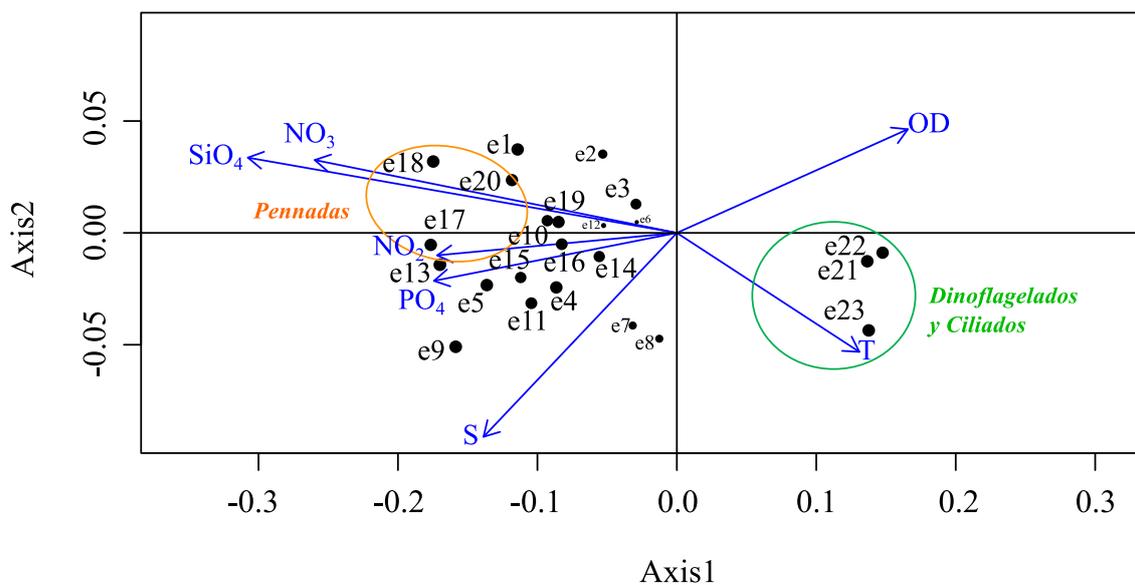


Figura 3.10 Mapa bi-factorial del Compromiso aplicando HJ-Biplot.

La Fig. 3.8 puso de manifiesto que la especie *Guinardia striata* (e8) asoció su mayor abundancia con valores por encima del valor promedio en salinidad y por debajo del valor promedio en oxígeno en los primeros 20m de profundidad. Sin embargo, con el análisis HJ-Biplot para los dos mismos ejes, se complementa este análisis al revelar que esa relación fue débil (e8; $QLRE_i = 0.632$) al compararla con la de otras especies como, por ejemplo, la especie *R. imbricata* (e9; $QLRE_i = 0.987$) la cual manifestó una relación más fuerte con las variaciones ambientales

generadas por el evento cálido extremo (Fig. 3.10). Los indicadores del HJ-Biplot se pueden consultar en el Anexo A7.

Para el caso de la estación norte de Esmeraldas, se observa que durante el evento cálido extremo la co-estructura más fuerte ($QLRE_i \geq 0.854$) ocurrió con las diatomeas céntricas *H. sinensis* (e4), *G. striata* (e8), *P. alata* (e10) y *S. costatum* (e12) cuando se registraron valores por encima del valor promedio de temperatura y por debajo del valor promedio del silicato, nitrito, oxígeno y salinidad; mientras que para las variables ambientales fue para el nitrato ($QLRE_i = 0.773$) (Fig. 3.11A).

En Manta durante el evento cálido extremo la co-estructura más fuerte ($QLRE_i \geq 0.96$) ocurrió con las diatomeas céntricas *D. fragilissimus* (e7), *R. imbricata* (e9), *L. borealis* (e13) y las pennadas *P. pungens* (e17) y *Navicula* sp. (e18) cuando se registraron valores por encima del valor promedio en nutrientes y bajos en temperatura y oxígeno; mientras que para el dinoflagelado *Gymnodinium* sp. (e21) y el ciliado *M. rubrum* (e23) ocurrió cuando prefirieron valores por encima del valor promedio en temperatura y por debajo del valor promedio en silicato y nitrito para incrementar su abundancia ($QLRE_i \geq 0.838$). Para el caso de las variables ambientales se destacó la explicación de la variabilidad de los nutrientes ($QLRE_j \geq 0.908$) (Fig. 3.11B).

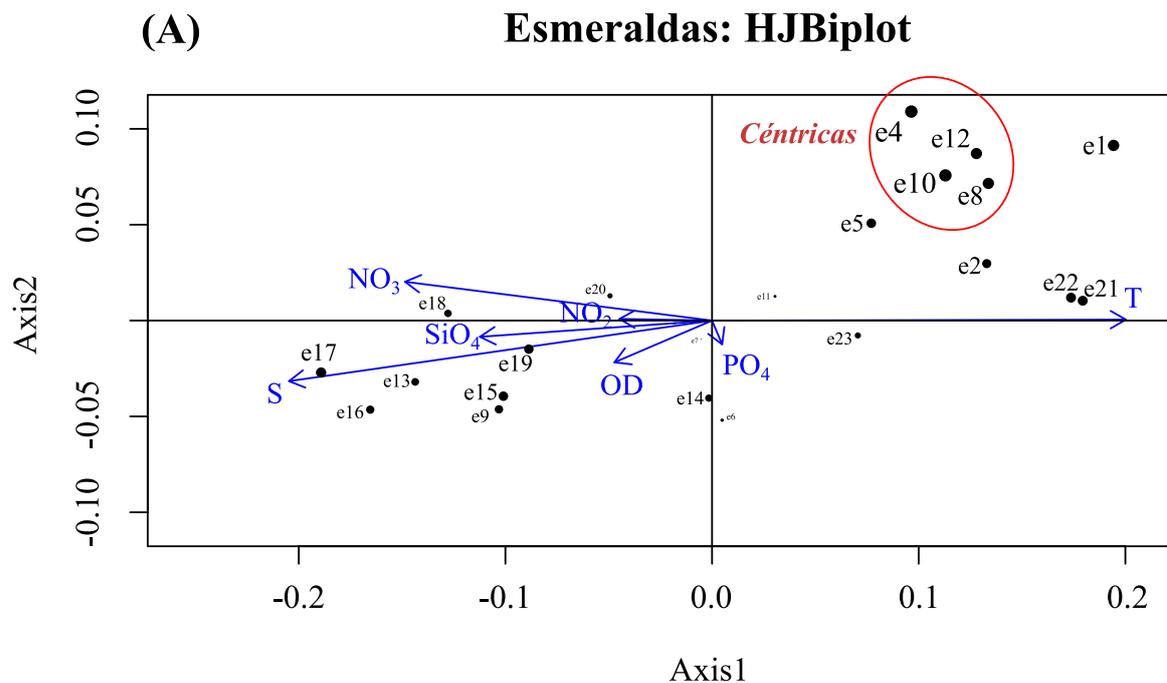
En La Libertad durante el evento cálido extremo la co-estructura más fuerte ($QLRE_i \geq 0.92$) ocurrió con las diatomeas céntricas *C. affinis* (e1), *H. sinensis* (e4), *L. danicus* (e5), *G. striata* (e8), *D. antarcticus* (e11) y *L. borealis* (e13) cuando se registraron valores por encima del valor promedio en salinidad y fosfato y por debajo del valor promedio en oxígeno; mientras que para los dinoflagelados *Gymnodinium* sp. (e21) y *Gyrodinium* sp. (e22) y el ciliado *M. rubrum* (e23) ocurrió cuando prefirieron valores por encima del valor promedio en oxígeno y por debajo del valor promedio en salinidad y nutrientes para incrementar su abundancia ($QLRE_i \geq 0.785$). Para el caso de las variables se destacó la

explicación de la variabilidad del silicato y el oxígeno ($QLRE_j \geq 0.825$) (Fig. 3.11C).

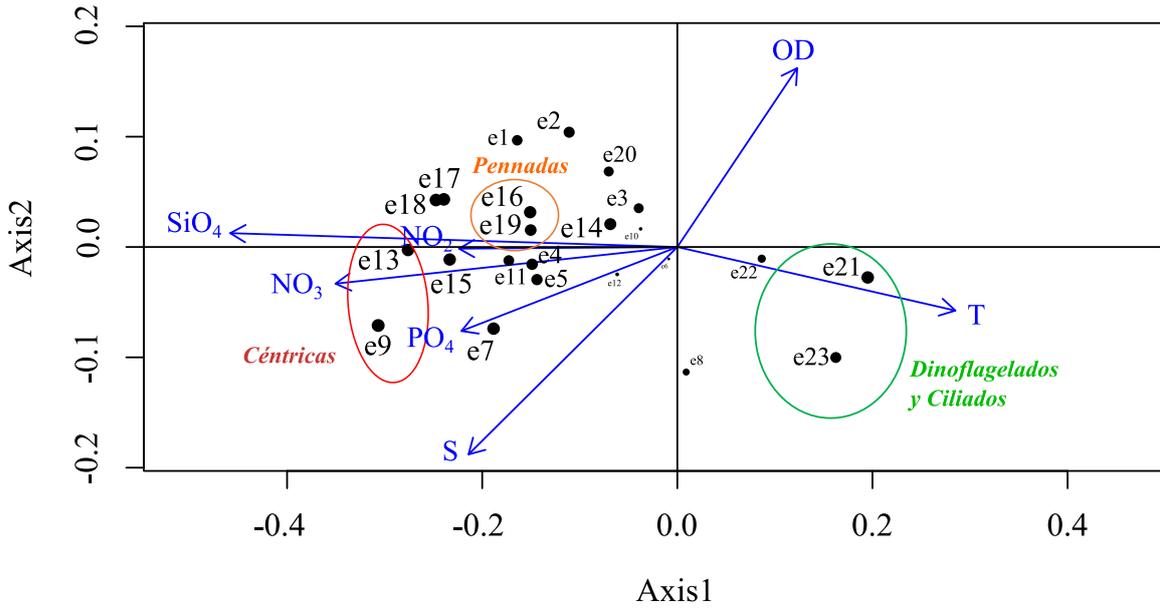
En Pto. Bolívar durante el evento cálido extremo la co-estructura más fuerte ($QLRE_i \geq 0.9$) ocurrió con la diatomea céntrica *L. borealis* (e13) y las pennadas *Navicula sp.* (e18) y *S. membranacea* (e19) cuando se registraron valores por encima del valor promedio en nitrato, silicato y fósforo pero por debajo del valor promedio en temperatura, salinidad y oxígeno ($QLRE_i \geq 0.866$). Para el caso de las variables ambientales se destacó la explicación de la variabilidad de los nutrientes y el oxígeno ($QLRE_j \geq 0.846$) (Fig. 3.10D).

Para las cuatro estaciones, el eje 1 es el que explica la mayor parte de la variabilidad para los nutrientes y para las especies se enfatiza la explicación de la variabilidad en la abundancia del dinoflagelado (e21).

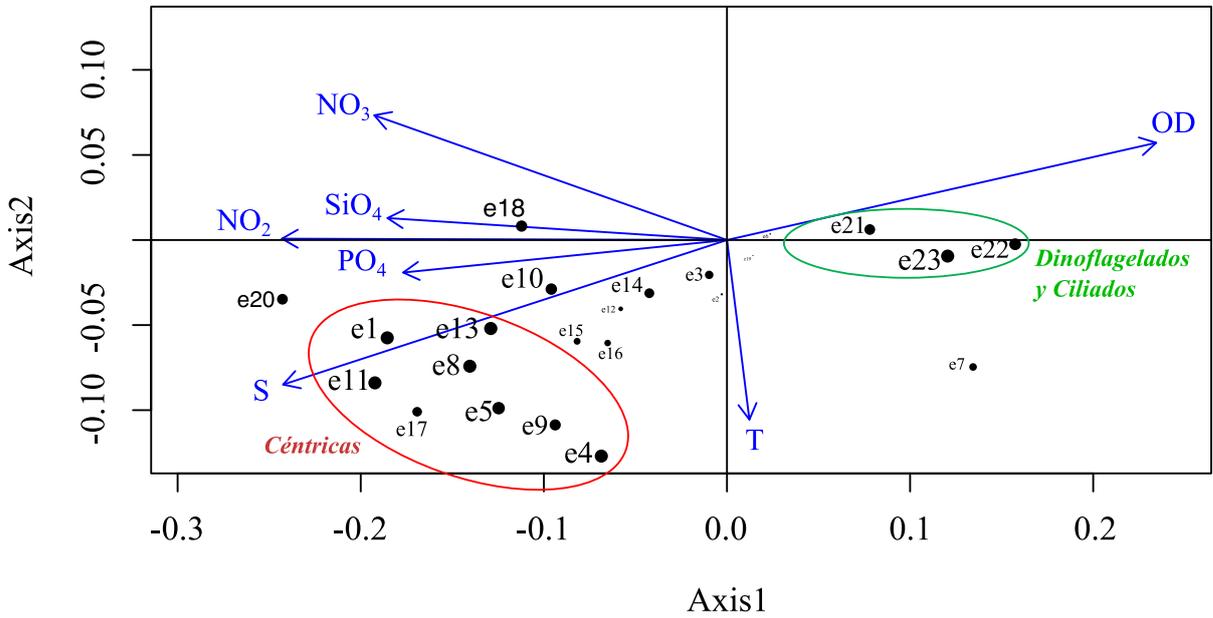
Los indicadores del HJ-Biplot para la intra-estructura se pueden consultar en el Anexo A8.



(B) Manta: HJBiplot



(C) La Libertad: HJBiplot



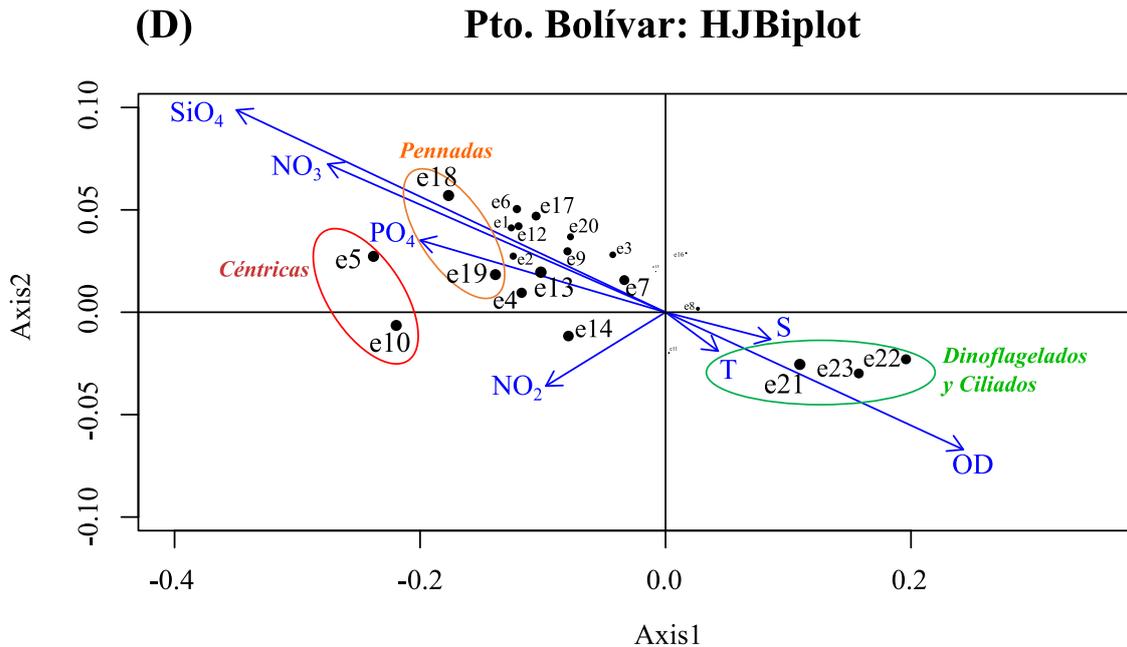


Figura 3.11 Mapa bi-factorial de la Intra-estructura aplicando HJ-Biplot para la estación de A) Esmeraldas, B) Manta, C) La Libertad y D) Pto. Bolívar.

Consultar la Tabla 3.1 para los códigos de las especies.

El método HJ-STATICO propuesto en este trabajo, ha permitido complementar el análisis que ofrece el método STATICO, al mostrar, además de las similitudes entre las K-matrices (inter-estructura), la relación especie-ambiente que puso de manifiesto las preferencias ambientales de las especies en toda la región de estudio (compromiso) y en cada estación de muestreo (intra-estructura). También nos permitió identificar de manera gráfica y cuantificar mediante índices, las variables ambientales y las especies que más información aportaron al análisis, así como también, identificar las variables ambientales y las especies que estaban mejor representadas y que nos proporcionan mayor información acerca del evento ocurrido. Adicionalmente, contribuyó a constatar la fuerte relación especies-ambiente durante la presencia de un evento cálido extremo.

CAPÍTULO 4

*Pon todas tus
preocupaciones
y ansiedades en
las manos de Dios,
porque él cuida de tí*
1 Pedro 5:7

CAPÍTULO 4: MÉTODO PARA ANALIZAR LAS RELACIONES DE DOS CONJUNTOS DE MATRICES DE DATOS MIXTOS: MixSTATICO

Las ventajas y aplicaciones del método STATIS y de cada una de sus vertientes son numerosas. Sin embargo, el análisis de la relación entre dos estructuras de matrices con una combinación de variables cuantitativas y cualitativas es un problema particular no cubierto por los métodos descritos hasta ahora en la literatura.

En muchas ocasiones, en estudios biológicos o de salud se debe analizar la interacción entre el ambiente y las unidades biológicas. Un caso muy frecuente es cuando se estudian parámetros abióticos como el clima y se desea analizar su relación o efecto sobre los parámetros bióticos tales como las categorías de presencia o ausencia de algún organismo. Ninguna de las anteriores técnicas de la familia STATIS se ajusta a este tipo de análisis.

En este capítulo, se presenta la propuesta metodológica de un nuevo método estadístico multivariante, que hemos denominado MixSTATICO, que permite obtener la parte estable de la co-estructura de dos matrices de tres vías que contienen datos cuantitativos y cualitativos (es decir, datos mixtos). Para este propósito nos apoyaremos en los procedimientos metodológicos de dos técnicas multivariantes de tres vías: 1) el DISTATIS ([Abdi et al., 2007](#)) que permite analizar un conjunto de matrices de distancias que representan la relación de los dos cubos de datos mixtos y 2) el STATICO ([Simier et al., 1999](#)) para calcular la inter-estructura, el compromiso y la intra-estructura de los datos. Adicionalmente,

se incorpora una variante en los pesos para la obtención de la matriz compromiso propuesta por [Abdi et al. \(2012\)](#). Esta técnica se encuentra publicada en la revista científica *Sustainability* 2021, 13(11) (<https://doi.org/10.3390/su13115924>), ver **Contribuciones Científicas**.

4.1 Descripción del método MixSTATICO.

Los datos de partida se organizan en dos matrices de datos de tres vías $\mathbb{X}_{[NPK]}$ y $\mathbb{Y}_{[NQK]}$, que están estructuradas con datos mixtos que pueden ser cuantitativos (discretos, o continuos) y/o cualitativos (binarios, ordinales, o nominales). Cada conjunto de matrices debe estar formada por K -matrices cada una de las cuales tiene en sus columnas las mismas variables mientras que en sus filas pueden estar los mismos o diferentes individuos. Entre pares de matrices se deben tener los mismos individuos, tal que:

$$\mathbb{X}_{[NPK]} = [\mathbf{X}_{[NP,1]} | \mathbf{X}_{[NP,2]} | \dots | \mathbf{X}_{[NP,K]}], \quad (4.1)$$

$$\mathbb{Y}_{[NQK]} = [\mathbf{Y}_{[NQ,1]} | \mathbf{Y}_{[NQ,2]} | \dots | \mathbf{Y}_{[NQ,K]}]. \quad (4.2)$$

En primer lugar se debe realizar un preprocesamiento de los datos $\mathbb{X}_{[NPK]}$ y $\mathbb{Y}_{[NQK]}$ de la siguiente forma:

- Las variables cuantitativas deben ser transformadas aplicando técnicas como por ejemplo: logaritmo, estandarización por columnas, centrado por columnas como lo sugieren [Legendre & Legendre \(1979\)](#) y [Kroonenberg \(2008\)](#);
- Las variables binarias deben ser codificadas con 0 y 1;
- Las variables ordinales deben ser codificadas de acuerdo a la escala en orden ascendente;

- Las variables nominales deben ser transformadas a una matriz disyuntiva, en la que cada categoría es una nueva variable dicotómica.

Como resultado, se obtendrán nuevas matrices de datos de tres vías con dimensión incrementada P' y Q' por la creación de la matriz disyuntiva para los casos de las variables nominales, tal como se detalla a continuación:

$$\mathbb{X}'_{[NP',K]} = \left[\mathbf{X}'_{[NP',1]} \mid \mathbf{X}'_{[NP',2]} \mid \dots \mid \mathbf{X}'_{[NP',K]} \right] \quad (4.3)$$

$$\mathbb{Y}'_{[NQ',K]} = \left[\mathbf{Y}'_{[NQ',1]} \mid \mathbf{Y}'_{[NQ',2]} \mid \dots \mid \mathbf{Y}'_{[NQ',K]} \right] \quad (4.4)$$

Entre los métodos comúnmente utilizados para analizar las relaciones entre dos conjuntos de variables medidas sobre los mismos individuos se tienen el CCA (ter Braak, 1986), el Análisis de la Redundancia (RDA; Rao, 1964), el Análisis de Correlación Canónica (CANCOR; Hotelling, 1936) y el análisis de la Co-Inercia (Co-IA; Dolédec & Chessel, 1994) que difiere del resto por su enfoque en maximizar la covarianza entre las variables de las dos matrices.

Para calcular la co-estructura en este nuevo método, se implementa la metodología basada en matrices de distancias de la técnica DISTATIS (Abdi et al., 2007). Para datos mixtos se aplica la medida de distancia de Gower (1971) entre las variables del par de matrices $\mathbf{X}'_{[NP',k]}$ y $\mathbf{Y}'_{[NQ',k]}$ que han sido previamente concatenadas por filas en cada k condición,

$$\mathbb{XY}_{[Nr',K]} = \left[\mathbf{XY}_{[Nr',1]} \mid \mathbf{XY}_{[Nr',2]} \mid \dots \mid \mathbf{XY}_{[Nr',K]} \right], \text{ donde } r' = P' + Q' \quad (4.5)$$

Posteriormente, se calcula de manera independiente la distancia de Gower d_{ij} para cada k-matriz:

$$d_{ij}^2 = 1 - s_{ij}, \quad (4.6)$$

tal que,

$$s_{ij} = \frac{\sum_{h=1}^{p_1} \left(1 - \frac{|xy_{ih} - xy_{jh}|}{G_h} \right) + a + \alpha}{p_1 + (p_2 - d) + p_3} \quad (4.7)$$

siendo,

s_{ij} el coeficiente de similaridad de Gower;

p_1 el número de variables cuantitativas continuas;

p_2 el número de variables binarias;

p_3 el número de variables cualitativas no-binarias;

a y d el número de coincidencias en las variables binarias (1,1) y (0,0), respectivamente;

α el número de coincidencias en variables cualitativas no-binarias;

G_h el rango de la h-ésima variable cuantitativa.

El análisis está basado solamente en las distancias entre las P variables y Q variables en las matrices $\mathbf{X}_{[NP,k]}$ y $\mathbf{Y}_{[NQ,k]}$. Por lo tanto, cuando al menos una de estas matrices es nominal, se calcula la distancia promedio de Gower entre las distancias de la variable no nominal y cada una de las categorías de la variable nominal que previamente se transformaron en variables dicotómicas, de la siguiente manera:

$$\bar{d}_{v,w} = \frac{\sum_{i=1}^{nc} d_{v,w_i}}{nc}, \quad (4.8)$$

tal que,

d es la distancia de Gower;

v representa la variable no-nominal;

w representa la variable nominal;

nc es el número de categorías de la variable nominal.

Si ambas variables son nominales, entonces se aplica el método del centroide, usando el promedio generalizado para todas las distancias.

Finalmente, se obtiene una nueva matriz de tres vías formada por medidas de distancias y de dimensión $(Q \times P \times K)$, que expresa la estructura en común de las dos matrices de tres vías iniciales, tal como sigue:

$$\mathbb{Z}_{[QPK]} = [\mathbf{Z}_{[QP,1]} | \mathbf{Z}_{[QP,2]} | , \dots , | \mathbf{Z}_{[QP,K]}] \quad (4.9)$$

Continuando con el enfoque adoptado por [Abdi et al. \(2007\)](#) en el DISTATIS, se procede a normalizar las matrices de distancia $\mathbf{Z}_{[QP,k]}$ utilizando un método similar al aplicado en el Análisis Factorial Múltiple (AFM; [Escofier & Pagès, 1998](#)) para obtener nuevas matrices de Co-IA ponderadas $\mathbf{Z}'_{[QP,k]}$, que son comparables en importancia a las matrices de distancia originales al relacionar las k-matrices en el plano bi-factorial.

Para realizar la normalización, se realiza la DVS de cada una de las k-matrices de distancia con rango L . Esta técnica permite descomponer matrices rectangulares $\mathbf{Z}_{[QP,k]}$ en factores $\mathbf{U}_{[QL]}$ y $\mathbf{V}_{[PL]}$. Las matrices contienen como columnas vectores singulares ortonormales asociados con los valores singulares γ_{ii} contenidos en la diagonal principal de la matriz $\Gamma_{[LL]}$. Las matrices $\mathbf{U}_{[QL]}$ y $\mathbf{V}_{[PL]}$ tienen la propiedad $(\mathbf{U}_{[QL,k]}^T \times \mathbf{U}_{[QL,k]}) = (\mathbf{V}_{[PL,k]}^T \times \mathbf{V}_{[PL,k]}) = \mathbf{I}$. Los vectores singulares de $\mathbf{U}_{[QL]}$ y $\mathbf{V}_{[PL]}$ son los vectores propios de las matrices $(\mathbf{Z}_{[QP,k]} \times \mathbf{Z}_{[QP,k]}^T)$ y

$(\mathbf{Z}_{[QP,k]}^T \times \mathbf{Z}_{[QP,k]})$ respectivamente, con valores propios $\lambda_{[k]} = \gamma_{[11,k]}^2$. Las matrices son normalizadas al multiplicar sus elementos por el escalar $\lambda_{[K]}^{-1}$.

Es decir;

$$DVS(\mathbf{Z}_{[QP,k]}) = U\Gamma V^T, \text{ donde } U \in \mathbb{R}^{Q \times L}, V \in \mathbb{R}^{P \times L}, \Gamma \in \mathbb{R}^{L \times L}, L = \text{Rango}(\mathbf{Z}_{[QP,K]}) \quad (4.10)$$

y la normalización de las k-matrices, tal como;

$$\mathbf{Z}'_{[QP,k]} = \lambda_{[k]}^{-1} \mathbf{Z}_{[QP,k]}. \quad (4.11)$$

Las $\mathbf{Z}'_{[QP,k]}$ forman un conjunto de K-matrices de distancia ponderadas,

$$\mathbb{Z}'_{[QPK]} = [\mathbf{Z}'_{[QP,1]} \mid \mathbf{Z}'_{[QP,2]} \mid \dots \mid \mathbf{Z}'_{[QP,K]}] \quad (4.12)$$

El siguiente paso, es calcular la matriz de productos cruzados $\mathbf{C}_{[KK]}$ conocida como la matriz de varianzas y covarianzas. Para lo cual se debe obtener primero la vectorización de las matrices $\mathbf{Z}'_{[QP,k]}$, como sigue:

$$\mathbb{Z}''_{[s'K]} = [\text{Vec}(\mathbf{Z}'_{[QP,1]}), \text{Vec}(\mathbf{Z}'_{[QP,2]}), \dots, \text{Vec}(\mathbf{Z}'_{[QP,K]})]^T; \text{ donde } s' = Q \times P \quad (4.13)$$

y posteriormente,

$$\mathbf{C}_{[KK]} = \mathbb{Z}''_{[s'K]} \times \mathbb{Z}''^T_{[s'K]} \quad (4.14)$$

Para calcular la inter-estructura, se utiliza la matriz obtenida en la Ec. (4.14) para obtener una matriz de productos escalares que contiene los coeficientes RV de Rayleigh (Yves Escoufier, 1973) conocida como la matriz de correlación vectorial:

$$\mathbf{RV}(k, k') = \frac{c_{[kk']}}{(c_{[kk]})^{1/2} (c_{[k'k']})^{1/2}} ; \text{ with } k = 1, \dots, K \text{ and } k' = 1, \dots, K \quad (4.15)$$

La matriz RV contiene las proximidades entre las K-matrices y su interpretación es similar al coeficiente de correlación de Pearson.

Abdi et al. (2012) aplican DVS a la matriz $\mathbf{C}_{[KK]}$ para analizar las similitudes entre las K-matrices porque cumple con los requisitos de simetría y es definida positiva. Sin embargo, Abdi et al. (2007) sugieren calcular los vectores y valores propios de la matriz $\mathbf{RV}_{[KK]}$ (método adoptado en esta investigación), tal que:

$$\mathbf{RV}_{[KK]} = \mathbf{U}\mathbf{\Xi}\mathbf{V}^T, \text{ con } \mathbf{U} \in \mathbb{R}^{K \times L}, \mathbf{V} \in \mathbb{R}^{K \times L}, \mathbf{\Xi} \in \mathbb{R}^{L \times L} \text{ y } L = \text{Rango}(\mathbf{RV}_{[KK]}) \quad (4.16)$$

Posterior a la descomposición de la matriz $\mathbf{RV}_{[KK]}$ y al cálculo de sus vectores y valores propios (método ACP), las K-matrices pueden ser proyectadas sobre el plano bi-factorial como puntos, permitiendo estudiar las similitudes entre ellas. Los elementos del primer vector propio tienen el mismo signo por ser una matriz semidefinida positiva.

A partir del primer vector propio de $\mathbf{U}_{[KK]}$ obtenemos el vector de pesos $\vec{\alpha}_{[K]}$, tal que $\sum_{k=1}^K \alpha_{[k]} = \mathbf{1}$:

$$\mathbf{U}_{[kk]} = \begin{bmatrix} u_{11} & \dots & u_{1K} \\ \vdots & \ddots & \vdots \\ u_{K1} & \dots & u_{KK} \end{bmatrix} \Rightarrow \begin{pmatrix} u_{11} \\ \vdots \\ u_{K1} \end{pmatrix} = \vec{u}_{[K,1]} \quad (4.17)$$

$$\vec{\alpha}_{[K]} = (\vec{u}_{[K,1]}^T \times \mathbb{1}_{[K]})^{-1} \vec{u}_{[K,1]} \quad (4.18)$$

$$\vec{\alpha}_{[K]} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_K \end{pmatrix} \Rightarrow \sum_{k=1}^K \alpha_k = 1 \quad (4.19)$$

Abdi et al. (2012) mencionó que las K-matrices de similaridades son representadas en el primer vector propio $\vec{u}_{[K,1]}$ que otorga mayor importancia a las matrices que mejor representan la Co-IA.

Para representar gráficamente la inter-estructura, se calculan las coordenadas de las k-matrices utilizando los dos primeros vectores propios de $\mathbf{U}_{[KK]}$. La primera columna debe ser positiva (ver Abdi et al. (2012)).

$$G = \mathbf{U}_{[K2]} \times \mathbf{D}_{[22]}^{1/2}, \text{ donde } \mathbf{D}_{[LL]} = \text{Diag}(\Xi) \quad (4.20)$$

Los valores de $\vec{\alpha}_{[K]}$, se utilizan como pesos para calcular la matriz compromiso, que representa la parte estable de la imagen promedio de las relaciones entre las matrices pareadas (Co-IA).

$$\mathbf{W}_{[QP]} = \sum_{k=1}^K \alpha_{[k]} \mathbf{Z}'_{[QP,k]} \quad (4.21)$$

La matriz compromiso $\mathbf{W}_{[QP]}$ es analizada utilizando un Análisis de Coordenadas Principales (PcoA) con doble centrado (Gower, 1966). Para representar gráficamente su configuración en un espacio de baja dimensión es necesario realizar una DVS:

$$\mathbf{W}_{[QP]} = \mathbf{U}'\mathbf{\Xi}'\mathbf{V}'^T, \text{ donde } \mathbf{U}' \in \mathbb{R}^{Q \times L}, \mathbf{V}' \in \mathbb{R}^{P \times L}, \mathbf{\Xi}' \in \mathbb{R}^{L \times L} \text{ y } L = \text{Rango}(\mathbf{W}_{[QP]}) \quad (4.22)$$

Para generar el gráfico de la matriz compromiso, se calculan las puntuaciones $\mathbf{F}_{[QL]}$ para las variables filas y las cargas $\mathbf{Q}_{[PL]}$ para las variables columnas:

$$\mathbf{F}_{[QL]} = \mathbf{U}'_{[QL]} \times \mathbf{D}'_{[LL]}^{1/2} \text{ y } \mathbf{Q}_{[PL]} = \mathbf{V}'_{[PL]} \times \mathbf{D}'_{[LL]}^{1/2}; \text{ con } \mathbf{D}'_{[LL]} = \text{Diag}(\mathbf{\Xi}') \quad (4.23)$$

Una vez analizado el compromiso, se lleva a cabo un análisis de la intra-estructura, proyectando las filas y columnas de las K-matrices de Co-IA sobre el plano compromiso (o sobre los ejes factoriales del compromiso).

Para ello se realiza un doble centrado sobre cada k-matriz de $\mathbf{Z}'_{[QP,k]}$, y se obtienen las FK (puntuaciones para las variables fila) y QK (cargas para las variables columna):

$$FK = \mathbf{Z}'_{[QP,k]} \times \mathbf{V}'_{[PL]} \quad \text{y} \quad QK = \mathbf{Z}'_{[QP,k]}^T \times \mathbf{U}'_{[QL]} \quad (4.24)$$

En la figura siguiente se representan mediante un esquema, los pasos a seguir para realizar el método MixSTATICO (Fig. 4.1):

n : individuals number
 p, q : variables number – original matrix
 p', q' : variables number – disjunctive matrix

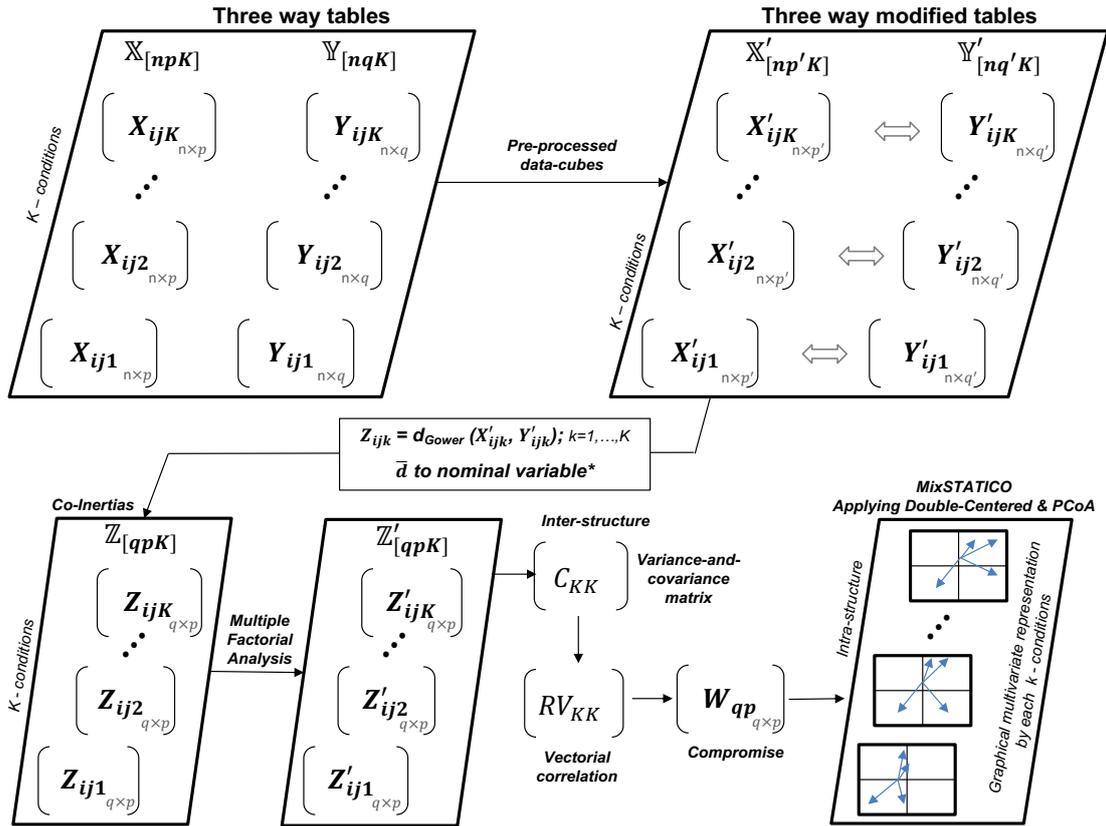


Figura 4.1. Esquema del método MixSTATICO.

A continuación, se detallan las ventajas que ofrece el nuevo método:

- Permite incorporar en un solo análisis variables cuantitativas y cualitativas (mixtas);
- Permite estudiar la relación entre conjuntos de matrices de datos que contienen variables cuantitativas y cualitativas (mixtas);
- Visualización de gráficos de forma independiente;
- Elección de diferentes ejes para presentar el mapa bi-factorial.

4.2 Aplicación del MixSTATICO a datos ilustrativos.

A continuación, se presentan algunos de los datos espacio-temporales que describen las características ambientales y biológicas de la región costera de Ecuador (se utiliza la misma base de datos del **Capítulo 3**, pero con una estructura diferente). Los datos analizados corresponden a las mediciones durante tres meses del año 2013 (características oceanográficas normales). Se muestrearon las mismas cuatro estaciones fijas, es decir: Esmeraldas (S1), Manta (S2), La Libertad (S3) y Pto. Bolívar (S4), en el punto de muestreo de 0-20m. Cada matriz se corresponde con un mes de muestreo (condición) (Fig. 4.2).

Hay dos matrices de tres vías, $X_{[NPK]}$ (ambiente) y $Y_{[NPK]}$ (especie), cada una con $K = 3$ meses (tiempo). Los individuos (filas) son las cuatro estaciones de muestreo, mientras que las variables (columnas) son parámetros fisicoquímicos y la abundancia de las especies de fitoplancton (Fig. 4.2).

		$X_{[NPK]}$			$Y_{[NPK]}$			
		T	OD	SiO ₄	e8	e22	e24	Grupo
Individuos	Junio_2013							
	Mayo_2013							
	Abril_2013							
	S1: 0_20m	27.4	4.4	Superior	0E+00	2E+03	3	Pennate
	S2: 0_20m	22.8	4.5	Superior	0E+00	0E+00	3	Ciliate
	S3: 0_20m	24.8	4.4	Superior	3E+04	8E+02	1	Pennate
	S4: 0_20m	23.1	4.2	Inferior	0E+04	8E+02	0	Pennate

Figura 4.2. Esquema de la estructura de los datos.

En el conjunto de matrices ambientales se han considerado las siguientes variables: la temperatura (T en ° C), la concentración de oxígeno disuelto (OD en mg.L⁻¹) (continuas-cuantitativas) y la concentración de silicato (binaria-cualitativa: para propósitos exclusivos de este trabajo, los valores de silicato se han dicotomizado

dependiendo de si se presentan concentraciones superiores o inferiores al promedio mensual de la estación de muestreo).

Las matrices de especies incluyen las abundancias de las especies de *G. striata* (e8) y *Gyrodinium* sp. (e22) expresadas en cel L⁻¹ (discreta-cuantitativa), y para la especie *M. rubrum* (e24) se ha tomado la información de forma ordinal-cualitativa (0 si $y_{ij} < Q_1$; 1 si $Q_1 < y_{ij} < Q_2$; 2 si $Q_2 < y_{ij} < Q_3$ y 3 si $y_{ij} \geq Q_3$ para los cuartiles calculados). Como últimas variables (tipo nominal), se utilizan el grupo taxonómico al que pertenece la especie más abundante en la estación y mes.

La Tabla 4.1 muestra los valores promedio mensuales para los primeros 20m de profundidad. Los datos son de tipo mixto y, por tanto, adecuados para probar la eficacia de este método.

Tabla 4.1. Datos para ejemplo: Variables ambientales y especies en diferentes momentos y lugares.

Condiciones	Tiempo	Espacio	Ambiente			Especie			
			Temperatura (T)	Oxígeno Disuelto (OD)	Silicato (SiO ₄)	<i>G. striata</i> (e8)	<i>Gyrodinium</i> sp. (e22)	<i>M. rubrum</i> (e24)	Grupo
k=1	Abril	S1	27.4	4.4	Superior	0.E+00	2.E+03	3	Pennadas
	Abril	S2	22.8	4.5	Superior	0.E+00	0.E+00	3	Ciliados
	Abril	S3	24.5	4.4	Superior	3.E+04	8.E+02	1	Pennadas
	Abril	S4	23.1	4.2	Inferior	0.E+00	8.E+02	0	Pennadas
k=2	Mayo	S1	26.1	4.6	Inferior	3.E+04	2.E+03	2	Céntricas
	Mayo	S2	24.7	4.1	Superior	6.E+03	0.E+00	0	Céntricas
	Mayo	S3	26.1	4.4	Inferior	2.E+04	8.E+02	2	Pennadas
	Mayo	S4	25.0	4.5	Superior	0.E+00	2.E+03	3	Ciliados
k=3	Junio	S1	26.3	4.6	Superior	2.E+05	0.E+00	1	Céntricas
	Junio	S2	20.5	3.2	Inferior	2.E+05	0.E+00	0	Céntricas
	Junio	S3	24.3	4.3	Inferior	1.E+05	0.E+00	0	Céntricas
	Junio	S4	22.2	5.1	Inferior	1.E+06	0.E+00	3	Céntricas

Las abreviaturas de las variables ambientales están detalladas entre paréntesis, los códigos de las especies se encuentran en la Tabla 3.1.

El primer paso es el preprocesamiento de las matrices. Las variables ambientales cuantitativas se centran y normalizan por variable (matriz total) para eliminar la escala de medición. La variable binaria tiene valores de 1 (más alto) y 0 (más bajo).

Las variables de especies se transforman con $\log_{10}(y_{ij} + 1)$ con la finalidad de eliminar el efecto de captura dominante durante el muestreo (Legendre & Legendre, 1979). Las variables ordinales se representan con códigos y las variables nominales se convierten en elementos de una matriz disyuntiva (Tabla 4.2).

Tabla 4.2. Preprocesamiento de matrices: Datos ambientales y especies son transformados para el análisis. Consultar la tabla 3.1 para los códigos de las especies.

Tiempo	Espacio	T	OD	SiO ₄	e8	e22	e24	Pennadas	Ciliados	Céntricas
Abril	S1	1.6	0.2	1	0.0	3.4	3	1	0	0
Abril	S2	-0.8	0.2	1	0.0	0.0	3	0	1	0
Abril	S3	0.0	0.0	1	4.5	2.9	1	1	0	0
Abril	S4	-0.7	-0.5	0	0.0	2.9	0	1	0	0
Mayo	S1	0.9	0.6	0	4.5	3.2	2	0	0	1
Mayo	S2	0.1	-0.6	1	3.8	0.0	0	0	0	1
Mayo	S3	0.9	0.0	0	4.2	2.9	2	1	0	0
Mayo	S4	0.3	0.4	1	0.0	3.2	3	0	1	0
Junio	S1	1.0	0.6	1	5.3	0.0	1	0	0	1
Junio	S2	-2.1	-2.7	0	5.2	0.0	0	0	0	1
Junio	S3	0.0	-0.1	0	5.0	0.0	0	0	0	1
Junio	S4	-1.2	1.8	0	6.2	0.0	3	0	0	1

A partir de los datos ambientales y de especies transformados (Tabla 4.2), se calculan las K-matrices de distancia de Gower entre las variables de cada par de matrices concatenadas $\mathbb{X}\mathbb{Y}_{[Nr'K]}$, obteniéndose el conjunto de K-matrices de Co-IA $\mathbb{Z}_{[QPK]}$. La variable Grupo (Tabla 4.3) contiene la distancia promedio de las categorías del grupo en cada variable ambiental (Ecs. (4.7), (4.8) y (4.9)). Luego se normaliza cada k-matriz multiplicándola por el inverso de su respectivo valor propio $\mathbf{1}/\lambda_{[K]}$ (Ec. (4.13)), como en el método AFM, y se obtiene la matriz de distancia ponderada $\mathbb{Z}'_{[QPK]}$.

Tabla 4.3. Matrices de distancias de Gower entre variables ambientales y especies. Consultar la tabla 3.1 para los códigos de las especies.

Tiempo	Especies	T	OD	SiO₄
Abril	e8	0.5	0.3	0.3
Abril	e22	0.6	0.6	0.5
Abril	e24	0.5	0.5	0.3
Abril	Grupo	0.3	0.1	0.2
Mayo	e8	0.6	0.8	0.7
Mayo	e22	0.5	0.6	0.6
Mayo	e24	0.3	0.4	0.4
Mayo	Grupo	0.1	0.1	0.1
Junio	e8	0.9	0.9	0.8
Junio	e22	0.2	0.2	0.0
Junio	e24	0.2	0.2	0.1
Junio	Grupo	0.2	0.2	0.1

$$\mathbf{1}/\lambda_{[K]} = \begin{pmatrix} 1/2.1 \\ 1/2.9 \\ 1/2.5 \end{pmatrix} \quad (4.25)$$

Posteriormente se vectoriza el conjunto de matrices $\mathbb{Z}'_{[QPK]}$, aplicando la Ec. (4.13) y se obtiene la matriz $\mathbf{C}_{[KK]}$ de la Ec. (4.14) (Tabla 4.4) y la matriz $\mathbf{RV}_{[KK]}$ de la Ec. (4.15) (Tabla 4.5).

Table 4.4. Matriz de varianzas y covarianzas vectoriales.

	Abril	Mayo	Junio
Abril	0.5	0.4	0.3
Mayo	0.4	0.3	0.3
Junio	0.3	0.3	0.4

Table 4.5. Matriz RV.

	Abril	Mayo	Junio
Abril	1.0	0.9	0.7
Mayo	0.9	1.0	0.8
Junio	0.7	0.8	1.0

Aplicando la DVS a la matriz RV (Ec. (4.16)), se puede visualizar la similitud de las relaciones entre el medio ambiente y las especies en un momento dado. En abril y mayo, las relaciones especies-ambiente fueron más similares que en junio (Fig. 4.3A). Esta asociación está relacionada con las dos estaciones climáticas de la región, la estación lluviosa (diciembre-mayo) y la estación seca (junio-noviembre), (INOCAR, 2012). Se muestra la interestructura, con 87.13% (CP1) y 11.91% (CP2) de inercia total explicada (Fig. 4.3B).

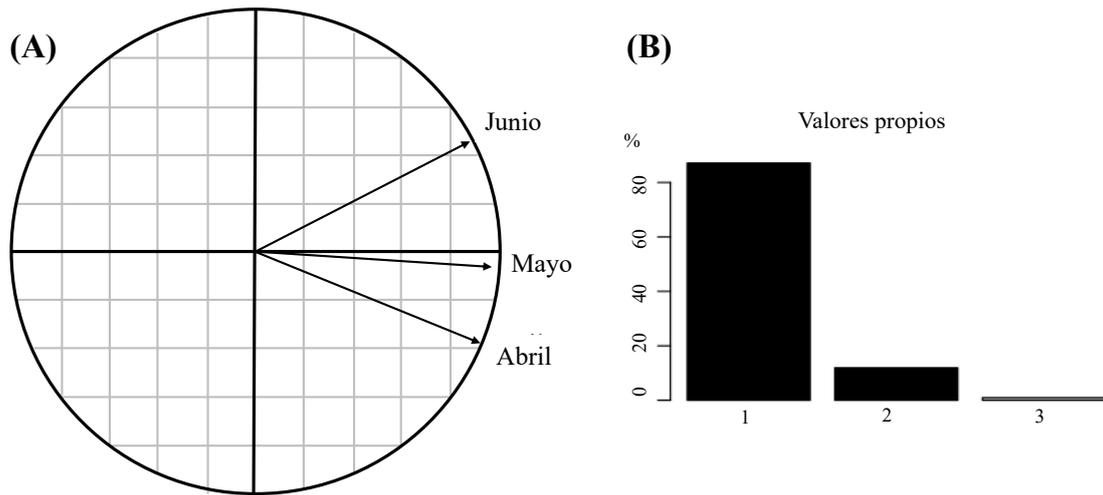


Figura 4.3. (A) Gráfico de la Inter-estructura y (B) valores propios (eigenvalues).

A partir de la factorización de la matriz RV, también se obtiene el vector de pesos $\vec{\alpha}_{[K]}$ (Ec. (4.19)) para utilizarlo posteriormente al realizar la combinación lineal multiplicando los pesos por las K-matrices de Co-IA (Ec. (4.21)), dando como resultado la matriz compromiso $\mathbf{W}_{[QP]}$ (Tabla 4.6):

$$\mathbf{W}_{[QP]} = \sum_{k=1}^K \alpha_{[k]} \cdot \mathbf{Z}'_{[QP,k]} = \sum_{k=1}^K \begin{pmatrix} 0.33_{k=1} \\ 0.35_{k=2} \\ 0.32_{k=3} \end{pmatrix} \cdot \mathbf{Z}'_{[QP,k]} \quad (4.26)$$

Continuando con la Ec. (4.22), se calcula la DVS de la matriz compromiso $\mathbf{W}_{[QP]}$ y se construye la imagen promedio de la relación especie-ambiente dada por la Ec.

(4.23). Los porcentajes de inercia explicada fueron 99.71% (CoP1) y 0.17% (CoP2) (Fig. 4.4C).

Tabla 4.6. Matriz compromiso de las relaciones especies-medioambiente.

Consultar la tabla 3.1 para los códigos de las especies.

	T	OD	SiO₄
e8	0.3	0.3	0.2
e22	0.2	0.2	0.2
e24	0.1	0.1	0.1
Grupo	0.1	0.1	0.1

A partir del análisis del compromiso, se puede observar para las variables ambientales (Fig. 4.4A) que el primer eje, CoP1, recoge un gradiente de temperatura y separa el oxígeno disuelto de la temperatura. El segundo eje separa el silicato (en la parte inferior) de las otras dos variables, mostrando una relación inversa entre ellos. Mientras que, para las variables de especies (Fig. 4.4B) el primer eje separa las especies *M. rubrum* (e24) y *Gyrodinium* sp. (e22) de la especie *G. striata* (e8) y de la variable Grupo taxonómico, presentando una relación inversa entre dichos grupos. El segundo eje separa, fundamentalmente *M. rubrum* (e24) de las otras dos especies. Al interpretar los dos graficos de forma conjunta, se observa que las especies *M. rubrum* (e24) y *Gyrodinium* sp. (e22) manifiestan una alta abundancia, la que estuvo asociada con valores por encima del valor promedio de oxígeno disuelto, pero por debajo del valor promedio de temperatura y silicato durante los tres meses de estudio. La especie *G. striata* (e8) asoció su alta abundancia con valores por encima del valor promedio de silicato, pero bajo en oxígeno disuelto. El Grupo taxonómico de las especies que presentaron la mayor abundancia en ese periodo de muestreo, estuvo asociado con valores por encima del valor promedio de temperatura.

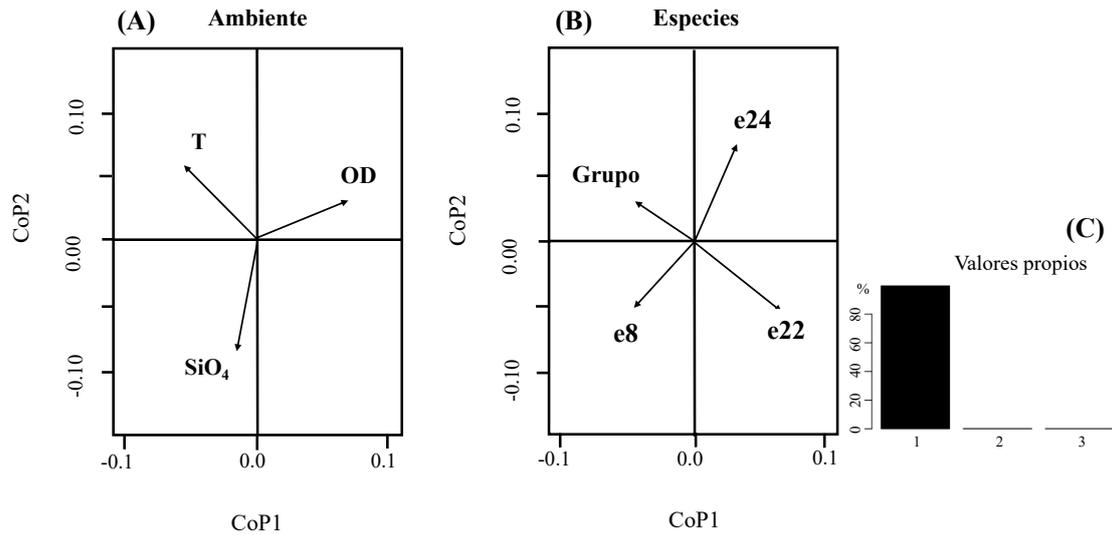


Figura 4.4. Compromiso: plano bi-factorial para (A) ambiente y (B) especies. (C) Valores propios. Consultar la Tabla 3.1 para los códigos de las especies.

Los valores de Cos^2 fueron 0.93 (abril), 0.99 (mayo) y 0.88 (junio), lo que indica que la matriz del compromiso representa mejor la relación especies-ambiente en mayo, seguida de abril (temporada de lluvias).

En la intra-estructura (Fig. 4.5), la relación especies-ambiente de cada k-matriz se proyecta en el mismo plano bi-factorial del compromiso (ver Ec. (4.24)).

En abril, las especies *Gyrodinium* sp. (e22) y *M. rubrum* (e24) mostraron mayor abundancia (vectores más largos) y está asociada con valores por debajo del promedio en temperatura y silicatos y valores por encima del promedio en oxígeno disuelto respectivamente (Fig. 4.5B-abril); mientras que *G. striata* (e8) y Grupo taxonómico, destacaron su preferencia ambiental con valores por encima del promedio en temperatura y silicato. Estas variables ambientales registraron valores de concentraciones mayores a lo que ocurrió en el mes de junio (Tabla 4.1). La especie *G. striata* (e8) y la variable Grupo taxonómico mostraron una relación inversa con *Gyrodinium* sp. (e22) y *M. rubrum* (e24), señalando que a medida que aumentaba la abundancia de estas dos especies (dinoflagelado y ciliados,

respectivamente), disminuía la abundancia de las diatomeas *G. striata* (e8) (céntricas) y Grupo (pennadas en su mayoría) en las estaciones de muestreo.

En mayo (mes de cambio de estación), la abundancia de las especies *G. striata* (e8), *Gyrodinium* sp. (e22) y *M. rubrum* (e24) asociaron su alta abundancia con valores por encima del valor promedio en silicato y oxígeno disuelto pero bajos en temperatura (Fig. 4.5A y B-mayo). En este mes las variables ambientales en las estaciones S2, S3 y S4 alcanzaron valores superiores a los registrados en junio (Tabla 4.1). La variable Grupo taxonómico (céntricas, pennadas y ciliados) mostró una relación inversa con las demás especies (Fig. 4.5B-mayo).

En junio las especies *G. striata* (e8) y *M. rubrum* (e24) registraron abundancia principalmente en las estaciones S1 y S4, mientras que *Gyrodinium* sp. (e22) no estuvo presente en absoluto (Tabla 4.1). La especie *Gyrodinium* sp. (e22) y Grupo taxonómico (diatomeas céntricas) mostraron una relación inversa a las especies *G. striata* (e8) y *M. rubrum* (e24) (Fig. 4.5B-junio) y asociaron su mayor abundancia con valores por encima del valor promedio de temperatura y silicatos.

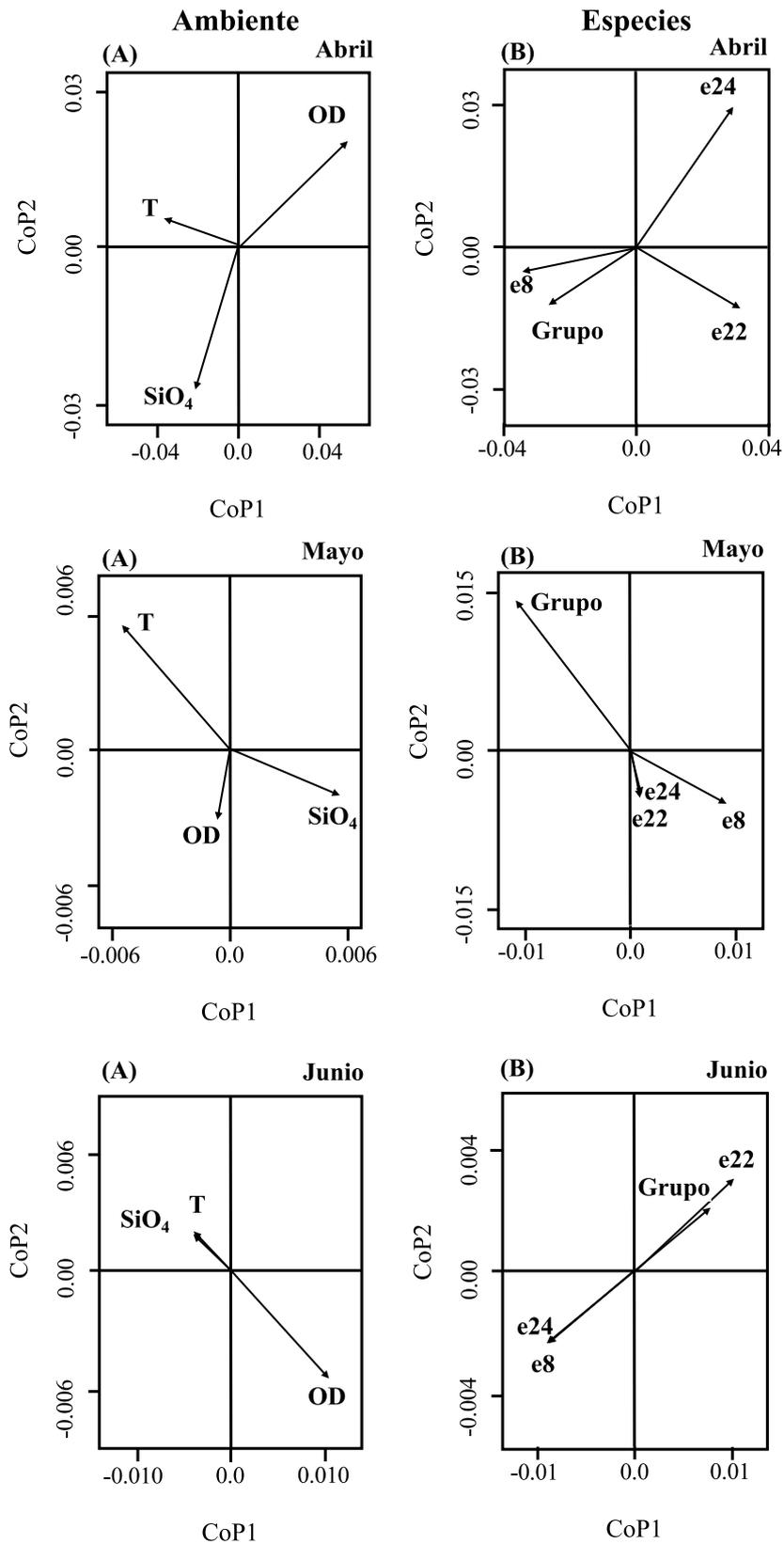


Figura 4.5. Intra-estructura: Proyecciones sobre el plano bi-factorial del compromiso (A) ambiente y (B) especies correspondientes a cada k-matriz (tiempo). Consultar la Tabla 3.1 para los códigos de las especies.

4.3 Aplicación del MixSTATICO a datos reales.

Se utiliza la misma base de datos que en la Sección 4.2, esta vez incorporando los 11 meses del año 2013 (la información de enero no estaba disponible). Además, se aumenta el número de parámetros ambientales y de especies. Para este análisis, se ingresan dos veces la abundancia de las especies “*S. membranacea* y *M. rubrum*” pero con diferentes tipos de variables: 1) como variables discretas-cuantitativas representando a la abundancia expresada en cel L⁻¹ y 2) como variables de tipo ordinal-cualitativa, adoptando el mismo procedimiento de los cuartiles detallado en el ejemplo ilustrativo 4.2 y como variable del tipo nominal “Grupo” se especifica el grupo taxonómico al que pertenece la especie que destacó por ser la más abundante en una estación y el respectivo mes de muestreo (Fig. 4.6), ver Anexos A9, A10 y A11.

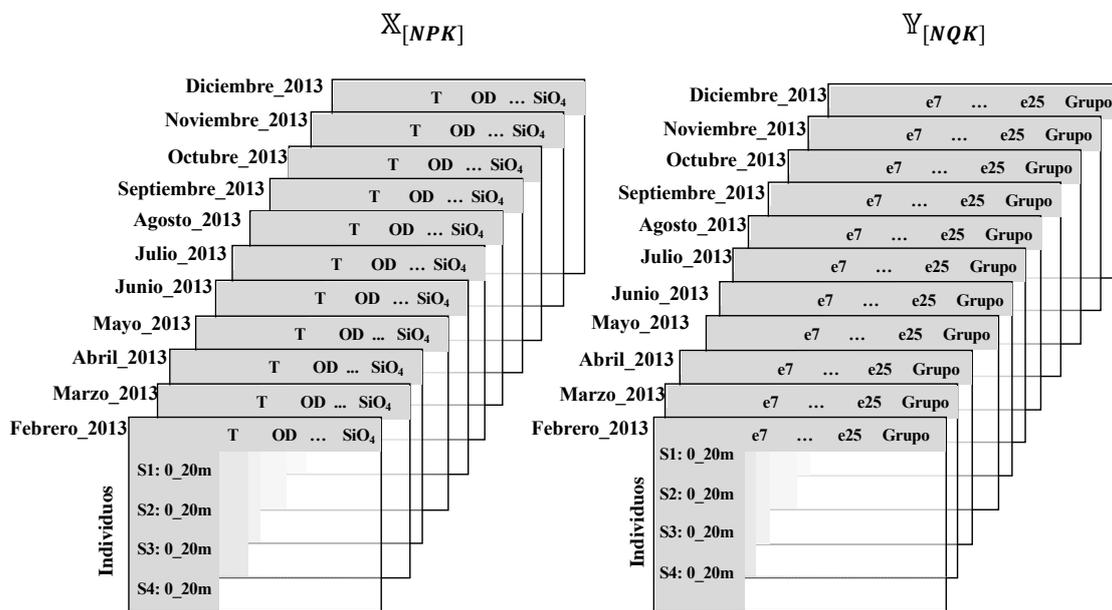


Figura 4.6. Esquema de la estructura de los datos. Variables ambientales (P = 7 cuantitativas), variables especies (Q = 10, que están clasificadas en: 7 cuantitativas, 2 ordinales y 1 nominal)

El gráfico de la inter-estructura muestra dos clases de comportamiento en la relación especies-ambiente, que varían según la estación (Fig. 4.7A). Los miembros de la primera clase manifestaron un comportamiento similar durante los meses de época seca (color rojo) y los de la segunda en los meses de época de lluvias (color azul), excepto en los meses de marzo y octubre. Las dos primeros componentes explicaron el 95.27% de la inercia (Fig. 4.7B). Noviembre tuvo mayor contribución a la formación del compromiso (peso = 0.094), seguido de agosto y septiembre (época seca). El compromiso representa más notablemente la relación especies-ambiente del mes de noviembre (valor más alto de \cos^2) (Fig. 4.7C).

Mayo es el mes con menor contribución al compromiso y es considerado un mes de cambio estacional, poniendo fin a la temporada de lluvias.

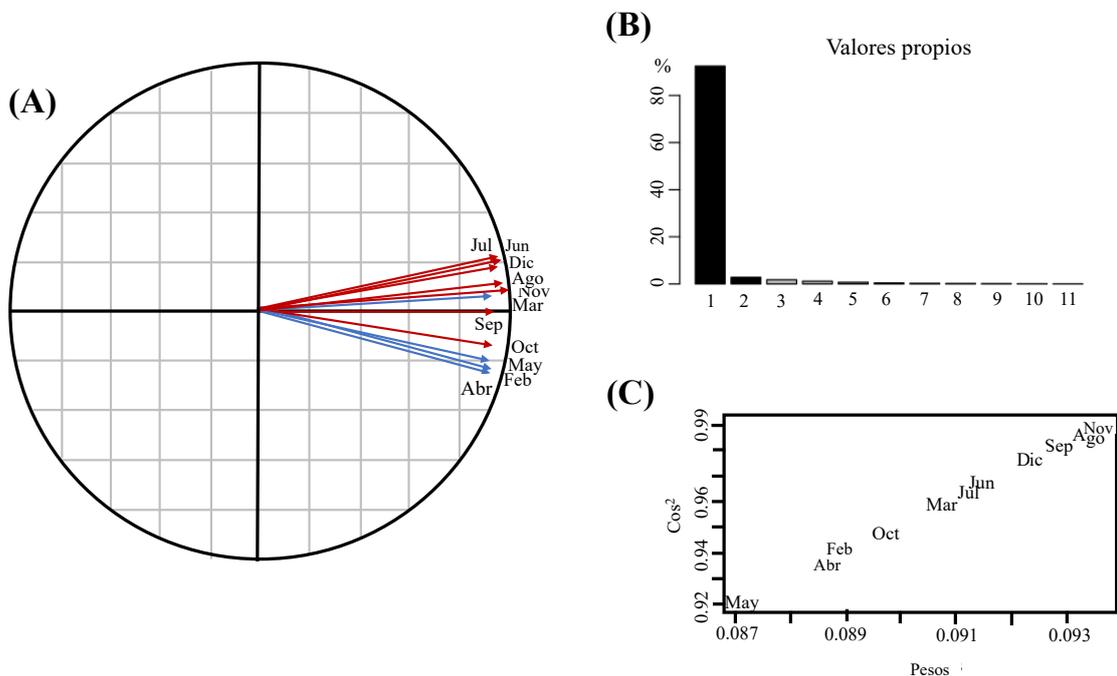


Figura 4.7. (A) Gráfico de la Inter-estructura (época seca: color rojo; época lluviosa: color azul). (B) Valores propios y (C) pesos (α) y \cos^2 para cada K-matriz.

Las dos primeras componentes del compromiso explican el 99.95% de la inercia (Fig. 4.8C). El compromiso para las variables ambientales muestra una fuerte correlación positiva entre la temperatura y el oxígeno disuelto, e inversa con el fosfato (Fig. 4.8A). Mientras que, las mismas especies (repetidas) mostraron una fuerte correlación directa, como: *S. membranacea* (variable cuantitativa e19) y *S. membranacea* (variable ordinal e25) localizada en el segundo cuadrante, mientras que *M. rubrum* (variable cuantitativa e23) y *M. rubrum* (variable ordinal e24) mostraron proximidad en la ubicación en el primer cuadrante (Fig. 4.8B).

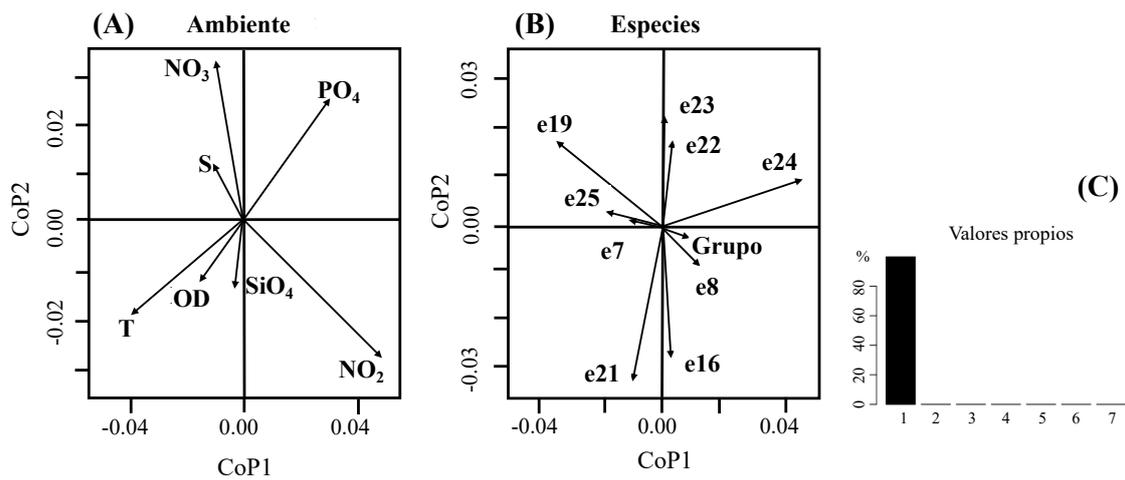


Figura 4.8. Compromiso: plano bi-factorial para (A) ambiente y (B) especies. (C) Valores propios. Consultar la Tabla 3.1 para los códigos de las especies.

El análisis de la intra-estructura nos permitió observar la proximidad en cada mes, entre las mismas especies que se analizaron utilizando diferentes tipos de variables, como por ejemplo en las proximidades de *S. membranacea* (e19) con *S. membranacea* (e25) y *M. rubrum* (e24) con *M. rubrum* (e23) mencionadas anteriormente (Fig. 4.9). Estos resultados evidenciaron la validez del método para analizar datos mixtos.

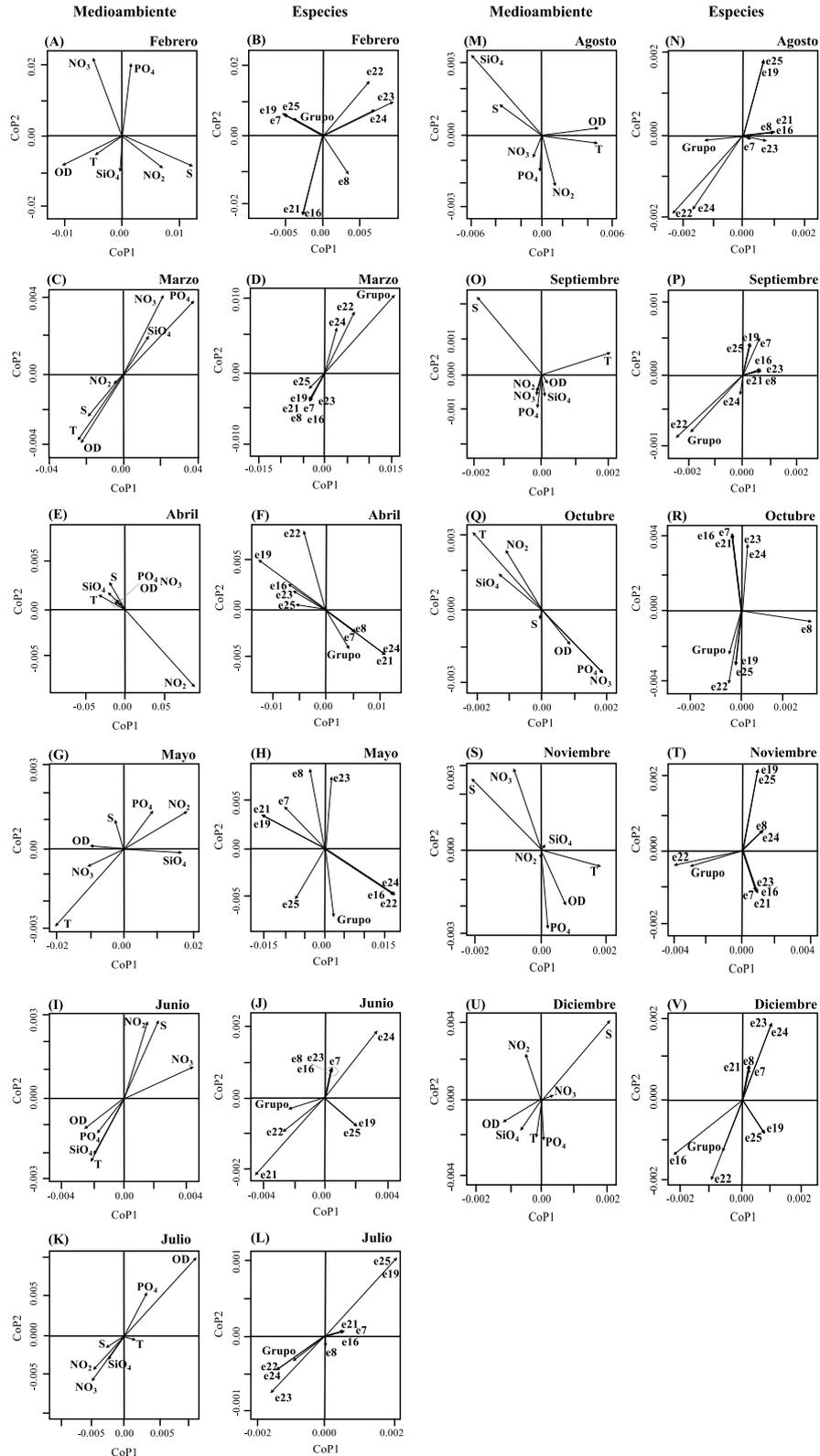


Figura 4.9. Intra-estructura: proyecciones de las K-matrices (tiempo) sobre el plano bi-factorial del compromiso: (A, C, E, G, I, K, M, O, Q, S, U) ambiente y (B, D, F, H, J, L, N, P, R, T, V) especies. Consultar la Tabla 3.1 para los códigos de las especies.

CAPÍTULO 5

*Y todo lo que hacéis, sea de palabra
o de hecho, hacedlo todo en el
nombre del Señor Jesús,
dando gracias a Dios Padre
por medio de él
Colosenses 3:17*

CAPÍTULO 5: HERRAMIENTAS INFORMÁTICAS DESARROLLADAS EN EL SOFTWARE R.

En este capítulo se presentan en detalle la descripción y las líneas de código realizadas en RStudio para la ejecución de la función HJSTATICO y del MixSTATICO.

5.1 Librería HJSTATICO utilizando el software R.

Se utilizó el programa RStudio (versión 1.3.1073), para el desarrollo de la nueva herramienta informática que permite aplicar la innovadora metodología denominada HJ-STATICO. Esta nueva herramienta consiste en la creación de una librería que adoptó el nombre de la metodología, y se denomina HJSTATICO. Las líneas de código para esta nueva librería se encuentran disponibles en el **Capítulo 5.1.2.**

Inicialmente el usuario debe asegurarse de ordenar el conjunto de matrices en orden yuxtapuesto vertical según la variable que defina como la tercera vía. Hay que tener en cuenta que los datos deben almacenarse en tres archivos: un archivo para los datos ambientales (solo con los datos de las variables), el segundo para los datos de especies (solo con los datos de las especies) y el tercero que contiene las etiquetas (variable factor que defina la tercera-vía) (ver Anexos A3, A4 y A5).

Para la ejecución de la librería HJSTATICO se utilizan varias funciones (desarrolladas en este trabajo) que realizan procesos como: el preprocesamiento de los conjuntos de matrices, el cálculo del HJ-Biplot, representación gráfica del

compromiso, representación gráfica para la intra-estructura. Previo a la aplicación de la función HJSTATICO el usuario debe asegurarse de ejecutar cada una de las funciones mencionadas. A continuación, se presenta una breve explicación del funcionamiento de las funciones:

Preprocesamiento: matrices de entrada x (ambiente), y (especies), condition (variable de tipo factor que definirá la tercera-vía en el análisis) y el tipo de método de estandarización a aplicarse en las matrices.

Prep.Tables=function(x=env,y=spe,condition=eti\$space,type=c("B.Partial", "B.Total", "A.Norma"))

HJ-Biplot: realiza el cálculo de los indicadores de calidad de representación y cantidad de contribución para el compromiso y la intra-estructura.

hjb=function(biplot=TRUE,F=F,Q=Q,dimx,dimy,compromise=compromise)

Gráfico del compromiso: realiza el gráfico clásico del ACP si biplot=FALSE en la función HJSATICO y realiza el gráfico considerando la perspectiva HJ-Biplot si biplot=TRUE

graphcomp=function(dimx,dimy,F,Q,biplot,compromise,HJB)

Gráficos de la intra-estructura: realiza el gráfico clásico del ACP si biplot=FALSE en la función HJSATICO y realiza el gráfico considerando la perspectiva HJ-Biplot si biplot=TRUE

graphintra=function(dimx=compx,dimy=compy,FK=FK,QK=QK,indexz1,biplot,compromise=compromise,k=k)

HJ-STATICO: solo en esta función, el usuario ingresará los argumentos tales como: matrices y especificaciones para realizar el análisis.

Ofrece resultados tipo lista.

Argumentos:

x: conjunto de matrices yuxtapuestas en orden vertical que contiene los datos de ambiente;

y: conjunto de matrices yuxtapuestas en orden vertical que contiene los datos de especies;

condition: variable tipo factor que definirá la tercera-vía para el análisis;

biplot: valor lógico TRUE por defecto que calculará y representará gráficamente el compromiso y la intra-estructura considerando la perspectiva del HJ-Biplot;

preprocessing: seleccionar el método de estandarización a utilizar. En los datos ambientales se tienen las opciones de B.Total (estandarización total), B.Parcial (estandarización parcial) y A.Norma (estandarización por norma). Para el caso de las especies la función realiza por default un centrado por columna a la matriz total;

comp_x y **comp_y**: indica las componentes que se utilizarán para la representación de los gráficos del compromiso y de la intra-estructura sobre el plano bi-factorial.

Valores:

X: presenta las k-matrices de ambiente estandarizadas en el preprocesamiento;

Y: presenta las k-matrices de especies estandarizadas en el preprocesamiento;

Crosstables: presenta las k-matrices cruzadas obtenidas con la Co-IA;

Compromise: presenta la matriz compromiso;

HJB: presenta los indicadores de calidad de representación y contribuciones obtenidos con el método HJ-Biplot para la matriz compromiso;

HJBK: presenta los indicadores de calidad de representación y contribuciones obtenidos con el método HJ-Biplot para la intra-estructura.

HJSTATICO=function(x=env,y=spe,condition=etiq\$space,biplot=TRUE,preprocessing="B.Total",compx=1,compy=2)

A continuación, en la Tabla 5.1 se presenta una comparación entre los beneficios que ofrece la librería **HJSTATICO** y la función “**statico**” disponible en el paquete **ade4** de R.

Tabla 5.1 Ventajas de la herramienta HJSTATICO en relación con el STATICO.

Diferencias	HJSTATICO	STATICO – ADE4
Mayor facilidad de realizar el análisis	<pre>resultado=HJSTATICO(x=env,y=spe,condition=eti iq\$space,biplot=TRUE,preprocessing = "B.Total",compX=1,compY=2)</pre>	<pre>data(meau) wit1 <- withinpca(env, etiq\$space, scan = FALSE, scal = "total") spepca <- dudi.pca(spe, scale = FALSE, scan = FALSE, nf = 2) wit2 <- wca(spepca, etiq\$space, scan = FALSE, nf = 2) kta1 <- ktab.within(wit1, colnames = etiq\$ID) kta2 <- ktab.within(wit2, colnames = etiq\$ID) statico1 <- statico(kta1, kta2, scan = FALSE) #realiza análisis statico plot(statico1) kplot(statico1)</pre>
Métodos de estandarización	Estandarización Total de Bouroche; Estandarización Parcial de Bouroche y Estandarización por Norma de Abdi.	Estandarización Total de Bouroche; Estandarización Parcial de Bouroche y
Visualización de los gráficos	Un gráfico por ventana para el compromiso y la intra-estructura.	Presenta varios gráficos en una sola ventana dificultando la interpretación.
Incorpora HJ-Biplot	Presenta los gráficos del compromiso e intra-estructura como un ACP y como un Biplot	Presenta los gráficos de compromiso e intra-estructura solo como un ACP
Indicadores HJ-Biplot	Ofrece indicadores que cuantifican la calidad de representación y contribución de los elementos de la matriz.	No lo realiza
Ejes del mapa bi-factorial	Permite generar los mapas bi-factoriales considerando diferentes dimensiones.	Permite generar mapas bi-factoriales considerando solo la dimensión 1 y dimensión 2

5.1.1 Algoritmo HJSTATICO

1. Preprocesamiento del conjunto de matrices iniciales $\mathbb{X}_{[NPk]}$ y $\mathbb{Y}_{[NQk]}$.

2. Cálculo de la co-estructura entre cada par de k-matrices

Para $k=1, \dots, K$

$$\{\mathbf{Z}'_{[QP]}^{(k)} = \mathbf{y}_{[NQ]}^{(k)T} \mathbf{D}_N^{(k)} \mathbf{X}_{[NP]}^{(k)}\}$$

3. Cálculo del PTA para $\mathbf{Z}'_{[QPk]}$

4. Para $k=1, \dots, K$

$$\{C_{[KK]} \mid c_{kl} = \text{Trace}(\mathbf{Z}'^T_{[QP,k]} \times \mathbf{Z}'_{[QP,l]})\}; \text{ siendo } k = l = 1, \dots, K\}$$

$$\{RV_{[KK]} = \frac{c_{kl}}{\sqrt{c_{kk} \times \sqrt{c_{ll}}}}\}$$

5. Cálculo de la inter-estructura.

DVS ($RV_{[KK]}$) para obtener el vector de pesos u_1

Cálculo de los pesos aplicando reescalamiento $\vec{\alpha}_{[K]} = \vec{u}_1 \times (\vec{u}_1^T \mathbf{1})^{-1}$

6. Cálculo del Compromiso.

$$\mathbf{Compromise}_{[QP]} = \sum_{k=1}^K \alpha_{[k]} \mathbf{Z}'_{[QP,k]}$$

7. DVS ($\mathbf{Compromise}_{[QP]}$).

Representación grafica del compromiso aplicando el método ACP clásico, o utilizando el método HJ-Biplot

8. Proyección de la Intra-estructura.

Representación grafica de la intra-estructura aplicando el método ACP clásico, o utilizando el método HJ-Biplot

5.1.2 Código fuente HJSTATICO

```
library(openxlsx)
```

```
library(pracma)
```

```
library(ade4)
```

```
library(MFAG)
```

```
env=read.xlsx("BD_Inocar2015.xlsx",sheet="env2015")
```

```
spe=read.xlsx("BD_Inocar2015.xlsx",sheet="esp2015")
```

```
etiq=read.xlsx("BD_Inocar2015.xlsx",sheet="label")
```

```
#####FUNCTION PREPROCESSING TABLES#####
```

```
Prep.Tables=function(x=env,y=spe,condition=etiq$space,type=c("B.Partial","B.Total","A.Norma")) {
```

```
  dimenx=dim(x); dimeny=dim(y)
```

```
  k=dim(table(condition))
```

```
  #Index matrix environment and species
```

```
  index=matrix(nrow=k,ncol=4)
```

```
  index=as.data.frame(index)
```

```
  colnames(index)=c("Tables","Start","Finish","n")
```

```
  index[1,1]=condition[1]
```

```
  index[1,2]=1
```

```
  i=1; count=1
```

```
  s=1; num=1
```

```
  for (i in 2:length(condition)) {
```

```
    if(condition[i]!=index[s,1]) {
```

```
      index[s,3]=count
```

```
      index[s,4]=num
```

```
      s=s+1
```

```
      num=1
```

```
      index[s,1]=condition[i]
```

```
      count=count+1
```

```

    index[s,2]=count}
else {
    count=count+1
    num=num+1
    index[s,3]=count
    index[s,4]=num
}
}
index
#standarization by column - whole table
mean=apply(x,2,mean)
sd=apply(x,2,sd)
sd= sd*sqrt((dimenx[1]-1)/dimenx[1])
z=matrix(nrow=dimenx[1],ncol=dimenx[2])
for (i in 1:dimenx[1]) {
  for (j in 1:dimenx[2]) {
    z[i,j]=(x[i,j]-mean[j])/sd[j] }
}
if (type=="A.Norma") {
#preprocesamiento de las tablas,centrar y normalizar cada tabla por columnas
for(i in 1:k){
  temp=matrix(nrow=index[i,4],ncol=dimenx[2])
  temp=z[index[i,2]:index[i,3],]
  temp=as.matrix(temp)
  temp2=temp%*%t(temp)
  sol=1/sqrt(norm(temp2,type=c("O")))#norma
  if (i==1) {
    X=z[index[i,2]:index[i,3],]*sol}
  else{
    temp3=z[index[i,2]:index[i,3],]*sol
    X=rbind(X,temp3)

```

```

    }
  }
}
if (type=="B.Partial") {
  avem=matrix(nrow=dimenx[1],ncol=dimenx[2])
  stdm=matrix(nrow=dimenx[1],ncol=dimenx[2])
  for (i in 1:k) {
    temp=z[index[i,2]:index[i,3],]
    tempav=apply(temp,2,mean)
    tempsd=apply(temp,2,sd)
    tempsd=tempsd*sqrt((index[i,4]-1)/index[i,4])
    ta=matrix(nrow=index[i,4],ncol=dimenx[2])
    tsd=matrix(nrow=index[i,4],ncol=dimenx[2])
    for (ni in 1:index[i,4]) {
      ta[ni,]=tempav
      tsd[ni,]=tempsd }
    if (i==1){ avem=ta
      stdm=tsd }
    else { avem=rbind(avem,ta)
      stdm=rbind(stdm,tsd) }
  }
  X=(z-avem)/stdm
}
if (type=="B.Total") {
  avem=matrix(nrow=dimenx[1],ncol=dimenx[2])
  for (i in 1:k) {
    temp=z[index[i,2]:index[i,3],]
    tempav=apply(temp,2,mean)
    ta=matrix(nrow=index[i,4],ncol=dimenx[2])
    for (ni in 1:index[i,4]) {
      ta[ni,]=tempav }
  }
}

```

```

    if (i==1){ avem=ta}
    else { avem=rbind(avem,ta)}
  }
  r=z-avem
  temp=apply(r,2,sd)
  temp=temp*sqrt((dimenx[1]-1)/dimenx[1])
  stdm=matrix(nrow=dimenx[1],ncol=dimenx[2])
  for (i in 1:dimenx[1]) {
    stdm[i,]=temp }
  X=r/stdm
}
#Species Matrix-average matrix
avematrixs=matrix(nrow=dimeny[1],ncol=dimeny[2])
for (i in 1:dimeny[1]) {
  avematrixs[i,]=apply(y,2,mean) }
Y=y-avematrixs
colnames(X)=colnames(x)
Lst=list(X=X,Y=Y,index=index)
}

#####FUNCTION HJBILOT#####
hjb=function(biplot=TRUE,F=F,Q=Q,dimx,dimy,compromise=compromise){
  ii=dimx; jj=dimy
  bp=biplot
  J=F^2
  dimJ=dim(J)
  H=Q^2
  dimH=dim(H)
  rowJ=as.vector(rep(1,dimJ[1]))
  alfaJ=as.vector(rep(1,dimJ[2]))
  rowH=as.vector(rep(1,dimH[1]))

```

```

alfaH=as.vector(rep(1,dimH[2]))
for (i in 1:dimJ[1]) {rowJ[i]=sum(J[,i])}
for (i in 1:dimJ[2]) {alfaJ[i]=sum(J[,i])}
for (i in 1:dimH[1]) {rowH[i]=sum(H[,i])}
for (i in 1:dimH[2]) {alfaH[i]=sum(H[,i])}
CRTi=as.vector(rep(1,dimJ[1])); #contribución relativa a la traza del elemento
CREiFl=matrix(nrow=dimJ[1],ncol=2);
row.names(CREiFl)=row.names(compromise); names(CREiFl[1])=ii;
names(CREiFl[2])=jj
CRFIEi=matrix(nrow=dimJ[1],ncol=2);
row.names(CRFIEi)=row.names(compromise); names(CRFIEi[1])=ii;
names(CRFIEi[2])=jj
row
for (i in 1:dimJ[1]) {CRTi[i]=rowJ[i]/sum(alfaJ)}
for (i in 1:dimJ[1]) {
  CREiFl[i,1]=J[i,ii]/alfaJ[ii]; CREiFl[i,2]=J[i,jj]/alfaJ[jj]
  CRFIEi[i,1]=J[i,ii]/rowJ[i]; CRFIEi[i,2]=J[i,jj]/rowJ[i] }
CRTj=as.vector(rep(1,dimH[1]))#contribución relativa a la traza de la variable
CREjFl=matrix(nrow=dimH[1],ncol=2);
row.names(CREjFl)=colnames(compromise); names(CREjFl[1])=ii;
names(CREjFl[2])=jj
CRFIEj=matrix(nrow=dimH[1],ncol=2);
row.names(CRFIEj)=colnames(compromise); names(CRFIEj[1])=ii;
names(CRFIEj[2])=jj
for (i in 1:dimH[1]) {CRTj[i]=rowH[i]/sum(alfaH)}
for(i in 1:dimH[1]){
  CREjFl[i,1]=H[i,ii]/alfaH[ii]; CREjFl[i,2]=H[i,jj]/alfaH[jj]
  CRFIEj[i,1]=H[i,ii]/rowH[i];CRFIEj[i,2]=H[i,jj]/rowH[i] }
QLREi=matrix(nrow=dimJ[1],ncol=1)
QLREj=matrix(nrow=dimH[1],ncol=1)

```

```

QLREi=CRFIEi[,1]+CRFIEi[,2]; names(QLREi)=row.names(CRFIEi) #
names(QLREi)=c("QLREi")
QLREj=CRFIEj[,1]+CRFIEj[,2]; names(QLREj)=row.names(CRFIEj)
#names(QLREi)=c("QLREj")

Lst=list(CRTi=round(CRTi,3),CREiF1=round(CREiF1,3),CRFIEi=round(CRFIEi,
3),CRTj=round(CRTj,3),CREjF1=round(CREjF1,3),CRFIEj=round(CRFIEj,3),Q
LREi=round(QLREi,3),QLREj=round(QLREj,3))
}

```

#####FUNCTION COMPROMISE GRAPHIC#####

```

graphcomp=function (dimx,dimy,F,Q,biplot,compromise,HJB) {
  x=paste("Dimension",dimx,sep="")
  y=paste("Dimension",dimy,sep="")
  i=dimx; j=dimy
  maxxQ=max(abs(Q[,i])); maxyQ=max(abs(Q[,j]))
  maxxF=max(abs(F[,i])); maxyF=max(abs(F[,j]))
  if (biplot=="FALSE") {
    par(mfrow=c(1,2))
    plot(Q[,i],Q[,j],pch=20,col="white",main="Compromise",xlab=x,
ylab=y,xlim=c(-maxxQ,maxxQ),ylim=c(-maxyQ,maxyQ))
    text(Q[,i],Q[,j],labels=colnames(compromise))
    arrows(x0=0,y0=0,x1=Q[,i],y1=Q[,j],length=0.1)
    abline(h=0,v=0)
    plot(F[,i],F[,j],pch=20,col="white",main="Compromise",xlab=x,
ylab=y,xlim=c(-maxxF,maxxF),ylim=c(-maxyF,maxyF))
    text(F[,i],F[,j],labels=row.names(compromise),adj=1)
    arrows(x0=0,y0=0,x1=F[,i],y1=F[,j],length=0.1)
    abline(h=0,v=0) }
  else {
    par(mfrow=c(1,1))

```

```

qq=cbind(Q[,i],Q[,j])
ff=cbind(F[,i],F[,j])
tt=rbind(qq,ff)
maxb=max(maxxQ,maxxF)
mayb=max(maxyQ,maxyF)
plot(qq[,1],qq[,2],pch=20,col="white",main="Compromise:HJBiplot",xlab=x,
ylab=y,xlim=c(-maxb,maxb),ylim=c(-mayb,mayb))
arrows(x0=0,y0=0,x1=qq[,1],y1=qq[,2],length=0.1,col="blue")
text(qq[,1],qq[,2],labels=colnames(compromise),col="blue",adj=1)
points(ff[,1],ff[,2],pch=16,cex=HJB$QLREi,col="black")
text(ff[,1],ff[,2],labels=row.names(compromise),cex=HJB$QLREi,adj=1)
abline(h=0,v=0)
}
}

```

#####FUNCTION INTRA-STRUCTURE GRAPHIC#####

```

graphintra=function(dimx=compX,dimY=compY,FK=FK,QK=QK,indexz1,biplot,
compromise=compromise,k=k) {
  dimx; dimy
  d1=dimx; d2=dimy
  x=paste("Dimension",dimx,sep="")
  y=paste("Dimension",dimy,sep="")
  if (biplot=="FALSE") {
    for (i in 1:k) {
      par(mfrow=c(1,2))
      tempFK=FK[,indexz1[i,2]:indexz1[i,3]]
      tempQK=QK[indexz1[i,2]:indexz1[i,3],]
      maxxQK=max(abs(tempQK[,d1])); maxyQK=max(abs(tempQK[,d2]))
      maxxFK=max(abs(tempFK[,d1])); maxyFK=max(abs(tempFK[,d2]))
      maxx=max(maxxQK,maxxFK); mayy=max(maxyQK,maxyFK)
    }
  }
}

```

```

    plot(tempQK[,d1],tempQK[,d2],pch=20,col="white",xlab=x,ylab=y,xlim =
c(-maxxQK,maxxQK),ylim=c(-maxyQK,maxyQK),main=indexz1[i,1])
    arrows(x0=0,y0=0,x1=tempQK[,d1],y1=tempQK[,d2],length=0.1)
    text(tempQK[,d1],tempQK[,d2],labels=colnames(compromise),adj=1)
    abline(h=0,v=0)
    plot(tempFK[,d1],tempFK[,d2],pch=20,col="white",xlab=x,ylab=y,xlim =
c(-maxxFK,maxxFK),ylim=c(-maxyFK,maxyFK),main=indexz1[i,1])
    arrows(x0=0,y0=0,x1=tempFK[,d1],y1=tempFK[,d2],length=0.1)
    text(tempFK[,d1],tempFK[,d2],labels=row.names(compromise),adj=1)
    abline(h=0,v=0) }
}
else {
  for(i in 1:k) {
    par(mfrow=c(1,1))
    tempFK=FK[indexz1[i,2]:indexz1[i,3]]; F=tempFK
    tempQK=QK[indexz1[i,2]:indexz1[i,3],]; Q=tempQK
    HJBK=hjb(biplot=TRUE, F=F,
Q=Q,dimx=d1,dimy=d2,compromise=compromise)
    qq=cbind(tempQK[,d1],tempQK[,d2])
    ff=cbind(tempFK[,d1],tempFK[,d2])
    tt=rbind(qq,ff)
    maxxQK=max(abs(tempQK[,d1])); maxyQK=max(abs(tempQK[,d2]))
    maxxFK=max(abs(tempFK[,d1])); maxyFK=max(abs(tempFK[,d2]))
    maxx=max(maxxQK,maxxFK); mayy=max(maxyQK,maxyFK)

plot(qq[,1],qq[,2],pch=20,col="white",main=paste(indexz1[i,1],"HJBiplot",sep="
:"),xlab=x,ylab=y,xlim=c(-maxx,maxx),ylim=c(-mayy,mayy))
    arrows(x0=0,y0=0,x1=qq[,1],y1=qq[,2],length=0.1,col="blue")
    text(qq[,1],qq[,2],labels=colnames(compromise),col="blue",adj=1)
    points(ff[,1],ff[,2],pch=16,cex=HJBK$QLREi,col="black")
    text(ff[,1],ff[,2],labels=row.names(compromise),cex=HJBK$QLREi,adj=1)

```

```

abline(h=0,v=0)
if(i==1){
  t1=0;t2=0;t3=0;t4=0;t5=0;t6=0;t7=0;t8=0
  t1=HJBK$CRTi; t2=HJBK$CREiF1; t3=HJBK$CRFIEi;
  t4=HJBK$CRTj; t5=HJBK$CREjF1; t6=HJBK$CRFIEj;
  t7=HJBK$QLREi; t8=HJBK$QLREj }
else{
  t1=c(t1,HJBK$CRTi); t2=rbind(t2,HJBK$CREiF1);
t3=rbind(t3,HJBK$CRFIEi);
  t4=c(t4,HJBK$CRTj); t5=rbind(t5,HJBK$CREjF1);
t6=rbind(t6,HJBK$CRFIEj);
  t7=c(t7,HJBK$QLREi); t8=c(t8,HJBK$QLREj) }}

Lst=list(CRTi=round(t1,3),CREiF1=round(t2,3),CRFIEi=round(t3,3),CRTj=roun
d(t4,3),CREjF1=round(t5,3),CRFIEj=round(t6,3),QLREi=round(t7,3),QLREj=ro
und(t8,3))
}
}

```

#####FUNCTION HJSTATICO#####

```

#preprocessing=c("B.partial","B.total","A.norma")
HJSTATICO=function(x=env,y=spe,condition=etiq$space,biplot=TRUE,preproc
essing="B.Total",compx=1,compy=2) {
tp=preprocessing
dimx=compx
dimy=compy
k=dim(table(condition))
dimenx=dim(x)
dimeny=dim(y)
n=dimenx[1]/k
p=dimenx[2]

```

```

q=dimeny[2]

prep=Prep.Tables(x=x,y=y,condition=condition,type=tp)
X=prep$X
Y=prep$Y
index=prep$index
#COINERTIA
for (i in 1:k) {
  dn=1/index[i,4]
  Dn=diag(dn,nrow=index[i,4], ncol=index[i,4])
  tempx=X[index[i,2]:index[i,3],]
  tempx=as.matrix(tempx)
  tempy=Y[index[i,2]:index[i,3],]
  tempy=as.matrix(tempy)
  tempz=t(tempy)%*%Dn%*%tempx
  if (i==1) {
    Z=tempz
    Z1=tempz}
  else {
    Z=rbind(Z,tempz)
    Z1=cbind(Z1,tempz) }
}

```

Z #matriz de coinercia que contiene las Zk tablas de coinercia

```

#crea matriz index guía para tablas Z - Coinertia
dimenz=dim(Z)
indexz=matrix(nrow=k,ncol=5)
indexz=as.data.frame(indexz)
colnames(indexz)=c("Tables","Start","Finish","rows","columns")
temp=dimenz[1]/k

```

```

indexz[1,1]=index[1,1]
indexz[1,2]=1
indexz[1,3]=temp
indexz[1,4]=(indexz[1,3]-indexz[1,2])+1
indexz[1,5]=dimenz[2]
for (i in 2:k) {
  indexz[i,1]=index[i,1]
  indexz[i,2]=indexz[i-1,3]+1
  indexz[i,3]=temp*i
  indexz[i,4]=(indexz[i,3]-indexz[i,2])+1
  indexz[i,5]=dimenz[2]
}
indexz

```

```

dimenz1=dim(Z1)
indexz1=matrix(nrow=k,ncol=5)
indexz1=as.data.frame(indexz1)
colnames(indexz1)=c("Tables","Start","Finish","rows","columns")
temp=dimenz1[2]/k
indexz1[1,1]=index[1,1]
indexz1[1,2]=1
indexz1[1,3]=temp
indexz1[1,4]=dimenz1[1]
indexz1[1,5]=dimenz1[2]/k
for (i in 2:k) {
  indexz1[i,1]=index[i,1]
  indexz1[i,2]=indexz1[i-1,3]+1
  indexz1[i,3]=temp*i
  indexz1[i,4]=dimenz1[1]
  indexz1[i,5]=dimenz1[2]/k
}

```

```
indexz1
```

```
#INTERSTRUCTURE-PTA
```

```
C=matrix(nrow=k,ncol=k)
```

```
for (i in 1:k) {
```

```
  for (j in 1:k) {
```

```
    temp1=Z[indexz[i,2]:indexz[i,3],]
```

```
    temp2=Z[indexz[j,2]:indexz[j,3],]
```

```
    C[i,j]=sum(temp1*temp2)
```

```
  }
```

```
}
```

```
C
```

```
#RV calculo
```

```
d=diag(C)
```

```
RV=matrix(nrow=k,ncol=k)
```

```
row.names(RV)=index[,1]
```

```
colnames(RV)=index[,1]
```

```
for (i in 1:k) {
```

```
  for (j in 1:k){
```

```
    RV[i,j]=C[i,j]/(sqrt(d[i])*sqrt(d[j]))
```

```
  }
```

```
}
```

```
myLabels=index[,1]
```

```
print("RV")
```

```
print(round(RV,3))
```

```
l=Rank(RV)
```

```
dvs=svd(RV,nu=min(l,l),nv=min(l,l))
```

```
dvs$d
```

```
dvs$u
```

```

U=dvs$u[,1:2]
if(min(U[,1])<0){
  U=U*-1}
GG=U%*%sqrt(diag(dvs$d[1:2]))
u1=(U[,1])
one=matrix(1,nrow=length(u1),ncol=1)
alfa=u1*((sum((u1)*one))^-1) #tables weight
tb1=matrix(nrow=length(dvs$d),ncol=4) #tabla de eigenvalues, %var y % var
acumulada
colnames(tb1)=c("No.", "Eigenvalues", "% Inercia", "% Acum Inercia")
tb1[,2]=round(t(dvs$d),3)
for (i in 1:length(dvs$d)) {
  tb1[i,1]=i
  tb1[i,3]=round(((dvs$d[i]/sum(dvs$d))*100),3)
  if (i==1){
    tb1[i,4]=round(tb1[i,3],3) }
  else {
    tb1[i,4]=round((tb1[i,3]+tb1[i-1,4]),3) }
}
print("Interstructure")
print(tb1)

#Obtener el compromiso
compromise=matrix(nrow=indexz[1,4],ncol=indexz[1,5])
for(i in 1:k) {
  temp=matrix(nrow=indexz[1,4],ncol=indexz[1,5])
  temp=Z[indexz[i,2]:indexz[i,3],]
  temp=as.matrix(temp)
  temp2=alfa[i]*temp #este si va por Abdi
  if (i==1){
    compromise=temp2 }
}

```

```

else{
  compromise=compromise+temp2
}
}

#obtener coseno2
C1=matrix(nrow=k,ncol=1)
for (i in 1:k) {
  temp1=Z[indexz[i,2]:indexz[i,3],]
  temp2=compromise
  C1[i,1]=sum(temp1*temp2)
}
C1
#coseno2 calculo
varcomp=sum(compromise*compromise)
cos2=matrix(nrow=k,ncol=1)
row.names(cos2)=index[,1]
for (i in 1:k) {
  cos2[i,1]=C1[i,1]/(sqrt(d[i])*sqrt(varcomp))
}

#analisis compromiso
dimenco=dim(compromise)
csvd=svd(compromise,nu=min(Rank(compromise),Rank(compromise)),nv=min(
Rank(compromise),Rank(compromise)))
delta=diag(csvd$d)
F=compromise%*%csvd$v
Q=t(compromise)%*%csvd$u

delta2=(csvd$d)^2
tb2=matrix(nrow=length(csvd$d),ncol=4)

```

```

colnames(tb2)=c("No.", "Eigenvalues", "% Inercia", "% Acum Inercia")
for (i in 1:length(csvd$d)){
  tb2[i,1]=i
  tb2[i,2]=round(delta2[i],3)
  tb2[i,3]=round(((delta2[i]/sum(delta2))*100),3)
  if (i==1){
    tb2[i,4]=round(tb2[i,3],3) }
  else {
    tb2[i,4]=round((tb2[i,3]+tb2[i-1,4]),3) }
}
print("Compromise")
print(tb2)

```

```

tb11=matrix(nrow=k,ncol=3)
colnames(tb11)=c("Tables", "Weights", "Cos2")
for(i in 1:k){
  tb11[i,1]=index[i,1]
  tb11[i,2]=round(alfa[i],3) #lo correcto es alfa[i]
  tb11[i,3]=round(cos2[i,1],3) }
tb11=as.data.frame(tb11)
print("K-tables Indexs")
print(tb11)

```

#####Interestructure, Compromise, weigths-cos2 graphics#####

#SALIDAS GRAFICAS RESULTADOS

```

plothjstat=function(dimxc=1,dimyc=2){
  mycolorI=as.vector(rep("gray",length(dvs$d)))
  for (i in 1:length(dvs$d)) {
    if (i== 1) {
      mycolorI[i]="black" }
    if (i == 2){

```

```

    mycolorI[i]="black" }
}
mycolorC=as.vector(rep("gray",length(csvd$d)))
for (i in 1:length(csvd$d)) {
  if (i== dimxc) {
    mycolorC[i]="black" }
  if (i ==dimyc){
    mycolorC[i]="black" }
}
par(mfrow=c(2,2))
s.corcircle(cbind(GG[,1],GG[,2]),sub="Interstructure",possub =
"topright",csub=1.5,label="")
text(GG[,1]+0.08, GG[,2], labels=myLabels, cex=0.8,col="black")
plot(alfa,cos2,xlab="Weights",ylab="Cos2",pch=19,main
="Index",col="black",xlim=c(min(alfa)-0.01,max(alfa)+0.01),ylim=c(min(cos2)-
0.01,max(cos2)+0.01)) #no es u1, es alfa
text(alfa,cos2,labels=index[,1],adj=1)#no es u1, lo correcto es alfa, solo estoy
probando
barplot(tb1[,3],col=mycolorI,main ="Eigenvalues Inter-
structure",ylab="%",axes=TRUE,names.arg = tb1[,1])
barplot(tb2[,3],col=mycolorC,main="Eigenvalues Compromise",
ylab="%",axes=TRUE,names.arg = tb2[,1])
}
plothjstat(1,2)

```

```

HJB=hjb(biplot=biplot,F=F,Q=Q,dimx=compX,dimy=compY,compromise=comp
romise)

```

```

graphcomp(dimx=compX,dimy=compY,biplot=biplot,F=F,Q=Q,compromise=co
mpromise,HJB=HJB)

```

```

#codigo interactivo intra-estructura - SI

```

```

FK=matrix(nrow=length(Z1[,1]),ncol=k*length(csvd$v[,1]))
QK=matrix(nrow=k*length(Z1[,1]),ncol=length(csvd$u[,1]))
for(i in 1:k){
  tempf=matrix(nrow=indexz1[i,4],ncol=length(csvd$v[,1]))
  tempf=(Z1[,indexz1[i,2]:indexz1[i,3]])%*%csvd$v
  tempq=matrix(nrow=indexz1[i,5],ncol=length(csvd$u[,1]))
  tempq=t(Z1[,indexz1[i,2]:indexz1[i,3]])%*%(csvd$u)
  if(i==1) {
    FK=tempf
    QK=tempq }
  else {
    FK=cbind(FK,tempf)
    QK=rbind(QK,tempq) }
}

```

```

FK
dim(FK)
QK
dim(QK)

```

#grafico de las trayectorias / intra-estructuras

```

intra=graphintra(dimx=compX,dimy=compY,FK=FK,QK=QK,biplot=biplot,inde
xz1=indexz1,compromise=compromise,k=k)
if (biplot=="FALSE") {

```

```

Lst=list(X=round(X,3),Y=round(Y,3),CrossTables=round(Z,3),Compromise=rou
nd(compromise,3))}
else {

```

```

Lst=list(X=round(X,3),Y=round(Y,3),CrossTables=round(Z,3),Compromise=rou
nd(compromise,3),HJB=HJB,HJBK=intra) }

```

```
}
```

```
resultado=HJSTATICO(x=env,y=spe,condition=eti$space,biplot=TRUE,prepro  
cessing = "A.Norma",compx=1,compy=2)  
resultado$Compromise  
resultado$HJB  
resultado$HJBK  
resultado$CrossTables  
resultado$X  
resultado$Y
```

5.2 Librería MixSTATICO utilizando el software R.

Para el desarrollo del código que permite realizar el análisis de datos mixtos aplicando el MixSTATICO, se utilizó el programa RStudio (versión 1.3.1073) con las funciones margarita (distancia de Gower) y svd (DVS). El código fuente desarrollado está disponible en el **Capítulo 5.2.2**.

Inicialmente el usuario debe asegurarse de ordenar el conjunto de matrices en orden yuxtapuesto vertical según la variable que defina la tercera vía y deben estar almacenados en tres archivos: un archivo para los datos ambientales, el segundo para los datos de especies y el tercero que contiene las etiquetas (variable factor que defina la tercera-vía). Los dos primeros archivos solo deben contener los datos de las variables (ver Anexos A9, A10 y A11).

5.2.1 Algoritmo MixSTATICO

1. Ingresar los dos conjuntos de matrices $\mathbb{X}'_{[NP',K]}$ y $\mathbb{Y}'_{[NQ',K]}$ previamente preprocesadas.
2. Concatenar por filas todas las matrices $\mathbf{X}'_{[NP',k]}$ y $\mathbf{Y}'_{[NQ',k]}$.

Para $k=1$ hasta K

$$\{ \mathbf{XY}_{[Nr',k]} = \text{concatenar} \left(\mathbf{X}'_{[Np',k]}, \mathbf{Y}'_{[Nq',k]} \right) \}; \text{ donde } r' = p' + q'$$

3. Calcular la distancia de Gower, considerando las Ecs. (5.6), (5.7) y (5.8).

Para $k=1$ hasta K

$$\{ \mathbb{Z}_{[QPK]} = \text{gower}(\mathbf{XY}_{[Nr',k]}) \}$$

4. Aplicar un AFM a $\mathbb{Z}_{[QPK]}$ para normalizar

Para $k=1$ hasta K

DVS ($\mathbf{Z}_{[QP,k]}$) para obtener $\lambda_{[k]}$ y multiplicar cada elemento de la matriz por

$$\mathbf{1}/\lambda_{[k]}$$

$$\{ \mathbb{Z}'_{[QPK]} = \frac{1}{\lambda_{[k]}} \mathbf{Z}_{[QP,k]} \}$$

5. Vectorizar $\mathbb{Z}'_{[QPK]}$

$$\mathbb{Z}''_{[s'K]} = [\text{Vec}(\mathbf{Z}'_{[QP,1]}), \text{Vec}(\mathbf{Z}'_{[QP,2]}), \dots, \text{Vec}(\mathbf{Z}'_{[QP,K]})]^T; \text{ donde } s' = Q \times$$

P

6. Calcular $\mathbf{C}_{[KK]}$

$$\mathbf{C}_{[KK]} = \mathbb{Z}''_{[s'K]} \times \mathbb{Z}''^T_{[s'K]}$$

7. Calcular $\mathbf{RV}_{[KK]}$

Para $k=1$ hasta K

$$\{ \mathbf{RV}_{[KK]} = \frac{c_{[kk']}}{\sqrt{c_{[kk]} \times \sqrt{c_{[k'k']}}} \}; \text{ con } k = 1, \dots, K \text{ y } k' = 1, \dots, K$$

8. DVS ($\mathbf{RV}_{[KK]}$) para obtener $\vec{\alpha}_{[K]}$

9. Calcular la matriz compromiso

$$\mathbf{W}_{[QP]} = \sum_{k=1}^K \alpha_{[k]} \mathbf{Z}'_{[QP,k]}$$

10. DVS ($\mathbf{W}_{[QP]}$).

11. Generar el gráfico del compromiso usando PCoA.

12. Generar el gráfico de la intra-estructura usando PCoA.

5.2.2 Código fuente *MixSTATICO*

```
library(openxlsx)
library(pracma)
library(ade4)
library(MFAG)
library(MVar)
library(cluster)
library(StatMatch)

x=read.xlsx("MIXSTATICO2013.xlsx",sheet="ENV_stand_all")
y=read.xlsx("MIXSTATICO2013.xlsx",sheet="SPE_log_all")
etiq=read.xlsx("MIXSTATICO2013.xlsx",sheet="LABEL_all")

condition=etiq$time
k=dim(table(condition))
#Index matrix environment and species
index=matrix(nrow=k,ncol=4)
index=as.data.frame(index)
colnames(index)=c("Tables","Start","Finish","n")
index[1,1]=condition[1]
index[1,2]=1
i=1; count=1
s=1; num=1
for (i in 2:length(condition)) {
  if(condition[i]!=index[s,1]) {
    index[s,3]=count
    index[s,4]=num
    s=s+1
    num=1
    index[s,1]=condition[i]
    count=count+1
  }
}
```

```

    index[s,2]=count}
else {
    count=count+1
    num=num+1
    index[s,3]=count
    index[s,4]=num
}
}
index

```

#yuxtaponer las matrices x1 y y1 (contienen el AFM de cada subtabla cuanti y cuali)

#aplicamos distancia de gower a las variables

```
xy=cbind(x,y)
```

```
for (i in 1:k) {
```

```
    tempx=xy[index[1,2]:index[1,3],]
```

```
    tempx=as.matrix(tempx)
```

```
    d=daisy(t(tempx),"gower")
```

```
    if (i==1){D=d}
```

```
    else {D=rbind(D,d)}
```

```
}
```

```
D
```

```
s=read.xlsx("MIXSTATICO2013.xlsx",sheet="costruc_all")
```

#crea matriz index guía para tablas s - gower

```
dimx=dim(x)
```

```
n=dim(s)
```

```
dimy=n[1]/k
```

```
nam=s[1:dimy,1]
```

```
etiq=s[,1]
```

```
s=s[,-1]
```

```

dimenz=dim(s)
indexz=matrix(nrow=k,ncol=5)
indexz=as.data.frame(indexz)
colnames(indexz)=c("Tables","Start","Finish","rows","columns")
temp=dimenz[1]/k
indexz[1,1]=index[1,1]
indexz[1,2]=1
indexz[1,3]=temp
indexz[1,4]=(indexz[1,3]-indexz[1,2])+1
indexz[1,5]=dimx[2]
for (i in 2:k) {
  indexz[i,1]=index[i,1]
  indexz[i,2]=indexz[i-1,3]+1
  indexz[i,3]=temp*i
  indexz[i,4]=(indexz[i,3]-indexz[i,2])+1
  indexz[i,5]=dimx[2]
}
Indexz

#AFM para cada subtabla cuanti y cuali en X y en Y
#se obtiene el lambda(1) que multiplica a cada subtabla analizada
#como lambda(1)^-1
for (i in 1:k) {
  tempx=s[indexz[i,2]:indexz[i,3],]
  tempx=as.matrix(tempx);dimt=dim(tempx)
  dvs=svd(tempx,nu=min(Rank(tempx),Rank(tempx)),nv=min(Rank(tempx),Rank
(tempx)))
  s1=tempx*(1/(dvs$d[1]^2))
  if (i==1){sw=s1}
  else {sw=rbind(sw,s1)}
}

```

```

sw
#INTERSTRUCTURE-PTA
#obtención de C vectorizando la matriz sw
r=dim(sw)
Z=matrix(nrow=(indexz[1,4]*indexz[1,5]),ncol=k)
j=1; col=1; fi=1;count=0;ki=1
for (l in 1:k){
for (i in 1:indexz[ki,5]) {
for (j in indexz[ki,2]:indexz[ki,3]) {
Z[fi,col]=sw[j,i]
fi=fi+1;
count=count+1
#print(Z); print(count)
}
if (count==(indexz[1,4]*indexz[1,5])) { #count==r[1]
count=0
col=col+1
fi=1
ki=ki+1
}
}}
#colnames(S)
Z
ZZ=t(Z)
C1=ZZ%*%t(ZZ)
C1

d=diag(C1)
RV=matrix(nrow=k,ncol=k)
row.names(RV)=index[,1]

```

```

colnames(RV)=index[,1]
for (i in 1:k) {
  for (j in 1:k){
    RV[i,j]=C1[i,j]/(sqrt(d[i])*sqrt(d[j]))
  }
}
RV
myLabels=index[,1]
print(RV)
l=Rank(RV)
dvs=svd(RV,nu=min(l,l),nv=min(l,l))
dvs$d
dvs$u
U=dvs$u[,1:2]
if(min(U[,1])<0){
  U=U*-1}
GG=U%*%sqrt(diag(dvs$d[1:2]))
u1=(U[,1])
one=matrix(1,nrow=length(u1),ncol=1)
alfa=u1*((sum((u1)*one))^-1) #tables weight
tb1=matrix(nrow=length(dvs$d),ncol=4) #tabla eigenvalues, %var y % var
acumulada
colnames(tb1)=c("No.", "Eigenvalues", "% Inercia", "% Acum Inercia")
for (i in 1:length(dvs$d)) {
  tb1[i,1]=i
  tb1[i,2]=round((dvs$d[i]^2),3)
  tb1[i,3]=round(((dvs$d[i]/sum(dvs$d))*100),3)
  if (i==1){
    tb1[i,4]=round(tb1[i,3],3) }
  else {
    tb1[i,4]=round(tb1[i,3]+tb1[i-1,4],3) }
}

```

```

}
print("Interstructure")
print(tb1)

par(mfrow=c(1,1))
s.corcircle(cbind(GG[,1],GG[,2]),sub="Interstructure",possub =
"topright",csub=1.5,label="")
text(GG[,1]+0.15, GG[,2], labels=myLabels, cex=0.5,col="black")

Z1=0
for (i in 1:k){
  temp=sw[indexz[i,2]:indexz[i,3],]
  if (i==1){Z1=temp}
  else {Z1=cbind(Z1,temp)}
}
Z1

dimenz1=dim(Z1)
indexz1=matrix(nrow=k,ncol=5)
indexz1=as.data.frame(indexz1)
colnames(indexz1)=c("Tables","Start","Finish","rows","columns")
temp=dimenz1[2]/k
indexz1[1,1]=index[1,1]
indexz1[1,2]=1
indexz1[1,3]=temp
indexz1[1,4]=dimenz1[1]
indexz1[1,5]=dimenz1[2]/k
for (i in 2:k) {
  indexz1[i,1]=index[i,1]
  indexz1[i,2]=indexz1[i-1,3]+1
  indexz1[i,3]=temp*i
}

```

```

indexz1[i,4]=dimenz1[1]
indexz1[i,5]=dimenz1[2]/k
}
indexz1

#compromiso
compromise=matrix(nrow=indexz1[1,4],ncol=indexz1[1,5])
for(i in 1:k) {
  temp=matrix(nrow=indexz[1,4],ncol=indexz[1,5])
  temp=sw[indexz[i,2]:indexz[i,3],]
  temp=as.matrix(temp)
  temp2=alfa[i]*temp
  if (i==1){
    compromise=temp2 }
  else{
    compromise=compromise+temp2
  }
}
compromise
colnames(compromise)=colnames(x)
row.names(compromise)=nam

#obtener coseno2
C11=matrix(nrow=k,ncol=1)#cov entre la ktabla y el compromiso
dd=matrix(nrow=k,ncol=1)
for (i in 1:k) {
  temp1=sw[indexz[i,2]:indexz[i,3],]
  dd[i,1]=sum(temp1*temp1)
  temp2=compromise
  C11[i,1]=sum(temp1*temp2)
}

```

C11

dd

```
#coseno2 calculo
```

```
varcomp=sum(compromise*compromise)
```

```
cos2=matrix(nrow=k,ncol=1)
```

```
row.names(cos2)=index[,1]
```

```
for (i in 1:k) {
```

```
  cos2[i,1]=C11[i,1]/(sqrt(d[i])*sqrt(varcomp))
```

```
}
```

```
dimenco=dim(compromise)
```

```
csvd=svd(compromise,nu=min(Rank(compromise),Rank(compromise)),nv=min(
Rank(compromise),Rank(compromise)))
```

```
delta=diag(csvd$d)
```

```
F=compromise%*%csvd$v
```

```
Q=t(compromise)%*%csvd$u
```

```
#summary compromise
```

```
delta2=(csvd$d)^2
```

```
tb2=matrix(nrow=length(csvd$d),ncol=4)
```

```
colnames(tb2)=c("No.", "Eigenvalues", "% Inercia", "% Acum Inercia")
```

```
for (i in 1:length(csvd$d)){
```

```
  tb2[i,1]=i
```

```
  tb2[i,2]=delta2[i]
```

```
  tb2[i,3]=(delta2[i]/sum(delta2))*100
```

```
  if (i==1){
```

```
    tb2[i,4]=tb2[i,3] }
```

```
  else {
```

```
    tb2[i,4]=tb2[i,3]+tb2[i-1,4] }
```

```
}
```

```
print("Compromise")
```

```

print(tb2)

tb11=matrix(nrow=k,ncol=3)
colnames(tb11)=c("Tables","Weights","Cos2")
for(i in 1:k){
  tb11[i,1]=index[i,1]
  tb11[i,2]=round(alfa[i],3)
  tb11[i,3]=round(cos2[i,1],3) }
tb11=as.data.frame(tb11)
print("K-tables Indexs")
print(tb11)

#####Interestructure, Compromise, weigths-cos2 graphics#####
#SALIDAS GRAFICAS RESULTADOS
plothjstat=function(dimxc=1,dimyc=2){
  mycolorI=as.vector(rep("gray",length(dvs$d)))
  for (i in 1:length(dvs$d)) {
    if (i== 1)  {
      mycolorI[i]="black" }
    if (i == 2){
      mycolorI[i]="black" }
  }
  mycolorC=as.vector(rep("gray",length(csvd$d)))
  for (i in 1:length(csvd$d)) {
    if (i== dimxc)  {
      mycolorC[i]="black" }
    if (i ==dimyc){
      mycolorC[i]="black" }
  }
  par(mfrow=c(2,2))

```

```

s.corcircle(cbind(GG[,1],GG[,2]),sub="Interstructure",possub =
"topright",csub=1.5,label="")
text(GG[,1]+0.08, GG[,2], labels=myLabels, cex=0.8,col="black")
plot(alfa,cos2,xlab="Weights",ylab="Cos2",cex=cos2,pch=".",main
="Index",col="black")
text(alfa,cos2,labels=index[,1])
barplot(tb1[,3],col=mycolorI,main="Eigenvalues Inter-
structure",ylab="%",axes=TRUE,names.arg = tb1[,1])
barplot(tb2[,3],col=mycolorC,main="Eigenvalues Compromise",
ylab="%",axes=TRUE,names.arg = tb2[,1])
}
plothjstat(1,2)

```

```

#graficos con coordenadas principales ACop
dimc=dim(compromise)
#REALIZANDO DOBLE CENTRADO
#A LA MATRIZ COMPROMISO
meanrow=apply(compromise,1,mean)
meancol=apply(compromise,2,mean)
meantot=mean(compromise)
compcenter=matrix(nrow=dimc[1],ncol=dimc[2])
for(i in 1:dimc[1]) {
  for(j in 1:dimc[2]){
    compcenter[i,j]=compromise[i,j]-meanrow[i]-meancol[j]+meantot
  }
}
compcenter
di=dim(compcenter)
desp=svd(compcenter,nu=min(di[1],di[2]),nv=min(di[1],di[2]))
Zacoopf=desp$u%*%diag(sqrt(desp$d))
Zaccopc=desp$v%*%diag(sqrt(desp$d))

```

```

#graficando el compromiso
i=1;j=2
maxxQ=max(abs(Zaccopc[,i])); maxyQ=max(abs(Zaccopc[,j]))
maxxF=max(abs(Zacoopf[,i])); maxyF=max(abs(Zacoopf[,j]))
par(mfrow=c(1,2))
plot(Zaccopc[,i],Zaccopc[,j],pch=".",col="black",main="Environment",xlab="CoP 1", ylab="CoP 2",xlim=c(-maxxQ,maxxQ),ylim=c(-maxyQ,maxyQ))
text(Zaccopc[,i],Zaccopc[,j],labels=colnames(compromise))
arrows(x0=0,y0=0,x1=Zaccopc[,i],y1=Zaccopc[,j],length=0.07)
abline(h=0,v=0)
plot(Zacoopf[,i],Zacoopf[,j],pch=".",col="black",main="Species",xlab="CoP 1", ylab="CoP 2",xlim=c(-maxxF,maxxF),ylim=c(-maxyF,maxyF))
text(Zacoopf[,i],Zacoopf[,j],labels=row.names(compromise))
arrows(x0=0,y0=0,x1=Zacoopf[,i],y1=Zacoopf[,j],length=0.07)
abline(h=0,v=0)

```

```

#graficar intra-estructura
#centrando cada k tabla sw
Z1center=matrix(nrow=dimenz1[1],ncol=dimenz1[2])
Zcenter=matrix(nrow=dimenz[1],ncol=dimenz[2])
for (i in 1:k){
  temp=Z1[,indexz1[i,2]:indexz1[i,3]]
  mf=apply(temp,1,mean)
  mc=apply(temp,2,mean)
  mt=mean(temp)
  dimzs=dim(temp)
  temp2=matrix(nrow=dimzs[1],ncol=dimzs[2])
  for (l in 1:dimzs[1]){
    for(j in 1:dimzs[2]){
      temp2[l,j]=temp[l,j]-mf[l]-mc[j]+mt
    }
  }
}

```

```

    }
    if (i==1){Z1center=temp2;Zcenter=temp2}
    else{
      Z1center=cbind(Z1center,temp2)
      Zcenter=rbind(Zcenter,temp2)}
  }
Z1center
Zcenter

FK=matrix(nrow=length(Z1center[,1]),ncol=k*length(desp$v[,1]))
QK=matrix(nrow=k*length(Z1center[,1]),ncol=length(desp$u[,1]))
for(i in 1:k){
  tempf=matrix(nrow=indexz1[i,4],ncol=length(desp$v[,1]))
  tempf=(Z1center[,indexz1[i,2]:indexz1[i,3]])%*%desp$v
  tempq=matrix(nrow=indexz1[i,5],ncol=length(desp$u[,1]))
  tempq=t(Z1center[,indexz1[i,2]:indexz1[i,3]])%*%(desp$u)
  if(i==1) {
    FK=tempf
    QK=tempq }
  else {
    FK=cbind(FK,tempf)
    QK=rbind(QK,tempq) }
}

FK
dim(FK)
QK
dim(QK)

d1=1; d2=2

```

```

dimx=d1; dimy=d2
x=paste("CoP",dimx,sep="")
y=paste("CoP",dimy,sep="")
for (i in 1:k) {
  par(mfrow=c(1,2))
  tempFK=FK[,indexz1[i,2]:indexz1[i,3]]
  tempQK=QK[,indexz1[i,2]:indexz1[i,3],]
  maxxQK=max(abs(tempQK[,d1])); maxyQK=max(abs(tempQK[,d2]))
  maxxFK=max(abs(tempFK[,d1])); maxyFK=max(abs(tempFK[,d2]))
  maxx=max(maxxQK,maxxFK); mayy=max(maxyQK,maxyFK)
  plot(tempQK[,d1],tempQK[,d2],pch=".",col="black",xlab=x,ylab=y,xlim = c(-
maxxQK,maxxQK),ylim=c(-maxyQK,maxyQK),main=indexz1[i,1])
  arrows(x0=0,y0=0,x1=tempQK[,d1],y1=tempQK[,d2],length=0.07)
  text(tempQK[,d1],tempQK[,d2],labels=colnames(compromise))
  abline(h=0,v=0)
  plot(tempFK[,d1],tempFK[,d2],pch=".",col="black",xlab=x,ylab=y,xlim = c(-
maxxFK,maxxFK),ylim=c(-maxyFK,maxyFK),main=indexz1[i,1])
  arrows(x0=0,y0=0,x1=tempFK[,d1],y1=tempFK[,d2],length=0.07)
  text(tempFK[,d1],tempFK[,d2],labels=row.names(compromise))
  abline(h=0,v=0) }

```


PARTE III

RESUMEN Y CONCLUSIONES EN INGLÉS

*Cuando pases por aguas profundas,
yo estaré contigo.
cuando pases por ríos de dificultad,
no te ahogaras*
Isaías 43:2

RESUMEN Y CONCLUSIONES EN INGLÉS

SUMMARY

Over time, scientific research has been one of the primary sources of data generation in the various fields of study. The fundamental purpose was to promote knowledge and provide solutions to real problems.

Through the data, it is possible to examine characteristics such as the origin, behavior, and effect of events or phenomena of interest by applying statistical techniques. Thus, data can be arranged as a matrix, with individuals in the rows (observations) and variables (measured in those individuals) in the columns. As this type of data is usually large (observations and variables), classical data analysis methods are insufficient and lacking to produce reliable results. Therefore, multivariate statistical analysis proves to be a crucial tool for the analysis of this type of data. In this way it is possible to analyze multiple variables simultaneously.

On the other hand, in addition to the large volume of this type of information, there is also the problem associated with the different sources of variability (such as spatial effect and evolution over time) that usually characterize these data. In this sense, when structuring the data matrix-wise, the structure always has more than two dimensions. Thus, in the field of multivariate statistical analysis, multi-way methods prove to be the most efficient techniques for studying this type of data. Among the various methods developed in the literature, those of the STATIS family stand out clearly.

STATIS methods include:

- The analysis of a set of data arrays (cube) with three dimensions-ways;
- Consisting of individuals (dimension 1);
- Variables (dimension 2);
- Conditions (dimension 3) which are the K-arrays that contain data from the different measurements in space and/or time.

These methods are based on calculating the average structure of the K-arrays and representing it in a low dimensional space. The interpretation of the results depends on the researcher subjective perception when observing the projection of individuals or variables on a bi-factorial plane. It should note that these methods analyze only one type of variables (that is, quantitative or qualitative).

Of the various research areas that generate a large volume of information with a three-way structure, in time or/and space, ecology stands out. Thus, in this investigation it is intended to answer two crucial problems that researchers usually face, that is, (i) quantify the amount of information expressed by variables when explaining the relationship between two three-way matrices and (ii) jointly analyze quantitative and qualitative data.

To solve these problems, this research aims to propose two new methodologies, namely:

- (i) the HJ-STATICO method, together with its computer tool, which is the result of the combination of STATICO and Biplot methods;

- (ii) the MixSTATICO method, that jointly analyzes quantitative and qualitative data, together with its computer tool, which is an extension of the STATICO methodology for the analysis of mixed variables.

Both methodologies were applied to marine environmental data (species and environmental parameters) sampled monthly, for different depths, during the period 2013-2015, for four sampling stations, in the coastal profile of Continental Ecuador.

HJ-STATICO allowed us through the analysis of the compromise, to identify the average structure of the species-environment relationship of the region under study. The method identified the environmental distribution of the species, according to their abundance. Besides, it was possible to identify species that highlighted a stronger relationship with the environmental variations of an extreme warm event, such as the centric diatom *Lauderia borealis* (e13), the Dinoflagellates *Gymnodinium* sp. (e21) and *Gyrodinium* sp. (e22) and the *Mesodinium rubrum* (e23) ciliate having the highest quality of representation ($QLRE_i \geq 0.966$) of all species. Also, it was possible to highlight the environmental parameters that most amounts of variance expressed, namely nitrate and silicate ($QLRE_j \geq 0.976$). A similar analysis for each sampling station was obtained with the intra-structure.

Concerning MixSTATICO, the results expressed the evolution over time of the species-environment relationship during the occurrence of a normal event in the region under study. The quantitative variables that have been used were the abundance of species, as well as the concentrations of environmental parameters. As qualitative variables, the categories "above" and "below" were used, in relation to the monthly average samples of species abundance and concentrations of environmental parameters, for the sampling stations. As a result, the method

showed a segmentation of the species-environment relationship's behavior into two groups, associated with the seasonal time of the region (dry and rainy season). In addition, MixSTATICO, highlighted that the presence of the climatic season influences the species-environment relationship. For method verification purposes, we used the same variables represented quantitatively and qualitatively, resulting in that they are projected very close to each other on the bi-factorial plane.

Objective

This thesis has as its general objective to propose new multivariate methodological contributions to analyze three-way data matrices. These contributions are decomposing in:

Specifics objectives

- The first one, consists in contributing to the analysis of the stable part of the common structure of two three-way data matrices by combining in one analysis the advantages of the STATICO method ([Simier et al., 1999](#); [Thioulouse et al., 2004](#)) and the HJ-Biplot ([Galindo-Villardón, 1986](#); [Galindo-Villardón & Cuadras, 1986](#)). Therefore, series of indices (contributions and qualities of representation) can be calculated, in order to provide an objective interpretation of the results of the analysis. In addition, it is intended to enable joint representation, in the same reference system and with the highest quality of representation, both for lines and columns (which in the context of this investigation are species and environmental variables).
- The second objective is to develop a freely available source code, using the open-source software RStudio ([R Core Team and contributors worldwide](#),

2018). In this way, it is intended to provide access to a tool that facilitates the interpretation of the results obtained, supported by graphs and indices that increase knowledge about the relationships between pairs of three-way matrices and, consequently, the increase in the research area in the technique is applied.

- The third objective of this research aims to propose a new methodology as an extension of STATICO, which allows analyzing the stable part of the co-structure of two three-way data matrices that are conformed by a mixture of variables measured at different scales. Thus, it will be possible to analyze quantitative and qualitative variables simultaneously, as well as to represent them in the habitual way that the STATICO method does.

CONCLUSIONS

1. The exhaustive bibliographic review shows the high demand for multivariate data analysis methods in Ecology and the importance of applying three-way techniques to study the behaviors of biotic and abiotic parameters as well as their relationships.
2. The numerous applications of the three-way methods that have carried out in recent years, have shown that the interpretation of the distribution of the environmental pattern (either from the point of view of species, environmental parameters or the relationship between both), relies, in most cases, on the visualization of graphic outputs. Therefore, quantification of the co-structure magnitude that can reveal the relationship between species-environmental is commonly overlooked.
3. The methodology proposed in this work, which we call HJ-STATICO, provides information on the average image of the co-structure between two three-way arrays. In addition, HJ-STATICO provides other options that increase the interpretation of results, through the quantification of the contributions and quality of representation of the species and variables in the compromise and intra-structure. These outputs provide the possibility of proposing biological indicators that highlight relevant associations between species and environment, as well as the species' response to environmental variability.

4. The source code for the HJSTATICO library was developed, in open-access RStudio, allowing all researchers to apply it and to obtain results.
5. With the application of HJ-STATICO to real data, it was possible to demonstrate the quantification of the magnitude of the relationship between species and the environment, underlying the stronger species-environment relationships such as dinoflagellates and ciliates throughout the coastal profile during the presence of one of the warmest events that have occurred recently.
6. The MixSTATICO method, proposed in this work, is an alternative to STATICO that allows the simultaneous use of quantitative and qualitative variables in two three-way arrays. This method bases its analysis on distance matrices, similar to DISTATIS method. In addition, shows a graphic representation of inter-structure, compromise and intra-structure on a bi-factorial plane using PCoA.
7. Through its application to real data, this work has revealed how the MixSTATICO method allows finding the average image of the co-structure between the environment and species, analyzing mixed data and its evolution over different conditions.
8. The method proved to be suitable for the simultaneous analysis of mixed data (that is, of a quantitative and qualitative nature, simultaneously), since the projections of two variables (which had the same information) proved to be very close.

9. To facilitate the use of MixSTATICO, a library was developed in open-access RStudio, called MixSTATICO. This will allow all researchers to apply it and validate their results.

CONTRIBUCIONES CIENTÍFICAS

*No temas porque yo estoy contigo;
no desmayes, porque
yo soy tu Dios
que te esfuerzo;
siempre te ayudaré,
siempre te sustentaré con
la diestra de mi justicia
Isaías 41:10*



Article

Study of Temporal Variations in Species–Environment Association through an Innovative Multivariate Method: MixSTATICO

Mariela González-Narváez ^{1,2,*}, María José Fernández-Gómez ^{3,4}, Susana Mendes ⁵, José-Luis Molina ⁶, Omar Ruiz-Barzola ² and Purificación Galindo-Villardón ^{1,4}

- ¹ Statistics Department, Faculty of Medicine, University of Salamanca (USAL), 37007 Salamanca, Spain; pgalindo@usal.es
 - ² Faculty of Life Sciences, Campus Gustavo Galindo, ESPOL Polytechnic University, Km 30.5 Vía Perimetral, Guayaquil 090902, Ecuador; oruiz@espol.edu.ec
 - ³ Statistics Department, Faculty of Biology, University of Salamanca (USAL), 37007 Salamanca, Spain; mjfg@usal.es
 - ⁴ Statistics Department, Institute for Biomedical Research (IBSAL), 37007 Salamanca, Spain
 - ⁵ MARE—Marine and Environmental Sciences Centre, ESTM, Polytechnic of Leiria, School of Tourism and Maritime Technology, 2520-641 Peniche, Portugal; susana.mendes@ipleiria.pt
 - ⁶ IGA Research Group, Higher Polytechnic School of Ávila, University of Salamanca (USAL), 50 Avenue Homos Caleros, 05003 Ávila, Spain; jlmolina@usal.es
- * Correspondence: margonzal@espol.edu.ec; Tel.: (+593)-982859868



Citation: González-Narváez, M.; Fernández-Gómez, M.J.; Mendes, S.; Molina, J.-L.; Ruiz-Barzola, O.; Galindo-Villardón, P. Study of Temporal Variations in Species–Environment Association through an Innovative Multivariate Method: MixSTATICO. *Sustainability* **2021**, *13*, 5924. <https://doi.org/10.3390/su13115924>

Academic Editor: Just T. Bayle-Sempere

Received: 25 March 2021
Accepted: 19 May 2021
Published: 24 May 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The study of biotic and abiotic factors and their interrelationships is essential in the preservation of sustainable marine ecosystems and for understanding the impact that climate change can have on different species. For instance, phytoplankton are extremely vulnerable to environmental changes and thus studying the factors involved is important for the species’ conservation. This work examines the relationship between phytoplankton and environmental parameters of the eastern equatorial Pacific, known as one of the most biologically rich regions in the world. For this purpose, a new multivariate method called MixSTATICO has been developed, allowing mixed-type data structured in two different groups (environment and species) to be related and measured on a space–time scale. The results obtained show how seasons have an impact on species–environment relations, with the most significant association occurring in November and the weakest during the month of May (change of season). The species *Lauderia borealis*, *Chaetoceros didymus* and *Gyrodinium* sp. were not observed in the coastal profiles during the dry season at most stations, while during the rainy season, the species *Dactyliosolen antarcticus*, *Proboscia alata* and *Skeletonema costatum* were not detected. Using MixSTATICO, species vulnerable to specific geographical locations and environmental variations were identified, making it possible to establish biological indicators for this region.

Keywords: MixSTATICO; mixed data; multi-way; species–environment; sustainability

1. Introduction

Water sustainability has often been treated as sustainable freshwater management for human consumption. However, from a more holistic point of view, water, both freshwater and saltwater, is an important resource and essential for sustaining ecosystems that generate or maintain the environmental conditions necessary for sustaining life in general [1]. The demand for water in rural agriculture, industrial operations, human consumption and other purposes is high, but in recent years, the water contamination caused by chemicals, pharmaceuticals and cleaning processes [2] have polluted rivers, creating serious problems in the marine ecosystems once this contamination reaches the sea.

For the United Nations (UN), water is an essential element in the adaptation to climate change [1], and is a significant problem that society and the environment are currently

experiencing. Climate change has led to alterations in the natural variability of the environment, which is reflected worryingly in extreme temperatures and precipitation, affecting pollution in the hydrological component due to the high nutrient values generated [3]. Environmental changes are considered one of the main elements that cause alterations in the environment and ecology, leading to vulnerability [4]. To assess ecological vulnerability, it is necessary to have quantitative and qualitative knowledge about ecosystems and to thus understand the factors that cause their vulnerability [5].

In the study of the hydrological component, mathematical models [6] have increasingly been applied through predictive and descriptive methods that are classified into deterministic and stochastic. Stochastics are characterized by using multivariate statistical methods [7]. These methods are considered helpful for studying water quality [8] using evaluation factors such as correlation, covariation, similarities and distances [7]. In addition, in order to analyse temporal hydrological behaviour from a stochastic perspective, it is necessary to study the temporal–spatial correlation of hydrological parameters [7].

In this way, it is of vital importance to study the sustainability of water associated with climate change, not necessarily in its most common environment, but rather from a holistic perspective.

It is known that climate change affects the sustainability of water and its ecosystem, whether inland waters, seawater or the open ocean, as well as the environment and society at large.

Multivariate statistical methods have been developed to study data matrices with multiple variables in order to reduce the dimensionality of a matrix. These methods have a long history in science. The first was Principal Component Analysis (PCA) [9–11], a technique that projects the observation matrix onto a bidimensional space built from several inter-correlated orthogonal quantitative variables called principal components. The components are linear combinations of the original variables. The first component represents the largest variance retained by the method, the second component represents the largest variance possible while orthogonal to the first, and so on for the other components.

Subsequent developments include Factorial Analysis (FA) [12,13], which projects the correlations between variables and among variables and factors; Correspondence Analysis (CA) [14,15], which studies the association between the categories of two qualitative variables; Multiple Correspondence Analysis (MCA) [16], which extends CA to several qualitative variables; and Canonical Correspondence Analysis (CCA) [17]. The latter is an extension of CA widely used in biology which relates community composition to known variation in the environment (two matrices). The ordination axes are linear combinations of the environmental variables.

These methods are useful in analysing data arrays with a two-way dimensional structure, as they have the constraint that the number of variables must be less than the number of individuals (PCA). However, this condition cannot be met in some areas of research that analyse several variables to describe communities and environments, and for ecological data sets that are for the most part multivariate [18]. At present, with the advancement of multivariate data analysis techniques, it is also possible to analyse data that consider multiple conditions, which give rise to arrays that have a k-way structure and multi-way data arrays. For the analysis of this new array set, multi-way multivariate methods are used that allow multiple conditions to be studied in combination, identifying underlying patterns that other methods are unable to identify. There are valuable contributions to these multivariate methods cited below.

1.1. Contributions of the French School

Multiple Factor Analysis (FMA) [19–21] analyses information from different variable sets (matrices) defined on the same set of individuals (observations). This method performs PCAs in two phases: first it normalizes each matrix (dividing its elements by the first eigenvalue obtained from applying a PCA to each matrix) and then merges the matrices into one that represents the overall structure, then a PCA is applied to that global matrix.

STATIS [22] is a technique used to study a set of K-way matrices and is also known as ACT-STATIS [23]. It analyses data matrices by calculating Euclidean distances between configurations of the same individuals measured by different variables and K conditions (times or situations). This method bases its analysis on operator matrices (covariance matrices); it is a generalization of PCA and begins by analysing the similarities between the K-matrices. From these, it obtains weights (first eigenvector) used to find the stable part of the common structure (compromise or consensus) of the set of matrices, which is calculated as the weighted linear combination of the K-matrices and represents it in a low dimension space. It also allows the original individuals of each k-matrix ($k = 1, \dots, K$) to be projected over the consensus dimensions obtained.

STATIS developments have had important and numerous contributions and can be classified according to the input data (same variables or individuals in the k-matrices), according to the weights assigned to the compromised matrix, when considering external information and if pairs of matrices are available in different situations or times. These methods started with the STATIS and STATIS-DUAL [22], to be continued with X-STATIS or Triadic Partial Analysis (PTA) [24], STATIS-CoA [25], STATICO (STATIS and COINERCIA) [26,27], DOUBLE STATIS (DO-ACT) [28], K + 1 STATIS [29], DISTATIS [30], STATIS-4 [31], Kernel-STATIS [32], COVSTATIS and COSTATIS (COINERCIA and STATIS) [33], POWER-STATIS [34], Anisotropic STATIS or ANISOSTATIS [35], STATIS-LDA [36], INTERSTATIS [37], Sir-STATIS [38], HiDISTATIS and DiDISTATIS [39], CATATIS [40], STATICO-CoA [41] and recently CLUSTATIS [42].

However, despite all these contributions, none of these methods analyse matrices with mixed data.

The methodological basis of this study is based on the methods of the French school, but it is also necessary to present the methods developed by other schools.

1.2. Contributions of the Anglo-Saxon School

There are the methods Tucker [43] and its variants Tucker 1 [44,45], Tucker 2 [45–47], Tucker 3 [43,47], PARAFAC/CANDECOMP [48,49], INDSCAL [49], Three Mode Scaling [44], IDIOSCAL [50], DEDICOM [51], TUCKALS2 and TUCKALS3 [47], CANDELINC [52], PARATUCK2 [53], PARARLIND [54], CONFAC [55], CP-RIDGE [56] and CP-LASSO [57].

1.3. Contributions of the Salmantina School

There are the methods METABLIPO [58,59], Multiple Biplot [60], CANOSTATIS [61], ACPR Three ways [62], Triadic Biplot [63], STATIS Dual Canónico [64], Biplot Consensus [65], Co-Tucker [66,67], Coupled Data AnalysisT3-PCA [68], Dynamic Biplot [69,70], Co-Tucker 3 [71,72] and C_{enet} Tucker [73].

STATIS methods are applied in different fields research as medicine [74,75], policy [76], quality control [77–80], sensory profiles [42,81–83], economy [84,85], customer research [86], education [87], quality of life [88], molecular biology [89–91] and other fields.

Among the numerous research works developed in similar fields to this study, these methods can be employed in environmental studies [41,66,92–98], ecology [99,100], sustainability [101–103] and hydrology [104,105].

This work proposes a multivariate alternative called MixSTATICO in order to analyse together the information provided by quantitative and qualitative variables, which allows us to identify, through a space–time analysis, the effect of variations of the physical and chemical variables of the ocean on certain phytoplanktonic species from a perspective similar to that of the STATICO method.

2. Materials and Methods

2.1. Materials

The study area consists of four sampling stations, located spatially from the north to the south of the coastal profile of Continental Ecuador, located 16 km from the coastline. The samples sites of Emeralds and Puerto Bolívar (Pto. Bolívar) are influenced by an

estuarine system, while those of Manta and La Libertad maintain a direct influence of the Pacific Ocean. The area has two climatic epochs: rainy (December–May) and dry (June–November).

Data were collected in 2013. At each station, multiple environmental and biological variables were collected monthly at seven depth levels of about 0, 10, 20, 30, 40, 50 and 75 m in the morning (8:00 a.m. to 12:00 p.m.) from February to December in 2013. For a further description of the data, consult González-Narváez [106].

2.1.1. Hydrography and Sampling

A vertical profile of conductivity–temperature–depth (CTD), SeaBird 9/11-plus was measured daily at a fixed position, providing temperature ($^{\circ}\text{C}$) and salinity (psu) profiles. Water samples were collected at the four sampling sites, using a cast of 3 L nontoxic sampling Van Dorn bottles to obtain water samples for dissolved oxygen concentration (mg L^{-1}) and analysis of nutrients including nitrite, nitrate, phosphate and silicate ($\mu\text{g at L}^{-1}$), as well as phytoplankton examination. To measure dissolved oxygen, the volumetric method was applied (PEE/LAB-DOQ/01) based on Eaton et al. [107].

2.1.2. Determination of Inorganic Nutrients and Ratios

Samples for nutrient analysis were filtered through washed glass-fibre filters with $0.45 \mu\text{m}$ (Whatman, GF/C). They were frozen for analysis 1–3 weeks later onshore. Ammonium, nitrite, nitrate, phosphate and silicic acid amounts were determined by the method of Strickland and Parsons [108].

2.1.3. Phytoplankton Analysis

Phytoplankton analysis was performed with approximately 250 cm^3 water samples in glass bottles fixed with Lugol's solution. $>A 25 \text{ cm}^3$ composite chamber was subsequently filled with the sample water and its contents allowed to settle for 24 h. At least two transects of the chamber bottom were observed with an inverted microscope [109] at $400\times$ magnification to count the small, frequently occurring phytoplankton forms. Additionally, the whole chamber bottom was examined at $125\times$ magnification to count larger, less frequent cells [110,111]. Data are expressed as cells per liter. Classification was performed at the genus or species level when possible, but many taxa could not be identified and were pooled into categories such as “small flagellates” or “small dinoflagellates”. Abbreviations were added for statistical reporting (see Table 1 for species lists); references used can be found in Jiménez [112], Pesantes [113], Balech [114], Tomas [115]), Taylor et al. [116], Tomas [117,118], Young et al. [119] and Jiménez [120].

Table 1. Environment variables (average \pm SD).

	Location (Both Seasons)				Season (Entire Coastal Profile)		Total
	North	Centre North	Centre South	South	Rainy	Dry	
Temperature (T — $^{\circ}\text{C}$)	25.77 ± 1.34	23.96 ± 1.18	22.31 ± 1.79	22.36 ± 1.03	24.03 ± 1.79	23.24 ± 2.04	23.6 ± 1.97
Salinity (S —psu)	32.92 ± 0.49	33.57 ± 0.38	34.03 ± 0.41	33.7 ± 0.72	33.49 ± 0.68	33.61 ± 0.63	33.56 ± 0.66
Dissolved Oxygen (DO — $\text{mg}\cdot\text{L}^{-1}$)	4.5 ± 0.33	4.38 ± 0.2	4.23 ± 0.61	4.16 ± 0.6	4.34 ± 0.49	4.3 ± 0.48	4.32 ± 0.49
Nitrate (NO_3^- — $\mu\text{g}\cdot\text{at}\cdot\text{L}^{-1}$)	1.28 ± 1.59	2.18 ± 1.81	3.13 ± 2.2	4.25 ± 2.74	3.53 ± 2.6	2.03 ± 1.97	2.71 ± 2.4
Nitrite (NO_2^- — $\mu\text{g}\cdot\text{at}\cdot\text{L}^{-1}$)	0.12 ± 0.11	0.17 ± 0.13	0.24 ± 0.15	0.27 ± 0.19	0.25 ± 0.18	0.16 ± 0.12	0.2 ± 0.16
Phosphate (PO_4^{3-} — $\mu\text{g}\cdot\text{at}\cdot\text{L}^{-1}$)	0.59 ± 0.53	0.67 ± 0.52	0.85 ± 0.31	0.69 ± 0.37	0.91 ± 0.51	0.53 ± 0.31	0.7 ± 0.45
Silicate (SiO_4^{4-} — $\mu\text{g}\cdot\text{at}\cdot\text{L}^{-1}$)	6.19 ± 4.31	5.33 ± 3.62	9.62 ± 3.76	9.01 ± 3.07	8.85 ± 4.3	6.44 ± 3.65	7.54 ± 4.14

Appendices B and C show the average values at 20 m (m) the depth that was initially analysed. The data are of mixed type and thus suitable for testing this method. There are two series of three-way matrices, $\bar{X}_{[ppk]}$ (environment) and $\bar{Y}_{[mqk]}$ (species), each with $k = 3$ (time—months). The individuals (rows) are the average values obtained within the first

20 m at four sampling stations, while the variables (columns) are the physical and chemical parameters and the abundances of phytoplankton species.

The species matrices include species abundance for *Stauroneis membranacea* (ordinal-qualitative: 0 if $y_{[ij]} < Q_1$; 1 if $Q_1 \leq y_{[ij]} < Q_2$; 2 if $Q_2 \leq y_{[ij]} < Q_3$ and 3 if $y_{[ij]} \geq Q_3$ for quartiles calculated using all data collected throughout the entire study period). As the last variable (nominal type), the location is used to indicate the sector on which the sampling station is situated: north for Esmeraldas, centre north for Manta, centre south for La Libertad and south for Pto. Bolívar. The seasons were indicated as “rainy from December to May” and “dry from June to November”.

2.2. Methodological Description of the MixSTATICO Method

Boldface lowercase letters (e.g., $\vec{v}_{[n,j,k]}$) will be used to represent vectors of dimensionality n corresponding to j -th column and k -th condition. Boldface uppercase letters $X_{[np,k]}$ will denote $(n \times p)$ matrices corresponding to the k -th condition. A set of K such matrices (a “data cube”) will be denoted by a boldface uppercase double-stroke letter such as $\mathbb{X}_{[npK]}$. The positions of matrix elements will be represented by boldface lowercase letters $\{x_{ijk} | i = 1, \dots, n; j = 1, \dots, p, k = 1, \dots, K\}$.

The transpose of a matrix is written $X_{[np,k]}^T$, its rank is $Rank(X_{[np,k]})$, its diagonalization matrix is $Diag(X_{[np,k]})$ and its vectorised form is $Vec(X_{[np,k]})$. The merger of two matrices is written as $XY_{[nr',k]}$, where there are n individuals and $r' = p + q$ variables. The identity matrix is $I_{[m]}$ and the unit vector is $1_{[m]}$. The multiplication of a matrix by a scalar $\lambda_{[11,k]}$ is represented by $\lambda_{[11,k]}X_{[np,k]}$, while the product of two matrices is symbolised by $X_{[np,k]} \times X_{[pr,k]}$.

Consider two sets of three-way matrices (data cubes) $\mathbb{X}_{[npK]}$ and $\mathbb{Y}_{[nqK]}$, structured by mixed data that may be quantitative (discrete, continuous) and/or qualitative (binary, ordinal, nominal). The data cubes have the characteristic that their matrices contain the same variables and that the individuals are the same between pairs of matrices for the k -th condition, that is:

$$\mathbb{X}_{[npK]} = [X_{[np,1]} | X_{[np,2]} | \dots | X_{[np,K]}], \quad (1)$$

$$\mathbb{Y}_{[nqK]} = [Y_{[nq,1]} | Y_{[nq,2]} | \dots | Y_{[nq,K]}]. \quad (2)$$

Subsequently, it is suggested to pre-treat $\mathbb{X}_{[npK]}$ and $\mathbb{Y}_{[nqK]}$ as follows:

- Quantitative variables must be applied a mathematical transformation as the suggest principally Kroonenberg [121] and Legendre and Legendre [122], such as: normalization by column, centre by column, logarithm;
- Binary variables must be coded with 0 and 1;
- Ordinal variables must be coded for each scale in ascending numerical order;
- Nominal variables must be transformed to a disjunctive matrix (each category is a new dichotomous variable).

With these modifications, the dimensions of the two data cubes are temporarily modified. The dimensions p and q belonging to the variables will increase by the creation of the disjunctive matrix for each nominal variable, resulting in:

$$\mathbb{X}'_{[np'K]} = [X_{[np',1]} | X_{[np',2]} | \dots | X_{[np',K]}], \quad (3)$$

$$\mathbb{Y}'_{[nq'K]} = [Y_{[nq',1]} | Y_{[nq',2]} | \dots | Y_{[nq',K]}]. \quad (4)$$

Among the methods commonly used to analyse the relationships between two sets of variables measured on the same individuals are CCA, Redundancy Analysis (RDA) [123], Canonical Correlation Analysis (CANCOR) [11] and Co-Inertia Analysis (Co-IA) [124]. Co-Inertia differs from the others in that it focuses on finding axes maximizing the covariance between the variables of the two tables.

To implement the methodology of analysing the co-structure between two sets of variables measured on the same individuals, a suggestion made by Abdi et al. [30] for the distance matrix based DISTATIS method is adopted. For mixed data, the distance metric of Gower [125] is used, between the variables of the paired matrices $X_{[pp,k]}$ and $Y_{[qq,k]}$, having previously merged these matrices in each K-condition to obtain the following set of K-distance matrices of dimension $(q' \times p' \times K)$:

$$\mathbb{X}Y_{[pp',k]} = [XY_{[pp',1]} | XY_{[pp',2]} | \dots | XY_{[pp',K]}], \text{ where } p' = p + q. \quad (5)$$

The Gower distance d_{ij} to each K-matrix is calculated independently from:

$$d_{ij}^2 = 1 - s_{ij}, \quad (6)$$

where

$$s_{ij} = \frac{\sum_{h=1}^{p_1} \left(1 - \frac{|x_{y_{ih}} - x_{y_{jh}}|}{G_h} \right) + a + \alpha}{p_1 + (p_2 - d) + p_3} \quad (7)$$

Here,

s_{ij} is the Gower similarity coefficient;
 p_1 is the number of continuous quantitative variables;
 p_2 is the number of binary variables;
 p_3 is the number of qualitative non-binary variables;
 a and d are the number of matches in binary variables (1,1) and (0,0), respectively;
 α is the number of matches in qualitative non-binary variables; and
 G_h is the rank of the h -th quantitative variable.

The analysis is based only on the distances between the variables p and q in $X_{[pp,k]}$ and $Y_{[qq,k]}$. Therefore, when at least one of these is nominal, the average Gower distance between the distances of the non-nominal variable and each of the categories of the nominal variable that were previously transformed into dichotomous variables can be obtained from:

$$\bar{d}_{v,w} = \frac{\sum_{i=1}^{nc} d_{v,w_i}}{nc}, \quad (8)$$

where

d is the Gower distance;
 v represents the non-nominal variable;
 w represents the nominal variable; and
 nc is the category number of the nominal variable.

If both variables are nominal, then the centroid method is applied, using the generalised average of all distances.

Finally, a set of K-distance matrices of dimension $(q \times p \times K)$ is obtained that expresses the common structure of the two initial data cubes:

$$\mathbb{Z}_{[qp,K]} = [Z_{[qp,1]} | Z_{[qp,2]} | \dots | Z_{[qp,K]}]. \quad (9)$$

Continuing to parallel the approach taken by Abdi et al. [30] to DISTATIS, a normalisation of the distance matrices, using a similar method to that applied in Factorial Multiple Analysis (FMA) [21,126], is developed to obtain new weighted co-inertia matrices $Z'_{[qp,k]}$ comparable in importance to the original distance matrices when comparing the K-matrices in the factorial plane.

To perform the normalisation, the Singular Value Decomposition (SVD) is applied to each K-distance matrix with rank L . This technique allows us to decompose rectangular matrices $Z_{[qp,k]}$ into factors $U_{[qL]}$ and $V_{[pL]}$, matrices containing as their columns orthonormal singular vectors associated with the singular values γ_{ii} contained in the prin-

cipal diagonal of the $\Gamma_{[LL]}$ matrix. $\mathbf{U}_{[qL]}$ and $\mathbf{V}_{[pL]}$ have the property $(\mathbf{U}_{[qL]}^T \times \mathbf{U}_{[qL,k]}) = (\mathbf{V}_{[pL,k]}^T \times \mathbf{V}_{[pL,k]}) = \mathbf{I}$. The singular vectors of $\mathbf{U}_{[qL]}$ and $\mathbf{V}_{[pL]}$ are eigenvectors of the matrices $(\mathbf{Z}_{[qp,k]} \times \mathbf{Z}_{[qp,k]}^T)$ and $(\mathbf{Z}_{[qp,k]}^T \times \mathbf{Z}_{[qp,k]})$, respectively, with eigenvalues $\lambda_{[11,k]} = \gamma_{[ii,k]}^2$. The matrices are normalised by multiplying their entries by the scalar $\lambda_{[11,k]}^{-1}$.

Formally,

$$\text{SVD}(\mathbf{Z}_{[qp,k]}) = \mathbf{U}\mathbf{\Gamma}\mathbf{V}^T, \text{ where } \mathbf{U} \in \mathbb{R}^{q \times L}, \mathbf{V} \in \mathbb{R}^{p \times L}, \mathbf{\Gamma} \in \mathbb{R}^{L \times L}, \text{ and } L = \text{Rank}(\mathbf{Z}_{[qp,k]}); \quad (10)$$

and the normalised K-matrices are:

$$\mathbf{Z}'_{[qp,k]} = \lambda_{[11,k]}^{-1} \mathbf{Z}_{[qp,k]}. \quad (11)$$

The $\mathbf{Z}'_{[qp,k]}$ form a set of weighted K-distance matrices:

$$\mathbf{Z}'_{[qp,K]} = [\mathbf{Z}'_{[qp,1]} | \mathbf{Z}'_{[qp,2]} | \dots | \mathbf{Z}'_{[qp,K]}]. \quad (12)$$

The next step is to calculate the cross-product matrix $\mathbf{C}_{[KK]}$, known as the variance-and-covariance matrix. It is first necessary to obtain the vectorisation of the matrices $\mathbf{Z}'_{[qp,k]}$:

$$\mathbf{Z}''_{[s'K]} = [\text{Vec}(\mathbf{Z}'_{[qp,1]}), \text{Vec}(\mathbf{Z}'_{[qp,2]}), \dots, \text{Vec}(\mathbf{Z}'_{[qp,K]})]^T; \text{ where } s' = q \times p. \quad (13)$$

Then,

$$\mathbf{C}_{[KK]} = \mathbf{Z}''_{[s'K]} \times \mathbf{Z}''_{[s'K]}^T. \quad (14)$$

To compute the inter-structure, Equation (14) is used as well as to calculate a scalar product matrix that contains the Rayleigh coefficients \mathbf{RV} [127], the vectorial correlation matrix:

$$\mathbf{RV}(k, k') = \frac{\mathbf{C}_{[kk']}}{(\mathbf{C}_{[kk]})^{1/2} (\mathbf{C}_{[k'k']})^{1/2}}; \text{ with } k = 1, \dots, K \text{ and } k' = 1, \dots, K \quad (15)$$

This matrix shows the proximity between K-matrices; its interpretation is similar to that of the Pearson linear correlation coefficient.

Abdi et al. [35] apply SVD to the $\mathbf{C}_{[KK]}$ matrix to analyse the similarities between K-matrices because it meets the symmetry requirements and is positive defined. Nevertheless, Abdi et al. [30] suggest calculating the eigenvectors and eigenvalues of the $\mathbf{RV}_{[KK]}$ matrix (the method adopted in this investigation), such that:

$$\mathbf{RV}_{[KK]} = \mathbf{U}\mathbf{\Xi}\mathbf{V}^T, \text{ with } \mathbf{U} \in \mathbb{R}^{K \times L}, \mathbf{V} \in \mathbb{R}^{K \times L}, \mathbf{\Xi} \in \mathbb{R}^{L \times L} \text{ and } L = \text{Rank}(\mathbf{RV}_{[KK]}) \quad (16)$$

After the decomposition of the $\mathbf{RV}_{[KK]}$ matrix and the computation of its eigenvectors and eigenvalues (the PCA method), the K-matrices can be projected onto the factorial plane as points, allowing the similarities between them to be studied. The elements of the first eigenvector have the same sign for being a positive semi-definite matrix.

The first eigenvector of $\mathbf{U}_{[KK]}$ supplies the weight vector $\vec{\alpha}_{[K]}$, subject to $\sum_{k=1}^K \vec{\alpha}_{[k]} = \mathbf{1}$:

$$\mathbf{U}_{[kk]} = \begin{bmatrix} u_{11} & \dots & u_{1K} \\ \vdots & \ddots & \vdots \\ u_{K1} & \dots & u_{KK} \end{bmatrix} \Rightarrow \begin{pmatrix} u_{11} \\ \vdots \\ u_{K1} \end{pmatrix} = \vec{u}_{[K,1]} \quad (17)$$

$$\vec{\alpha}_{[K]} = \left(\vec{u}_{[K,1]}^T \times \mathbf{1}_{[K]} \right)^{-1} \vec{u}_{[K,1]} \quad (18)$$

$$\vec{\alpha}_{[K]} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_K \end{pmatrix} \Rightarrow \sum_{k=1}^K \alpha_k = 1 \quad (19)$$

Abdi et al. [35] mention that K-matrix similarities are represented in the first eigenvector $\vec{u}_{[K,1]}$ which provides greater importance at matrices that better represent the co-inertia.

To show the inter-structure graphically, one calculates the coordinates of the K-matrices using the first two vectors of $U_{[KK]}$ (the first columns of which must be positive; see Abdi et al. [35]). From these, one obtains:

$$G = U_{[K2]} \times D_{[22]}^{1/2}, \text{ where } D_{[LL]} = \text{Diag}(\Xi) \quad (20)$$

With the $\vec{\alpha}_{[k]}$ values, the consensus matrix is calculated, which represents the stable part or average image of the relations between the paired matrices (co-inertia).

$$W_{[qp]} = \sum_{k=1}^K \vec{\alpha}_{[k]} Z'_{[qp,k]} \quad (21)$$

The consensus matrix $W_{[qp]}$ is analysed using Principal Coordinates Analysis (PcoA) [128] with a double centre. To represent its configuration graphically in a low dimension space, it is necessary to apply SVD:

$$W_{[qp]} = U' \Xi' V'^T, \text{ where } U' \in \mathbb{R}^{q \times L}, V' \in \mathbb{R}^{p \times L}, \Xi' \in \mathbb{R}^{L \times L} \text{ and } L = \text{Rank}(W_{[qp]}) \quad (22)$$

To generate the graph of the consensus matrix, the scores $F_{[qL]}$ for the row variables and the loadings $Q_{[pL]}$ for the column variables are calculated:

$$F_{[qL]} = U'_{[qL]} \times D_{[LL]}^{1/2} \text{ and } Q_{[pL]} = V'_{[pL]} \times D_{[LL]}^{1/2}; \text{ with } D'_{[LL]} = \text{Diag}(\Xi') \quad (23)$$

In the same way, the intra-structure is studied, which shows the Euclidean image of the consensus and consists of projecting the co-inertia of the K-matrices onto the consensus axes. A double centring is carried out on each K-matrix of $Z'_{[qp,k]}$, and the FK (scores for the row variables) and QK (loadings for the column variables) are obtained:

$$FK = Z'_{[qp,k]} \times V'_{[pL]} \quad \text{and} \quad QK = Z'^T_{[qp,k]} \times U'_{[qL]} \quad (24)$$

The procedure of the method is summarized in Figure 1 and the algorithm is shown in Appendix A.

The algorithm that summarizes the MixSTATICO method steps is in Appendix A.

2.3. Case Study

In Ecuador, there is limited information regarding previous studies carried out on this subject. Therefore, it is necessary to do a space-time analysis to identify the effect that typical environmental variations have on this region. Increases in temperature and nutrient concentrations can occur, owing to the effect of the geographical location and the climatic season of this area on the abundance of phytoplanktonic, which is predominately the diatoms group.

The coast of Continental Ecuador is located in the eastern equatorial Pacific (Figure 2), with the north influenced by the warm current of Panama (heading south and producing an increase in SST and decrease in nutrients; begins in December and intensifies in February and April) and the south by the cold stream of Humboldt (heading north, its waters are coldest, high in salinity and rich in nutrients; it mainly presents in July and November, weakening in December), and there are two climatic seasons (dry and rainy) [129]. The southern part is considered to be the most significant tropical estuary on the west coast of South America [120]. However, much of its coastline is contained in the Niño Region

1 + 2 (0°–10° S, 90°–80° W), an area sensitive to the variations produced by the surface sea temperature from the Central Pacific due to the presence of ENSO events (El Niño—La Niña), which are events that alter the physical and chemical parameters of the sea. On the west coast of South America, the most evident signs of ENSO are manifested in ocean–atmospheric systems and by the impact it has on natural ecosystems [130].

n: individuals number
p, *q*: variables number – original matrix
p', *q'*: variables number – disjunctive matrix

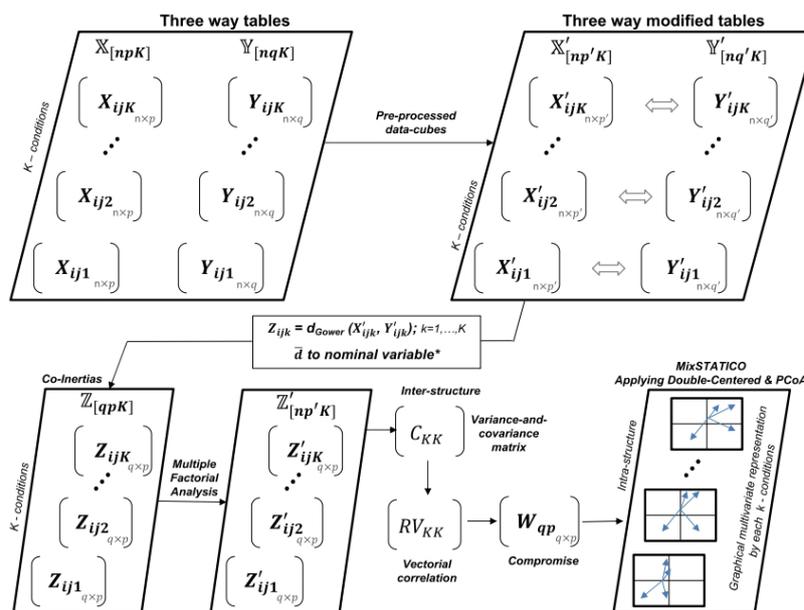


Figure 1. Scheme of the MixSTATICO method. For this application, $X_{[npK]}$ is the environment and $Y_{[mpK]}$ is the species.

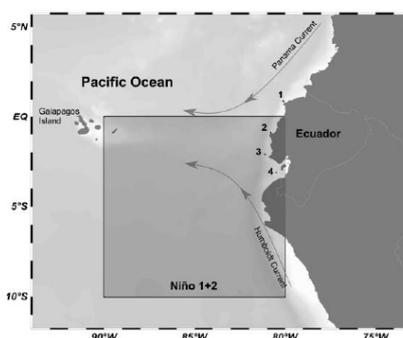


Figure 2. Localization of the sample stations—1. Esmeraldas (0.92° N, 80.11° W), 2. Manta (0.88° S, 80.83° W), 3. La Libertad (2.08° S, 81.12° W), 4. Pto. Bolívar (3.11° S, 80.49° W). The stations 2, 3 and 4 are contained in Region Niño 1+2. Self-authorship using Ocean Data View [131].

Due to the importance of this area, constant monitoring of the chemical and biological conditions of the sea is carried out in space and in time.

3. Results

All data were analysed using the software R Studio version 1.2.1335 and the functions daisy (Gower distance) and svd (Singular Value Decomposition).

Below is a summary of the spatial–temporal data describing the environmental and biological characteristics of the coastal region of Ecuador (see Appendices B and C for all the environmental and species data).

Tables 1 and 2 show the average value and standard deviation (SD) of the environmental and biological variables, respectively, according to the location of the sampling stations and the climatic period.

The stations in the north (Esmeraldas and Manta) presented the highest average values with regard to temperature and the lowest average values for nutrients. In addition, the environmental variables (except salinity) showed the highest average values during the rainy season and the Esmeraldas estuary had the warmest waters of the entire coast (Table 1).

The phytoplankton species had different average abundance values that varied according to the geographical location of each station (which had different environmental characteristics) and the influence of the season (Table 2).

Table 2. Species variables (average \pm SD).

Group	Species Phytoplankton Abundance Cells L ⁻¹	Location (Both Seasons)				Season (Entire Coastal Profile)		Total
		North	Centre North	Centre South	South	Rainy	Dry	
CD	<i>Dactylosolen fragilissimus</i> (e1)	2.78 \pm 2.14	3.88 \pm 1.34	2.89 \pm 2.22	1.94 \pm 2.15	2.41 \pm 2.21	3.26 \pm 1.94	2.87 \pm 2.11
CD	<i>Guinardia striata</i> (e2)	4.29 \pm 1.42	4.23 \pm 1.43	3.91 \pm 1.9	2.84 \pm 2.3	3.52 \pm 1.82	4.07 \pm 1.91	3.82 \pm 1.89
CD	<i>Rhizosolenia imbricata</i> (e3)	3.76 \pm 1.34	3.99 \pm 1.41	3.17 \pm 2.47	2.16 \pm 2	2.37 \pm 2	4.02 \pm 1.65	3.27 \pm 1.99
CD	<i>Dactylosolen antarcticus</i> (e4)	2.76 \pm 2.11	3.48 \pm 1.74	2.07 \pm 2.38	0.93 \pm 1.97	1.26 \pm 1.94	3.18 \pm 2.15	2.31 \pm 2.27
CD	<i>Proboscia alata</i> (e5)	3.46 \pm 1.22	2.55 \pm 1.97	2.11 \pm 1.99	0.84 \pm 1.78	1.61 \pm 1.82	2.76 \pm 2	2.24 \pm 2.01
CD	<i>Skeletonema costatum</i> (e6)	1.97 \pm 2.17	1.33 \pm 2.21	1 \pm 1.67	0.96 \pm 2.08	1.12 \pm 2.02	1.48 \pm 2.12	1.32 \pm 2.08
CD	<i>Lauderia borealis</i> (e7)	0.62 \pm 1.33	0.4 \pm 1.28	1.17 \pm 1.92	1.17 \pm 1.91	0.8 \pm 1.62	0.88 \pm 1.71	0.84 \pm 1.67
PD	<i>Nitzschia longissima</i> (e8)	3.84 \pm 1.29	4.48 \pm 0.64	3.75 \pm 1.83	4.46 \pm 0.5	3.96 \pm 1.74	4.28 \pm 0.49	4.13 \pm 1.24
PD	<i>Nitzschia</i> sp. (e9)	2.36 \pm 1.8	3.41 \pm 1.2	2.33 \pm 1.88	2.97 \pm 1.46	1.97 \pm 2.04	3.43 \pm 0.83	2.77 \pm 1.67
PD	<i>Pseudo-nitzschia pungens</i> (e10)	3.25 \pm 1.63	2.93 \pm 1.82	2.05 \pm 1.9	2.82 \pm 2.23	2.85 \pm 1.97	2.69 \pm 1.95	2.76 \pm 1.96
CD	<i>Leptocylindrus danicus</i> (e11)	4.38 \pm 0.47	4.48 \pm 0.77	3.25 \pm 2.03	2.74 \pm 2.11	3.41 \pm 1.78	3.97 \pm 1.59	3.71 \pm 1.7
CD	<i>Thalassiosira</i> sp. (e12)	1.89 \pm 1.75	2.73 \pm 1.69	2.98 \pm 1.84	2.77 \pm 1.38	2.21 \pm 1.85	2.91 \pm 1.54	2.59 \pm 1.73
CD	<i>Chaetoceros affinis</i> (e13)	3.51 \pm 1.72	4.58 \pm 0.41	3 \pm 1.9	2.8 \pm 2.18	3.29 \pm 2.01	3.62 \pm 1.65	3.47 \pm 1.83
CD	<i>Chaetoceros curvisetus</i> (e14)	1.6 \pm 2.14	1.44 \pm 1.91	0 \pm 0	0.87 \pm 1.85	1.08 \pm 1.89	0.89 \pm 1.75	0.98 \pm 1.82
CD	<i>Chaetoceros didymus</i> (e15)	0 \pm 0	0.71 \pm 1.51	0.34 \pm 1.08	0.37 \pm 1.16	0.78 \pm 1.56	0 \pm 0	0.35 \pm 1.12
CD	<i>Hemiaulus sinensis</i> (e16)	3.05 \pm 1.99	3.04 \pm 1.9	2.87 \pm 1.86	1.5 \pm 2.08	2.32 \pm 1.94	2.86 \pm 2.13	2.62 \pm 2.07
PD	<i>Thalassionema nitzschioides</i> (e17)	2.31 \pm 1.77	2.75 \pm 1.72	1.83 \pm 2.02	2.06 \pm 1.92	1.84 \pm 1.86	2.56 \pm 1.85	2.24 \pm 1.89
D	<i>Gymnodinium</i> sp. (e18)	3.23 \pm 1.53	3.5 \pm 1.21	3.64 \pm 1.23	3.22 \pm 1.57	2.89 \pm 1.71	3.82 \pm 0.89	3.4 \pm 1.41
D	<i>Gyrodinium</i> sp. (e19)	1 \pm 1.66	0.91 \pm 1.53	0.61 \pm 1.29	0.55 \pm 1.18	1.53 \pm 1.73	0.13 \pm 0.64	0.77 \pm 1.44
C	<i>Mesodinium rubrum</i> (e20)	3.29 \pm 1.07	2.92 \pm 1.39	3.18 \pm 1.13	3.72 \pm 0.55	3.34 \pm 1.21	3.23 \pm 1.03	3.28 \pm 1.12
CD	<i>Ditylum brightwellii</i> (e21)	1.46 \pm 1.61	0.79 \pm 1.29	0.63 \pm 1.34	0.89 \pm 1.45	0.79 \pm 1.38	1.07 \pm 1.52	0.94 \pm 1.46
PD	<i>Navicula</i> sp. (e22)	1.69 \pm 1.87	2.77 \pm 1.36	1.89 \pm 1.76	1.9 \pm 1.79	2 \pm 1.85	2.12 \pm 1.68	2.06 \pm 1.76
PD	<i>Stauroneis membranacea</i> (e23)	2.49 \pm 1.91	2.55 \pm 1.94	1.66 \pm 1.82	1.42 \pm 1.89	1.72 \pm 1.92	2.28 \pm 1.95	2.03 \pm 1.96

CD: Centric Diatom, PD: Pennate Diatom, D: Dinoflagellate, C: Ciliate.

The inter-structure graph, the result of the first step of the MixSTATICO analysis, showed that the environment–species association had two types of behaviour, which varied according to the season (Figure 3A). The members of the first type presented a similar behaviour during the dry season (red in Figure 3A), while those of the second type behaved similarly during the months of the rainy season (blue), except for October and December. The two first components explained 93.07% of the inertia (Figure 3B).

Moreover, there was a solid typical structure between the environment and species, as it was observed that all K-matrices (months) reached values close to the unitary cycle. They also had high values for weights. Therefore, all matrices substantially contributed

toward building the consensus matrix (Figure 3A,C). November contributed the most to the consensus (weight \times 0.094), followed by August and September (dry season), but the consensus of the environment–species association was more notably represented during November (highest value of \cos^2) (Figure 3C).

April, February and May (rainy season) made the lowest contributions to the consensus, while May was considered a month of seasonal change and the end of the rainy season.

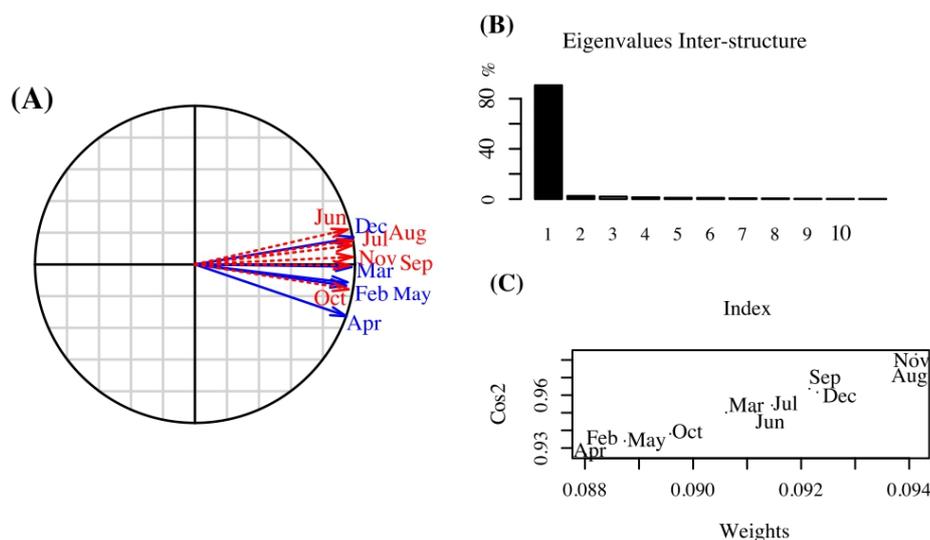


Figure 3. (A) Inter-structure graph (dry season: red colour; rainy season: blue colour). (B) Eigenvalues and (C) weights (α) and \cos^2 for each K matrix.

In relation to the consensus analysis, which showed the stable part of the species–environment association, the two first components explained 99.92% of the inertia (Figure 4C). The consensus for the environment variables showed a strong positive correlation among temperature, dissolved oxygen and salinity, and a strong negative correlation with phosphate, silicate and nitrate (Figure 4A).

In 2013, the average temperature value was 23.6 °C, with the highest values recorded at the stations located to the north, while the highest average nutrient values were recorded at the stations south of the coastal profile (Table 1).

The ordination of species in the consensus (Figure 4B) established a high level of association between *S. membranacea* (e23—continuous variable) and *S. membranacea* (e24 ordinal variable), where both variables contain the same information, showing the suitability of the method when different variable types are presented. The abundance of centric diatoms *Dactylosolen antarcticus* (e4), *Proboscia alata* (e5), *Skeletonema costatum* (e6), *Lauderia borealis* (e7), *Chaetoceros didymus* (e15) and *Ditylum brightwellii* (e21) and dinoflagellate *Gyrodinium* sp. (e19) were associated (same direction) with the location and season variables. This indicated that the vulnerability of the species varied according to the geographic location of the sampling stations and the season, with different levels of average abundance (Table 2). In addition, these species had the lowest average abundance throughout the whole coastline (Total column of Table 2), being absent during some of the months and stations (Table A2). In addition, the species with stable abundance levels (ordered in the opposite direction to the location and season variables), during both periods and in the entire coastal profile,

were mainly *P. pungens* (e10), *Gymnodinium* sp. (e18) and *Dactyliosolen fragilissimus* (e1), *S. membranacea* (e23) and *S. membranacea* (e24).

The vulnerable species indicated in Figure 4B (located on the factorial environmental plane Figure 4A) show preferential environmental conditions with below-average values in relation to temperature, dissolved oxygen and salinity (opposite direction to these variables).

The following species where those with the highest abundance (longer vector) over the entire coastline: the pennate diatoms *Pseudo-nitzschia pungens* (e10) and *N. longissima* (e8) and the dinoflagellate *Gymnodinium* sp. (e18). These species preferred above-average values in relation to temperature, dissolved oxygen and salinity. *Dactyliosolen antarcticus* (e4), on the other hand, was detected during the dry season at the stations located in the north (Esmeraldas and Manta) (Figure 3A,B).

The trajectories of both variables and species for the rainy and the dry seasons are presented separately for the purpose of comparison.

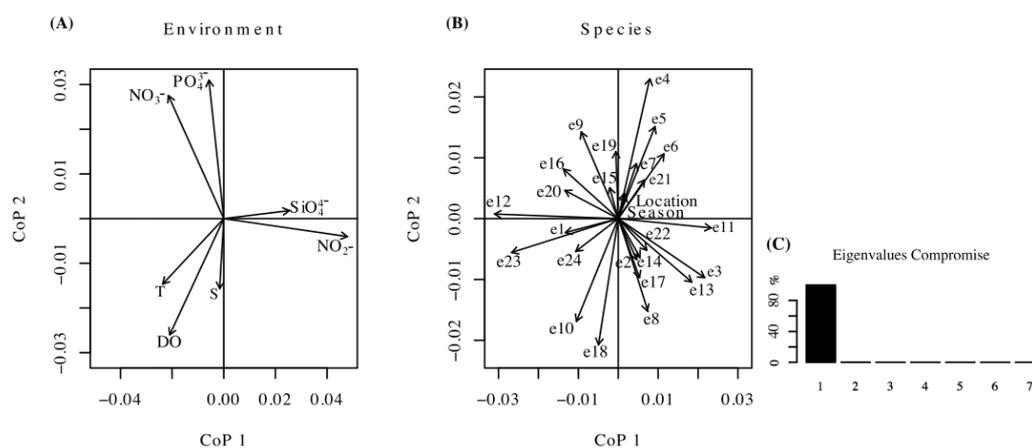


Figure 4. Consensus: factorial plane to (A) environment and (B) species; (C) eigenvalues. Species names: *Dactyliosolen fragilissimus* (e1), *Guinardia striata* (e2), *Rhizosolenia imbricata* (e3), *Dactyliosolen antarcticus* (e4), *Proboscia alata* (e5), *Skeletonema costatum* (e6), *Lauderia borealis* (e7), *Nitzschia longissima* (e8), *Nitzschia* sp. (e9), *Pseudo-nitzschia pungens* (e10), *Leptocylindrus danicus* (e11), *Thalassiosira* sp. (e12), *Chaetoceros affinis* (e13), *Chaetoceros curvisetus* (e14), *Chaetoceros didymus* (e15), *Hemiaulus sinensis* (e16), *Thalassionema nitzschioides* (e17), *Gymnodinium* sp. (e18), *Gyrodinium* sp. (e19), *Mesodinium rubrum* (e20), *Ditylum brightwellii* (e21), *Navicula* sp. (e22), *Stauroneis membranacea* (e23—continuous variable), *Stauroneis membranacea* (e24—ordinal variable).

The same interpretation is applied to the intra-structure.

During the rainy months (Figure 5A,C,E,G,I), it was observed that the environmental and species variables showed a non-stable management pattern. For February and March, temperature, dissolved oxygen and salinity showed a direct and inverse association with phosphate and nitrate, while in April and December, that association became weaker, mainly being between temperature and dissolved oxygen. The association pattern in May (month of change of season) differs from the rest of the months since it was observed that temperature does not directly associate with any other variable. At this time, the highest average values for temperature and nutrients were recorded (Table 1).

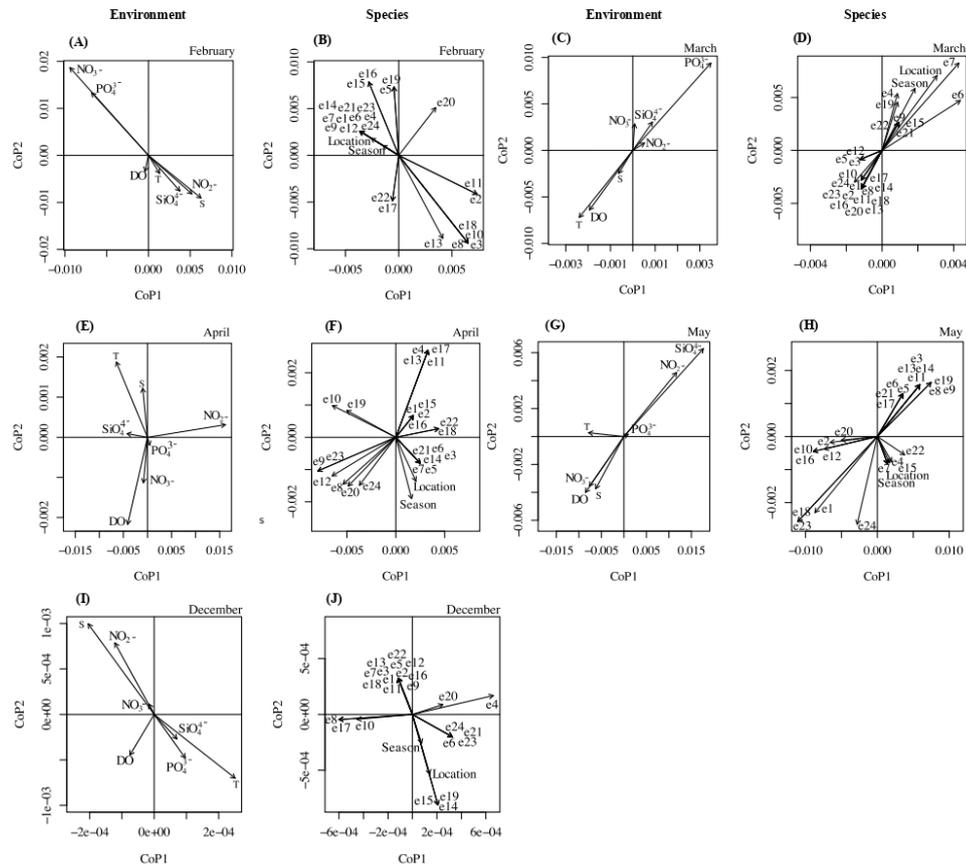


Figure 5. Intra-structure—rainy season: projections onto factorial plane of the consensus to environment and species corresponding each matrix (time). In the months of the rainy season, the environmental and species variables showed a non-stable management pattern. Distinct environmental correlations and weak patterns of association between species are observed. Species names: *Dactylosolen fragilissimus* (e1), *Guinardia striata* (e2), *Rhizosolenia imbricata* (e3), *Dactylosolen antarcticus* (e4), *Proboscia alata* (e5), *Skeletonema costatum* (e6), *Lauderia borealis* (e7), *Nitzschia longissima* (e8), *Nitzschia* sp. (e9), *Pseudo-nitzschia pungens* (e10), *Leptocylindrus danicus* (e11), *Thalassiosira* sp. (e12), *Chaetoceros affinis* (e13), *Chaetoceros curvisetus* (e14), *Chaetoceros didymus* (e15), *Hemiaulus sinensis* (e16), *Thalassionema nitzschioides* (e17), *Gymnodinium* sp. (e18), *Gyrodinium* sp. (e19), *Mesodinium rubrum* (e20), *Ditylum brightwellii* (e21), *Navicula* sp. (e22), *Stauroneis membranacea* (e23—continuous variable), *Stauroneis membranacea* (e24—ordinal variable).

The species with the highest abundance during the rainy season (Figure 5B,D,F,H,J) were mainly the species *Leptocylindrus danicus* (e11), *Chaetoceros affinis* (e13) and *Gymnodinium* sp. (e18). In February to April (they showed a longer vector), their high abundance was associated with above-average temperature and salinity values and low abundance was associated with nitrate values, among other nutrients.

Each sampling month showed the species that had a decrease in their abundance or disappeared due to the influence of the sampling station location and the variability of the corresponding climatic feature in that month. At this time of year (except in December), the main vulnerable species were the centric diatoms *D. antarcticus* (e4), *P. alata* (e5) and *S. costatum* (e6), which were found mainly at the southern station, and *D. brightwellii* (e21) at

the stations in the centre of the coastal profile that were influenced directly by the ocean. These species seemed to thrive more in waters with below-average temperature and salinity values (Table 2).

During the dry season, like during the months of the rainy season, a non-stable environmental ratio pattern (Figure 6A,C,E,G,I,K) was displayed, changing according to the month the samples were taken. In addition, lower values were obtained for nutrients and temperature.

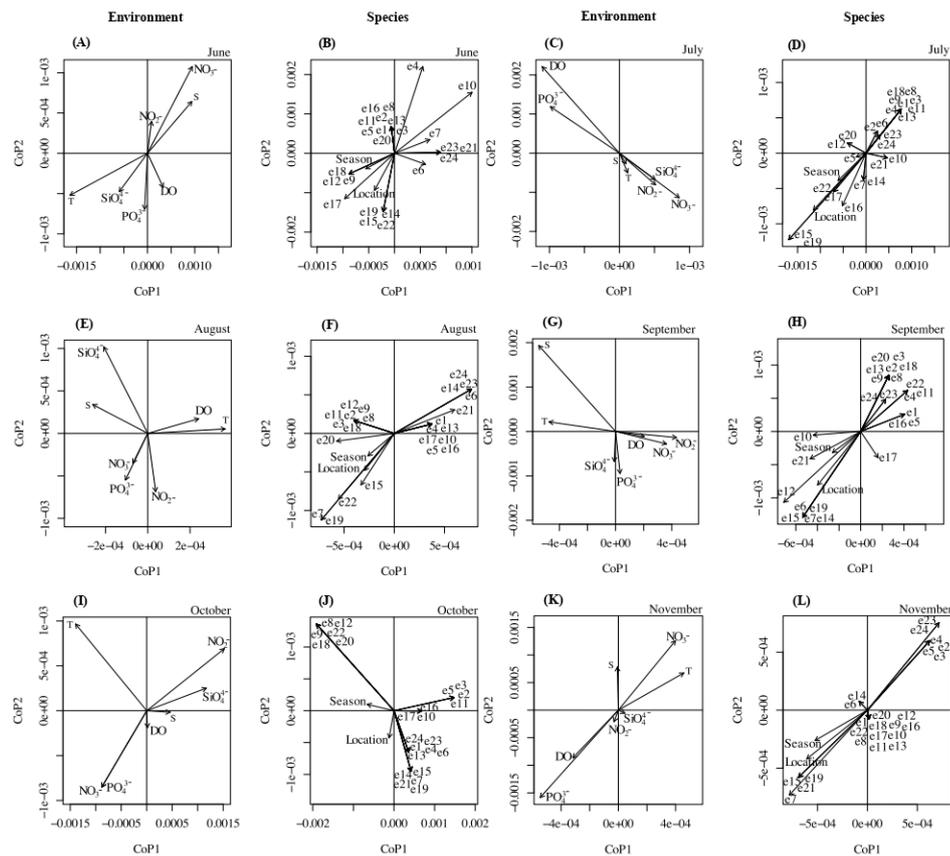


Figure 6. Intra-structure—dry season: projections onto factorial plane of the consensus to environment and species corresponding each matrix (time). In the months of the dry season, there are more apparent patterns of association between species with a non-stable environmental ratio pattern. Species names: *Dactylosolen fragilissimus* (e1), *Guinardia striata* (e2), *Rhizosolenia imbricata* (e3), *Dactylosolen antarcticus* (e4), *Proboscia alata* (e5), *Skeletonema costatum* (e6), *Lauderia borealis* (e7), *Nitzschia longissima* (e8), *Nitzschia* sp. (e9), *Pseudo-nitzschia pungens* (e10), *Leptocylindrus damicus* (e11), *Thalassiosira* sp. (e12), *Chaetoceros affinis* (e13), *Chaetoceros curvisetus* (e14), *Chaetoceros didymus* (e15), *Hemiaulus sinensis* (e16), *Thalassionema nitzschioides* (e17), *Gymnodinium* sp. (e18), *Gyrodinium* sp. (e19), *Mesodinium rubrum* (e20), *Ditylum brightwellii* (e21), *Navicula* sp. (e22), *Stauroneis membranacea* (e23—continuous variable), *Stauroneis membranacea* (e24—ordinal variable).

The species that stood out as being more abundant during these months were centric diatoms *D. antarcticus* (e4), *S. costatum* (e6), *P. pungens* (e10) and *D. brightwellii* (e21), among others, and their preferred environmental conditions varied according to the month. The

species that were vulnerable to environmental variations and geographical location were mainly *Lauderia borealis* (e7), *Chaetoceros curvisetus* (e14), *C. didymus* (e15), *Gyrodinium* sp. (e19) and *Navicula* sp. (e22), as they had the lowest average abundance values (see Table 2). However, the species that had a steady abundance were *Guinardia striata* (e2), *Rhizosolenia imbricata* (e3), *D. antarcticus* (e4), *P. alata* (e5), *S. membranacea* (e24) and *S. membranacea* (e23) (Figure 6B,D,F,H,J,L).

Moreover, there are more apparent patterns of association between species in the dry season (except in June, the start of the dry season) as compared to the rainy season.

4. Discussion

Water sustainability is an overriding issue for ecosystem support [1].

Human actions and natural events are factors that cause alterations in the environment, affecting ecosystem biotic and abiotic components [132–134]. Therefore, planktonic components are of great importance in assessing the water quality of aquatic systems, as they are subject to environmental variations [135] in the area and the geographical locations where they grow [95]. This was established by the analysis of the consensus and the intra-structure, where phytoplankton species vulnerable to climatic variability and environmental conditions of the estuarine and oceanic stations were identified. These stations are seen to have a positive effect on the ecosystem, which we believe gives support to the importance of studying the hydrological variations and the effects they have on biota and biotic processes. The estuarine and oceanic stations are known as dynamic environments that exhibit spatial and temporal variations in their biotic and abiotic parameters [136]. This dynamic behaviour was classified into two groups based on the months of the year (inter-structure), where the rainy season was established as February–May and the dry season as June–December [129], except for October and December. December was the start of the rainy season of 2014, the year in which a weak El Niño event occurred [137] with aforementioned anomalies >1.5 (Oceanic Niño Index—ONI) that decreased as the month passed.

In addition to the intra-structure, differences in species population patterns were observed between both seasons, with the dry season showing a more pronounced pattern of association among the species. The reason for this is that during the rainy season, the species–environment co-structure was weaker and the environmental variables presented unstable behaviour due to the presence of high outlier values for phosphate, nitrate, nitrite and salinity during February, March and April (in the different sample stations). This could be due to an increased amount of runoff originating from sedimentary basins (agriculture fields wash off). However, regarding the temperature, it only increases in August, a month that tends to experience upwellings along the Ecuador coast at the northern station (Esmeraldas) [106]. The highest average temperature and the lowest nutrient values were recorded in the stations located to the north (Esmeraldas and Manta). In contrast, an opposite effect occurred in the southern stations (La Libertad and Pto. Bolívar) (see Table 2). This behaviour coincides with the presence of warm waters coming from the north caused by the influence of the Panama current. However, in the south, oceanic waters are influenced by the cold Humboldt current, as well as upwellings that favour planktonic productivity [111].

During most of the months of the dry season (intra-structure), the species *D. antarcticus* (e4) showed a high abundance, which was associated with above-average nitrate values and low temperatures [138], which occurred mainly in June and July (Table A2). This is due to the presence of a cold Humboldt current, rich in nutrients, which is consistent with optimal environmental conditions to support the high productivity of diatoms. In the intra-structure chart for February, March, May and December, the dinoflagellate *Gyrodinium* sp. (e19) was found to be the species that associated its above-average abundance with geographical location (north, Esmeraldas and Manta—Table 2) and climate (rainy—Table 2). In 2013, other authors [139] stated that nutrient concentrations are the drivers of dinoflagellate productivity, drawing attention to species such as *Gymnodinium cf. catenatum*, *Oxytoxum*

turbo and *Prorocentrum micans*, which we found were associated with warm waters (typical of the Panama current) with low salinity and nutrients during the rainy season in the northern station. This may have occurred because dinoflagellates trophic behaviour is not dependent on nutrient levels, and as such is considered an indicator of species of warm ocean waters with low nutrient levels [140,141].

In the intra-structure, it was possible to identify the high abundance species, such as *P. alata* (e5) in November, which was associated with above-average nitrate and temperature values in the north station, Esmeraldas (Table 2). In August, the species *D. fragilissimus* (e1) was abundant with above-average values in temperature and oxygen association values in the north centre station, Manta (Table 2), but whose presence was found to be associated with nutrient availability [137]. In February, March, April and May, the abundance of *P. pungens* (e10) was found to be associated with above-average temperature and oxygen values in the south centre station, La Libertad (Table 2), which is also consistent with the warmer temperatures associated with conditions during the rainy season. Finally, throughout the rainy season (Feb–Dec), the species *Hemiaulus sinensis* (e16) associated its high abundance to above-average nitrate, nitrite and salinity values in the south station, Pto. Bolívar. These results reaffirm the reports of other studies [106] that describe how *P. alata* (e5) is the most abundant species in Esmeraldas and *D. fragilissimus* (e1) in Manta during the dry season, and *P. pungens* (e10) in La Libertad and *H. sinensis* (e16) in Pto. Bolívar during the rainy season. Nevertheless, the findings of this study highlight the association of phytoplankton growth with location and seasons, where *P. alata* (e5) was found to be associated with these variables. However, in the case of *D. fragilissimus* (e1), *P. pungens* (e10) and *H. sinensis* (e16), these species were found to have had a stable level of abundance throughout all the seasons and along the entire coastal profile.

This study has shown the utility of this new multivariate method in analysing associations between biota abundance and environmental variables such as location and season within ENSO neutral conditions. In addition, this work highlights the seasonal patterns and possible association with freshwater run-off, which proves that the method could be useful for the assessment of other environmental contexts.

5. Conclusions

In this work, using the MixSTATICO multi-way mixed multivariate method, the need for additional analysis has been reduced. In this sense, descriptive statistics techniques such as PCA and CCA were employed and applied independently to the data collected at each sampling station. In addition, this work introduces geographical location and climatic periods as qualitative variables and analyses them in conjunction with the other quantitative variables, namely temperature, dissolved oxygen, salinity, nutrients and phytoplankton abundance.

Our findings support the effectiveness of the proposed method, which reduces the amount of analysis required to obtain the same results. This is conducted through the analysis of data cubes that contain sets of quantitative and qualitative variables. In this sense, the STATICO method calculates the stable part of the common structure between two data cubes.

This innovative method has identified the phytoplankton species showing poor, stable and high levels of abundance according to climatic variability and the environmental conditions typical of these geographical locations and seasons. In addition, this work has identified the more vulnerable species in the coast profile as *D. antarcticus*, *P. alata*, *S. costatum*, *L. borealis*, *C. didymus* and *D. brightwellii* (centric diatoms) and *Gyrodinium* sp. (dinoflagellate). During the dry season, it was possible to observe more apparent patterns of association between species.

These results provide proof of the validity of this method for analysing mixed data, and even offer the possibility of introducing qualitative characteristics to the study. In this sense, the results obtained are further enriched for the knowledge of the ecosystem. This is performed from a quantitative and qualitative analysis perspective. Furthermore, this

method, supported by statistical data, efficiently introduces the insight of environmental assessment in terms of environmental parameters' relationships.

Author Contributions: Conceptualization, M.J.F.-G. and O.R.-B.; methodology, M.G.-N. and O.R.-B.; software, M.G.-N.; validation, S.M.; formal analysis, M.J.F.-G.; investigation, S.M. and M.J.F.-G.; resources, S.M., M.J.F.-G., P.G.-V. and J.-L.M.; data curation, M.G.-N.; writing—original draft preparation, M.G.-N.; writing—review and editing, P.G.-V., J.-L.M., S.M. and M.J.F.-G.; visualization, M.G.-N.; supervision, P.G.-V., M.J.F.-G., S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in Appendices B and C.

Acknowledgments: We acknowledge the invaluable collaboration of INOCAR (Oceanographic Institute of the Navy, Ecuador), CPG-EM Edwin Pinto Uscocovich, and Gladys Torres, who have provided us with oceanographic data from a project that studied El Niño under the regional project framework El Niño-ERFEN-ECUADOR. We have no business and/or financial interest in the project; we have disclosed our interests fully and have in place an approved plan for managing any potential conflicts arising from this arrangement.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. MixSTATICO Algorithm

1. Input two pre-processed data-cubes $\mathbb{X}_{[npK]}$ and $\mathbb{Y}_{[nqK]}$.
2. Merge all $X_{[np,k]}$ and $Y_{[nq,k]}$ matrices.
For $k = 1$ to K

$$\{XY_{[nr',k]} = merge(X_{[np,k]}, Y_{[nq,k]})\}; \text{ where } r' = p + q$$

3. Compute the Gower distance, considering Equations (6)–(8).
For $k = 1$ to K

$$\{Z_{[qpK]} = gower(XY_{[nr',k]})\}$$

4. Apply FMA to normalize
For $k = 1$ to K

SVD ($Z_{[qp,k]}$) to obtain $\lambda_{[11,k]}$ and multiply each matrix element by $\lambda_{[11,k]}^{-1}$

$$\{Z'_{[qpK]} = \lambda_{[11,k]}^{-1} Z_{[qp,k]}\}$$

5. Vectorize $Z'_{[qpK]}$

$$Z''_{[s'K]} = [Vec(Z'_{[qp,1]}), Vec(Z'_{[qp,2]}), \dots, Vec(Z'_{[qp,K]})]^T; \text{ where } s' = q \times p$$

6. Calculate $C_{[KK]}$

$$C_{[KK]} = Z''_{[s'K]} \times Z''_{[s'K]}^T$$

7. Calculate $RV_{[KK]}$.
For $k = 1$ to K

$$\{RV_{[KK]} = \frac{C_{[kk']}}{\sqrt{C_{[kk]} \times \sqrt{C_{[k'k']}}}\}; \text{ with } k = 1, \dots, K \text{ and } k' = 1, \dots, K$$

8. SVD ($RV_{[KK]}$) to obtain $\vec{\alpha}_{[K]}$
9. Calculate consensus matrix

$$W_{[qp]} = \sum_{k=1}^K \vec{\alpha}_{[k]} Z'_{[qp,k]}$$

10. SVD ($W_{[qp]}$).
11. Generate consensus graph using PCoA.
12. Generate intra-structure graph using PCoA.

Appendix B. Table Environmental Data

Table A1. Text Values for physical and chemical variables.

Space	Time	Temperature (T—C)	Salinity (S—psu)	Dissolved Oxygen (DO—mg.L ⁻¹)	Nitrate (NO ₃ ⁻ — μg—at.L ⁻¹)	Nitrite (NO ₂ ⁻ — μg—at.L ⁻¹)	Phosphate (PO ₄ ⁻ — μg—at.L ⁻¹)	Silicate (SiO ₄ ⁻ — μg—at.L ⁻¹)
Esmeraldas	Feb	26.74	32.65	5.29	0.77	0.08	0.71	10.42
Manta	Feb	24.02	33.24	4.35	6.91	0.16	1.62	5.12
La Libertad	Feb	25.25	33.11	5.38	0.73	0.28	0.58	11.88
Pto. Bolivar	Feb	22.23	33.47	4.35	8.76	0.22	1.31	7.52
Esmeraldas	Mar	21.92	34.11	4.28	5.95	0.39	2.00	14.36
Manta	Mar	23.04	33.69	4.37	2.73	0.43	1.75	13.73
La Libertad	Mar	23.58	34.29	4.62	3.64	0.30	1.33	5.24
Pto. Bolivar	Mar	21.17	31.56	3.35	9.36	0.29	0.97	12.87
Esmeraldas	Apr	27.45	33.24	4.45	0.50	0.06	0.71	11.06
Manta	Apr	24.47	34.15	4.36	1.55	0.28	0.38	8.07
La Libertad	Apr	22.79	34.47	4.46	4.96	0.09	1.26	10.08
Pto. Bolivar	Apr	23.12	33.56	4.16	3.86	0.73	1.02	8.14
Esmeraldas	May	26.11	33.21	4.60	1.80	0.19	0.70	3.22
Manta	May	26.07	33.40	4.35	3.85	0.15	0.75	2.10
La Libertad	May	24.65	33.64	4.12	1.70	0.56	1.17	19.10
Pto. Bolivar	May	24.95	33.81	4.53	4.92	0.02	0.51	10.66
Esmeraldas	Jun	26.33	33.17	4.60	2.02	0.20	0.38	10.11
Manta	Jun	24.31	33.95	4.32	0.65	0.34	0.31	1.63
La Libertad	Jun	20.45	34.41	3.20	7.75	0.41	0.95	9.55
Pto. Bolivar	Jun	22.18	34.22	5.12	0.36	0.11	0.33	2.65
Esmeraldas	Jul	25.50	32.91	4.73	1.13	0.03	1.02	3.66
Manta	Jul	23.90	33.56	4.62	0.99	0.10	0.84	3.16
La Libertad	Jul	22.68	33.78	4.82	0.39	0.04	0.83	9.34
Pto. Bolivar	Jul	22.80	33.99	5.01	1.47	0.24	1.25	8.87
Esmeraldas	Aug	25.49	32.89	4.62	0.24	0.01	0.25	1.92
Manta	Aug	22.73	33.59	4.47	2.27	0.05	0.71	6.73
La Libertad	Aug	19.09	34.38	3.89	4.04	0.23	1.00	12.21
Pto. Bolivar	Aug	21.60	33.93	3.90	3.19	0.04	0.56	12.46
Esmeraldas	Sep	26.33	32.68	4.32	0.72	0.04	0.26	1.16
Manta	Sep	21.35	34.20	4.53	1.09	0.01	0.38	8.84
La Libertad	Sep	21.85	33.96	4.00	2.43	0.09	0.58	5.01
Pto. Bolivar	Sep	22.54	33.84	4.41	3.36	0.25	0.50	4.66
Esmeraldas	Oct	26.09	32.20	4.23	0.58	0.18	0.10	3.73
Manta	Oct	24.31	33.43	3.81	2.59	0.23	0.18	5.11
La Libertad	Oct	23.18	33.72	4.51	0.29	0.13	0.33	6.87
Pto. Bolivar	Oct	22.60	33.81	4.17	1.54	0.38	0.46	11.88
Esmeraldas	Nov	26.06	32.41	4.39	0.12	0.08	0.11	2.23
Manta	Nov	24.93	32.99	4.56	0.30	0.05	0.13	2.82
La Libertad	Nov	20.30	34.38	3.27	5.03	0.26	0.61	9.11
Pto. Bolivar	Nov	21.21	34.24	3.58	6.08	0.31	0.63	10.89
Esmeraldas	Dec	25.45	32.65	3.96	0.26	0.03	0.21	6.24
Manta	Dec	24.38	33.09	4.48	1.03	0.03	0.37	1.27
La Libertad	Dec	21.64	34.23	4.25	3.52	0.29	0.69	7.41
Pto. Bolivar	Dec	21.56	34.23	3.13	3.81	0.37	0.09	8.48

Appendix C. Table Species Data

Table A2. Text The abundance of the phytoplankton species.

Space	Time	Phytoplankton Species 1—Abundance Cells L ⁻¹																								Localization	Sea-son	
		e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13	e14	e15	e16	e17	e18	e19	e20	e21	e22	e23	e24			
Esmeraldas Manta La Libertad Pro. Bolívar	Feb	0.00	4.62	3.74	0.00	3.80	0.00	4.36	0.00	2.90	4.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.94	4.44	3.74	0.00	0.00	0.00	0.00	0	North	Rainy
	Feb	0.00	3.37	3.85	0.00	0.00	0.00	0.00	3.59	0.00	4.04	3.67	0.00	0.00	0.00	0.00	0.00	0.00	2.90	0.00	0.00	0.00	0.00	0.00	0.00	0	Centre North	Rainy
	Feb	0.00	4.46	0.00	0.00	0.00	0.00	0.00	0.00	4.37	0.00	3.59	0.00	3.74	3.67	0.00	0.00	0.00	0.00	0.00	0.00	3.37	0.00	0.00	0.00	0	Centre South	Rainy
Esmeraldas Manta La Libertad Pro. Bolívar	Mar	4.33	4.17	4.66	0.00	2.90	0.00	4.62	0.00	4.69	4.36	3.20	3.74	3.80	0.00	3.37	3.20	3.80	0.00	3.50	0.00	3.50	0.00	4.13	3	North	Rainy	
	Mar	4.01	4.28	4.15	3.20	3.20	0.00	5.58	2.90	4.54	3.74	4.34	4.90	4.55	3.50	4.10	3.37	4.22	4.24	3.80	2.90	2.90	2.90	4.10	3	Centre North	Rainy	
	Mar	5.35	4.52	3.80	0.00	2.90	0.00	4.52	0.00	4.28	4.41	2.90	0.00	0.00	3.20	0.00	4.22	3.50	3.67	0.00	3.50	0.00	0.00	3.50	2	Centre South	Rainy	
Esmeraldas Manta La Libertad Pro. Bolívar	Apr	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	South	Rainy	
	Apr	4.69	4.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.51	3.37	4.01	5.22	3.74	5.25	0.00	4.29	4.10	4.04	3.80	2.90	3.37	4.01	0	North	Rainy	
	Apr	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.46	0.00	3.74	0.00	3	Centre North	Rainy	
Esmeraldas Manta La Libertad Pro. Bolívar	May	0.00	4.51	2.90	0.00	2.90	0.00	4.45	3.37	4.59	4.07	2.90	5.46	4.58	0.00	4.73	3.85	0.00	3.20	3.67	3.67	0.00	0.00	0.00	0	North	Rainy	
	May	4.01	4.22	3.59	0.00	0.00	0.00	0.00	0.00	3.94	4.41	0.00	2.90	0.00	4.47	3.67	0.00	0.00	0.00	2.90	3.59	0.00	2.90	0.00	0	Centre North	Rainy	
	May	3.80	3.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.90	0.00	0.00	0.00	3.20	1	Centre South	Rainy	
Esmeraldas Manta La Libertad Pro. Bolívar	Jun	4.94	5.25	4.34	4.74	3.74	4.72	0.00	3.74	0.00	4.15	4.87	0.00	4.34	0.00	0.00	5.88	3.50	0.00	3.37	3.20	0.00	0.00	4.04	3	North	Dry	
	Jun	4.76	5.03	5.12	3.90	4.89	0.00	4.47	2.90	0.00	4.53	3.20	4.73	0.00	0.00	4.75	3.37	2.90	0.00	3.20	0.00	0.00	0.00	0.00	0	Centre North	Dry	
	Jun	4.67	5.22	4.41	0.00	4.62	0.00	4.61	2.90	0.00	4.40	4.01	4.58	0.00	0.00	4.93	3.94	3.80	0.00	4.41	3.20	0.00	3.20	0.00	4.01	3	Centre South	Dry
Esmeraldas Manta La Libertad Pro. Bolívar	Jul	5.01	6.16	4.01	5.06	4.99	0.00	4.34	4.61	3.59	4.15	4.95	3.20	4.10	0.00	0.00	5.87	0.00	3.67	0.00	4.01	3.20	0.00	4.10	3	South	Dry	
	Jul	4.86	5.14	4.99	4.49	4.41	4.28	3.67	4.40	3.74	3.74	4.68	0.00	4.49	5.27	0.00	4.44	0.00	3.90	0.00	0.00	0.00	3.37	0.00	3.20	1	North	Dry
	Jul	2.90	0.00	4.29	4.78	0.00	0.00	0.00	0.00	4.07	3.37	3.20	4.34	3.59	4.20	0.00	0.00	3.20	4.04	3.20	2.90	2.90	2.90	4.13	3	Centre North	Dry	
Esmeraldas Manta La Libertad Pro. Bolívar	Aug	4.39	5.04	4.66	4.37	4.13	3.94	0.00	4.28	3.97	4.46	4.51	3.50	4.56	3.94	0.00	3.67	3.67	4.01	0.00	3.20	2.90	0.00	3.67	2	North	Dry	
	Aug	4.86	5.22	5.84	5.57	4.24	4.40	0.00	4.64	3.85	4.59	5.18	3.80	4.67	3.80	0.00	4.45	4.56	3.37	0.00	0.00	0.00	3.50	4.45	3	Centre North	Dry	
	Aug	4.07	4.94	5.83	5.71	4.71	0.00	4.13	3.94	3.20	3.67	5.01	4.51	4.51	0.00	0.00	4.68	4.65	4.37	3.20	3.80	0.00	2.90	0.00	0	Centre South	Dry	
Esmeraldas Manta La Libertad Pro. Bolívar	Sep	3.37	4.93	4.66	3.50	4.26	0.00	0.00	0.00	4.07	3.50	0.00	4.44	0.00	0.00	3.20	3.20	3.20	4.04	0.00	3.37	0.00	3.90	3.80	2	North	Dry	
	Sep	4.53	5.02	4.95	3.94	3.94	0.00	4.92	3.94	3.74	5.12	0.00	4.41	0.00	0.00	0.00	3.37	0.00	3.90	0.00	2.90	2.90	3.59	3.67	2	Centre North	Dry	
	Sep	0.00	4.33	4.82	2.90	0.00	0.00	0.00	4.41	3.20	4.10	4.34	0.00	4.01	0.00	0.00	0.00	0.00	4.17	0.00	3.37	0.00	2.90	0.00	0	Centre South	Dry	
Esmeraldas Manta La Libertad Pro. Bolívar	Oct	0.00	3.37	3.20	0.00	0.00	0.00	0.00	0.00	4.59	3.67	0.00	2.90	3.80	0.00	0.00	0.00	3.37	4.46	0.00	3.90	0.00	0.00	4.40	3	South	Dry	
	Oct	0.00	3.85	3.97	0.00	3.20	0.00	0.00	3.37	3.59	0.00	3.37	3.20	0.00	0.00	0.00	0.00	0.00	4.04	0.00	3.59	0.00	3.59	0.00	0	North	Dry	
	Oct	4.10	4.69	3.80	4.10	3.59	4.29	0.00	3.74	3.85	4.22	4.39	3.59	4.40	0.00	0.00	3.67	3.50	4.51	0.00	3.74	0.00	3.20	3.50	2	Centre North	Dry	
Esmeraldas Manta La Libertad Pro. Bolívar	Nov	4.46	5.03	4.20	4.55	4.50	0.00	0.00	3.80	3.97	4.45	4.62	4.07	3.80	0.00	4.43	3.90	4.33	0.00	3.90	0.00	4.36	4.31	3	North	Dry		
	Nov	4.22	5.08	4.07	3.80	4.40	5.94	0.00	4.67	4.39	4.07	5.62	3.97	5.09	3.80	0.00	4.69	4.40	4.15	0.00	3.74	0.00	3.90	3.85	2	Centre North	Dry	
	Nov	4.45	5.31	6.14	4.07	3.50	0.00	4.37	5.12	4.24	4.15	5.12	3.50	4.57	0.00	3.80	4.26	4.31	0.00	3.94	0.00	3.37	0.00	0.00	0	Centre South	Dry	
Esmeraldas Pro. Bolívar	Nov	3.80	0.00	0.00	0.00	0.00	0.00	0.00	4.87	3.85	3.20	4.13	3.50	3.74	0.00	3.20	4.31	3.37	0.00	3.20	0.00	0.00	2.90	0.00	0	South	Dry	

Table A2. Cont.

Space	Time	Phytoplankton Species ¹ —Abundance Cells L ⁻¹																								Localization	Sea-son
		e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13	e14	e15	e16	e17	e18	e19	e20	e21	e22	e23	e24		
Esmeraldas	Dec	4.24	4.64	3.20	4.75	4.20	3.97	3.20	0.00	3.85	3.37	4.78	3.94	4.26	0.00	0.00	3.85	0.00	3.97	0.00	3.85	2.90	3.80	4.26	3	North	Rainy
Manta	Dec	4.65	5.07	4.26	4.68	3.80	0.00	4.44	4.15	4.50	0.00	4.59	3.74	4.57	0.00	0.00	4.33	3.97	4.13	0.00	3.59	0.00	4.15	0.00	0	Centre North	Rainy
La Libertad	Dec	4.91	5.09	5.21	4.28	4.52	4.43	5.35	5.35	4.31	4.75	4.26	4.26	4.59	0.00	0.00	4.77	3.74	3.74	0.00	3.20	4.13	3.85	2	Centre South	Rainy	
Pto. Bolívar	Dec	4.40	5.43	4.61	0.00	4.22	0.00	4.01	5.25	4.29	3.74	4.75	4.01	4.15	0.00	0.00	3.74	3.85	3.94	0.00	3.90	0.00	4.49	0.00	0	South	Rainy

¹ *Dactylosolen fragilissimus* (e1), *Grinardia striata* (e2), *Rhizosolenia imbricata* (e3), *Dactylosolen antarcticus* (e4), *Probesia alata* (e5), *Skatlonema costatum* (e6), *Lauderia borealis* (e7), *Nitzschia longissima* (e8), *Nitzschia* sp. (e9), *Pseudo-nitzschia pungens* (e10), *Leptocylindrus danicus* (e11), *Thalassosira* sp. (e12), *Chaetoceros affinis* (e13), *Chaetoceros ditymus* (e15), *Hemiallus sinensis* (e16), *Thalassionema nitzschoides* (e17), *Gymnodinium* sp. (e18), *Gyrodinium* sp. (e19), *Mesodinium rubrum* (e20), *Ditylum brightwellii* (e21), *Nauticula* sp. (e22), *Stauroneis membranacea* (e23)—continuous variable), *Stauroneis membranacea* (e24—ordinal variable).

References

- United Nations Agua. Available online: <https://www.un.org/es/sections/issues-depth/water/index.html> (accessed on 10 March 2021).
- Checa Artos, M.; Castillo, D.; Ruiz-Barzola, O.; Barcos-Arias, M. Presence of pharmaceutical products in water and its impact on the environment. *Bionatura* **2021**, *6*, 1618–1627. [[CrossRef](#)]
- Sperotto, A.; Molina, J.L.; Torresan, S.; Critto, A.; Pulido-Velazquez, M.; Marcomini, A. Water quality sustainability evaluation under uncertainty: A multi-scenario analysis based on bayesian networks. *Sustainability* **2019**, *11*, 4764. [[CrossRef](#)]
- Adger, W.N. Vulnerability. *Glob. Environ. Chang.* **2006**, *16*, 268–281. [[CrossRef](#)]
- Ding, Q.; Shi, X.; Zhuang, D.; Wang, Y. Temporal and spatial distributions of ecological vulnerability under the Influence of natural and anthropogenic factors in an eco-province under construction in China. *Sustainability* **2018**, *10*, 3087. [[CrossRef](#)]
- Gyamfi, C.; Ndambuki, J.M.; Salim, R.W. Simulation of sediment Yield in a semi-arid river Basin under changing land Use: An integrated approach of hydrologic modelling and Principal Component Analysis. *Sustainability* **2016**, *8*, 1133. [[CrossRef](#)]
- Molina, J.L.; Zazo, S.; Martín-Casado, A.M.; Patino-Alonso, M.C. Rivers' temporal sustainability through the evaluation of predictive runoff methods. *Sustainability* **2020**, *12*, 1720. [[CrossRef](#)]
- Lake, Q.; Gu, Q.; Zhang, Y.; Ma, L.; Li, J.; Wang, K.; Zheng, K.; Zhang, X. Assessment of reservoir water quality using multivariate statistical techniques: A case study. *Sustainability* **2016**, *8*, 243. [[CrossRef](#)]
- Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **1901**, *2*, 559–572. [[CrossRef](#)]
- Hotelling, H. Analysis of complex statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [[CrossRef](#)]
- Hotelling, H. Simplified calculation of principal components. *Psychometrika* **1936**, *1*, 27–35. [[CrossRef](#)]
- Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [[CrossRef](#)]
- Thurstone, L.L. *Multiple Factor Analysis*; University of Chicago Press: Chicago, IL, USA, 1947.
- Benzécri, J.P. L'analyse des données: L'analyse des correspondances. In *L'Analyse des Données: Leçons sur L'analyse Factorielle et la Reconnaissance des Formes et Travaux*, 2nd ed.; Benzécri, J.P., Ed.; Dunod: Paris, France, 1973; ISBN 9782040072254.
- Hill, M.O. Reciprocal averaging: An eigenvector method of ordination. *J. Ecol.* **1973**, *61*, 237–249. [[CrossRef](#)]
- Tenenhaus, M.; Young, F.W. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika* **1985**, *50*, 91–119. [[CrossRef](#)]
- Ter Braak, C.J.F. Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Gradient Anal. Author Cajo J. F. Ter Braak Source Ecol. Ecol.* **1986**, *67*, 1167–1179. [[CrossRef](#)]
- Legendre, P.; Legendre, L. *Numerical Ecology*, 2nd ed.; Elsevier: Amsterdam, The Netherlands, 1998; ISBN 9780444538680.
- Escofier, B.; Pagès, J. L'analyse factorielle multiple. *Cah. Bur. Univ. Rech. Opérationnelle Série Rech.* **1984**, *42*, 3–68.
- Escofier, B.; Pagès, J. Analyses factorielles simples et multiples: Objectifs. *Méthodes Interpret.* **1990**, *1*, 284.
- Escofier, B.; Pagès, J. Multiple factor analysis (AFMULT package). *Comput. Stat. Data Anal.* **1994**, *18*, 121–140. [[CrossRef](#)]
- L'Hermier Des Plantes, H. *Structuration des Tableaux a Trois Indices de la Statistique: Theorie et Application d'une Methode D'analyse Conjointe*; Université des Sciences et Techniques du Languedoc: Montpellier, France, 1976.
- Lavit, C.; Escoufier, Y.; Sabatier, R.; Traissac, P. The ACT (STATIS method). *Comput. Stat. Data Anal.* **1994**, *18*, 97–119. [[CrossRef](#)]
- Jaffrenou, P.A. Sur L'analyse des Familles Finies de Variables Vectorielles: Bases Algébriques et Application a la Description Statistique. Ph.D Thesis, Université Lyon 1, Villeurbanne, France, 1978.
- Gaertner, J.C.; Chessel, D.; Bertrand, J. Stability of spatial structures of demersal assemblages: A multitable approach. *Aquat. Living Resour.* **1998**, *11*, 75–85. [[CrossRef](#)]
- Simier, M.; Blanc, L.; Pellegrin, F.; Nandris, D. Approche simultanée de K couples de tableaux: Application à l'étude des relations pathologie végétale-environnement. *Rev. Stat. Appliquée* **1999**, *47*, 31–46.
- Thioulouse, J.; Simier, M.; Chessel, D. Simultaneous analysis of a sequence of paired ecological tables. *Ecology* **2004**, *85*, 272–283. [[CrossRef](#)]
- Vivien, M.; Sabatier, R. A generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables. *Comput. Stat. Data Anal.* **2004**, *46*, 155–171. [[CrossRef](#)]
- Sauzay, L.; Hanafi, M.; Qannari, E.M.; Schlich, P. *Analyse de K+ 1 Tableaux a L'aide De La Méthode STATIS: Application en Évaluation Sensorielle*; 9ieme Journées Européennes Agro-Industrie et Méthodes Statistiques; Société Française de Statistique (SFdS): Montpellier, France, 2006.
- Abdi, H.; Valentin, D.; Chollet, S.; Chrea, C. Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Qual. Prefer.* **2007**, *18*, 627–640. [[CrossRef](#)]
- Sabatier, R.; Vivien, M. A new linear method for analyzing four-way multiblocks tables: STATIS-4. *J. Chemom.* **2008**, *22*, 399–407. [[CrossRef](#)]
- Marcondes Filho, D.; Fogliatto, F.S.; de Oliveira, L.P.L. Multivariate control charts for monitoring non-linear batch processes. *Producao* **2011**, *21*, 132–148. [[CrossRef](#)]
- Thioulouse, J. Simultaneous analysis of a sequence of paired ecological tables: A comparison of several methods. *Ann. Appl. Stat.* **2011**, *5*, 2300–2325. [[CrossRef](#)]
- Bénasséni, J.; Bennani Dosse, M. Analyzing multiset data by the Power STATIS-ACT method. *Adv. Data Anal. Classif.* **2012**, *6*, 49–65. [[CrossRef](#)]

35. Abdi, H.; Williams, L.J.; Valentin, D.; Bennani-Dosse, M. STATIS and DISTATIS: Optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdiscip. Rev. Comput. Stat.* **2012**, *4*, 124–167. [[CrossRef](#)]
36. Sabatier, R.; Vivien, M.; Reynès, C.; Myrtille, V.; Christelle, R. Une nouvelle proposition, l'Analyse Discriminante Multitableaux: STATIS-LDA. *J. Soc. Fr. Stat. Rev. Stat. Appliquée* **2013**, *154*, 31–43.
37. Rodríguez, O.; Corrales, D. Interstat: The STATIS method for interval valued data. *Rev. Mat. Teory Appl.* **2014**, *21*, 73–83.
38. Sautron, V.; Chavent, M.; Viguier, N.; Villa-Vialaneix, N. Multiway-SIR for Longitudinal Multi-Table Data Integration. In Proceedings of the 22nd International Conference on Computational Statistics (COMPSTAT), Oviedo, Spain, 23–26 August 2016; p. 1.
39. Kriegsmann, M.A. Discriminant DiSTATIS: A Multi-Way Discriminant Analysis for Distance Matrices, Illustrations with the Sorting Task. Ph.D. Thesis, University of Texas at Dallas, Richardson, TX, USA, 2018.
40. Llobell, F.; Cariou, V.; Vigneau, E.; Labenne, A.; Qannari, E.M. A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *Food Qual. Prefer.* **2019**, *72*, 31–39. [[CrossRef](#)]
41. Méricot, B.; Gaertner, J.C.; Brind'amour, A.; Carbonara, P.; Esteban, A.; Garcia-Ruiz, C.; Gristina, M.C.; Imzilen, T.; Jadaud, A.; Joksimovic, A.; et al. Stability of the relationships among demersal fish assemblages and environmental-trawling drivers at large spatio-temporal scales in the northern mediterranean sea. *Sci. Mar.* **2019**, *83*, 153–163. [[CrossRef](#)]
42. Llobell, F.; Cariou, V.; Vigneau, E.; Labenne, A.; Qannari, E.M. Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. *Food Qual. Prefer.* **2020**, *79*, 1–9. [[CrossRef](#)]
43. Tucker, L.R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **1966**, *31*, 279–311. [[CrossRef](#)]
44. Tucker, L.R. Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika* **1972**, *37*, 3–27. [[CrossRef](#)]
45. Tucker, L.R. Three-mode Factor Analysis Applied to Multidimensional Scaling. In Proceedings of the US-Japan Conference on Multidimensional Scaling, La Jolla, CA, USA, August 1975.
46. Israelsson, A. Three-way (or second order) component analysis. *Nonlinear Iterative Partial Least-Sq. Estim. Proced. Bull. Int. Stat. Inst.* **1969**, *43*, 29–51.
47. Kroonenberg, P.M.; De Leeuw, J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **1980**, *45*, 69–97. [[CrossRef](#)]
48. Harshman, R.A. Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multimodal factor analysis. *UCLA Work. Pap. Phon.* **1970**, *16*, 1–84.
49. Carroll, J.D.; Chang, J.J. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika* **1970**, *35*, 283–319. [[CrossRef](#)]
50. Carroll, J.D. IDIOSCAL (Individual Difference In Orientation SCALing): A generalization of INDSCAL allowing IDIOSyncratic reference systems as well as analytic approximation to INDSCAL. In Proceedings of the Paper Presented at Meeting of Psychometric Society, Princeton, NJ, USA, 20 August 1972.
51. Harshman, R.A. Models for analysis of Asymmetrical Relationships Among N Objects or Stimuli. In Proceedings of the First Joint Meeting of the Psychometric Society and the Society of Mathematical Psychology, Hamilton, ON, Canada, 30 March 1978.
52. Carroll, J.D.; Pruzansky, S.; Kruskal, J.B. CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika* **1980**, *45*, 3–24. [[CrossRef](#)]
53. Harshman, R.A.; Lundy, M.E. Uniqueness proof for a family of models sharing features of Tucker’s three-mode factor analysis and PARAFAC/CANDECOMP. *Psychometrika* **1996**, *61*, 133–154. [[CrossRef](#)]
54. Bahram, M.; Bro, R. A novel strategy for solving matrix effect in three-way data using parallel profiles with linear dependencies. *Anal. Chim. Acta* **2007**, *584*, 397–402. [[CrossRef](#)]
55. de Almeida, A.L.F.; Favier, G.; Mota, J.C.M. A constrained factor decomposition with application to MIMO antenna systems. *IEEE Trans. Signal Process.* **2008**, *56*, 2429–2442. [[CrossRef](#)]
56. Giordani, P.; Rocci, R. Candecomp/Parafac with ridge regularization. *Chemom. Intell. Lab. Syst.* **2013**, *129*, 3–9. [[CrossRef](#)]
57. Giordani, P.; Rocci, R. Constrained CP via the Lasso. *Psychometrika* **2013**, *78*, 669–684. [[CrossRef](#)]
58. Martín-Rodríguez, J. Contribuciones a la Integración de Subespacios Desde una Perspectiva Biplot. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 1996.
59. Martín-Rodríguez, J.; Galindo-Villardón, M.P.; Vicente-Villardón, J.L. Comparison and integration of subspaces from a biplot perspective. *J. Stat. Plan. Inference* **2002**, *102*, 411–423. [[CrossRef](#)]
60. Bacalá, N. Contribuciones al Análisis de Matrices de Datos Multivía: Tipología de las Variables. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2004.
61. Vallejo-Arboleda, A.; Vicente-Villardón, J.L.; Galindo-Villardón, M.P. Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology. *Comput. Stat. Data Anal.* **2007**, *51*, 4193–4205. [[CrossRef](#)]
62. Cortés Saud, Á. Contribuciones al Análisis de Tablas de Tres Vías Restringido. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2005.
63. Basso, L.C. Análisis Conjunto de Varias Matrices de Datos: Contribuciones a la Tipología de los Individuos. Ph.D. Thesis, Universidad Salamanca, Salamanca, Spain, 2006.
64. Vallejo, A.; Vicente, J.L.; Galindo, P.; Fernández, M.; Fernández, C.; Bécarea, E. Análisis de la evolución en el tiempo para datos con estructura de grupos: STATIS dual canónico y modelo de medidas repetidas doblemente multivariantes. *Rev. Colomb. Estad.* **2008**, *31*, 321–340.

65. Pinzón Sarmiento, L.M. Biplot Consenso para Análisis de Tablas Múltiples. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2011.
66. Mendes, S.; Fernández-Gómez, M.J.; Marques, S.C.; Pardal, M.Á.; Azeiteiro, U.M.; Galindo-Villardón, M.P. CO-tucker: A new method for the simultaneous analysis of a sequence of paired tables. *J. Appl. Stat.* **2017**, *44*, 2729–2755. [[CrossRef](#)]
67. Mendes, S. Métodos Multivariantes para Evaluar Patrones de Estabilidad y Cambio Desde Una Perspectiva Biplot. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2011.
68. Frutos Bernal, E. Análisis de Datos Acoplados: Modelo T3-PCA. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2014.
69. Egido, J.; Galindo, P. Dynamic Biplot. Evolution of the Economic Freedom in the European Union. *Br. J. Appl. Sci. Technol.* **2015**, *11*, 1–13. [[CrossRef](#)]
70. Egido Miguélez, J.F. Biplot Dinámico. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2015.
71. Rodríguez Rosa, M. Contribuciones al Análisis de la Sostenibilidad Internacional, Desde Una Perspectiva Algebraica Multivariante Comparada. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2016.
72. Rodríguez-Rosa, M.; Gallego-Álvarez, I.; Galindo-Villardón, M.P. Spatio-temporal analysis of economic, social, and environmental issues in the framework of sustainable development in worldwide countries. *Sustain. Dev.* **2019**, *27*, 429–447. [[CrossRef](#)]
73. González García, N. Análisis Sparse de Tensores Multidimensionales. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2019.
74. Lavit, C.; Perrin, M.-O. Multivariate and longitudinal data on growing children: Solution using STATIS. In *Data Analysis. The Ins and Outs of Solving Real Problems*; Janssen, J., Marcotorchino, F., Proth, J.M., Eds.; Plenum: New York, NY, USA, 1987; pp. 13–29.
75. Arcidiacono, C.; Sarnacchiaro, P.; Velleman, R. Testing fidelity to a new psychological intervention for family members of substance misusers during implementation in Italy. *J. Subst. Use* **2008**, *13*, 361–381. [[CrossRef](#)]
76. Bono, F.; Giacomarra, M. The photovoltaic growth in the European Union requires stronger RES support. *J. Policy Model.* **2016**, *38*, 324–339. [[CrossRef](#)]
77. Ramos-Barberán, M.; Hinojosa-Ramos, M.V.; Ascencio-Moreno, J.; Vera, F.; Ruiz-Barzola, O.; Galindo-Villardón, M.P. Batch process control and monitoring: A Dual STATIS and Parallel Coordinates (DS-PC) approach. *Prod. Manuf. Res.* **2018**, *6*, 470–493. [[CrossRef](#)]
78. Niang, N.; Fogliatto, F.S.; Saporta, G. Batch process monitoring by three-way data analysis approach. In Proceedings of the XIIIth International Conference Applied Stochastic Models and Data Analysis ASMDA, Vilnius, Lithuania, 30 June–3 July 2009.
79. Niang, N.; Fogliatto, F.; Saporta, G. Non parametric on-line control of batch processes based on STATIS and clustering. *J. Soc. Fr. Stat. Stat.* **2013**, *154*, 124–142.
80. Scepi, G. Parametric and non parametric multivariate quality control charts. In *Multivariate Total Quality Control*; Lauro, C., Antoch, J., Esposito Vinzi, V., Sporta, G., Eds.; Physica-Verlag HD: New York, NY, USA, 2002; pp. 163–189. ISBN 9788578110796.
81. Génard, M.; Souty, M.; Holmes, S.; Reich, M.; Breuils, L. Correlations among quality parameters of peach fruit. *J. Sci. Food Agric.* **1994**, *66*, 241–245. [[CrossRef](#)]
82. Meyners, M.; Kunert, J.; Qannari, E.M. Comparing generalized procrustes analysis and statis. *Food Qual. Prefer.* **2000**, *11*, 77–83. [[CrossRef](#)]
83. Chaya, C.; Perez-Hugalde, C.; Judez, L.; Wee, C.S.; Guinard, J.X. Use of the STATIS method to analyze time-intensity profiling data. *Food Qual. Prefer.* **2003**, *15*, 3–12. [[CrossRef](#)]
84. Lavit, C.; Perez Hugalde, C. Application de la méthode STATIS à des données économiques: Évolution des secteurs agricoles et non agricoles des provinces espagnoles de 1960 à 1979. *Quatr. J. Int. Anal. des Données Inform.* **1985**, *9*, 11.
85. Lavit, C. Application de la méthode STATIS. *Stat. Anal. Données* **1985**, *10*, 103–116.
86. Sandoval Solís, L.; Linares Fleites, G.; Quiroz Clemente, A.; Romero Vázquez, M.G. Implementacion del Método Statis en R y su aplicación al estudio de satisfacción de usuarios del sistema bibliotecario de la BUAP. *Rev. Investig. Oper.* **2012**, *32*, 56–66.
87. Ubertalli, B.; Perrin, M.-O. Utilisation d' une méthode multi-tableaux en sciences sociales. Une étude longitudinale de carrières: Les 12 premières promotions de l' école d' infirmières de Roanne. *Population* **1990**, *45*, 1092–1100. [[CrossRef](#)]
88. Volkova, M. Russian and European Population's Quality of Life Analysis with the Instruments of Common Principal Components (CPC). *Ekonom. Mat. Metod.* **2019**, *55*, 34–46. [[CrossRef](#)]
89. Raymond, O.; Fiasson, J.L.; Jay, M. Synthetic taxonomy of Rosa races using ACT-STATIS. *Z. Naturforsch. C.* **2000**, *55*, 399–409. [[CrossRef](#)]
90. Grossi, C.; Raymond, O.; Sanlaville-Boisson, C.; Jay, M. Rosa taxonomy and hierarchy of markers defined by ACT STATIS. *Z. Naturforsch. C* **1999**, *54*, 25–34. [[CrossRef](#)]
91. Coquet, R.; Troxler, L.; Wipff, G. The STATIS method: Characterization of conformational states of flexible molecules from molecular dynamics simulations in solution. *J. Mol. Graph.* **1996**, *14*, 206–212. [[CrossRef](#)]
92. L'Hermier Des Plantes, H.; Thiébaud, B. Étude de la pluviométrie au moyen de la méthode STATIS. *Rev. Stat. Appliquée* **1977**, *25*, 57–81.
93. Márquez, J.L.; Díaz de Pascual, A.; Defives, G. Aplicación del método Statis: Factores físico-químicos del agua del embalse Uribante. *Economía* **1992**, *17*, 35–58.
94. Stanimirova, I.; Walczak, B.; Massart, D.L.; Simeonov, V.; Saby, C.A.; Di Crescenzo, E. STATIS, a three-way method for data analysis. Application to environmental data. *Chemom. Intell. Lab. Syst.* **2004**, *73*, 219–233. [[CrossRef](#)]

95. Mendes, S.; Marques, S.C.; Azeiteiro, U.M.; Fernández-Gómez, M.J.; Galindo-Villardón, M.P.; Maranhão, P.; Morgado, F.; Leandro, S.M. Zooplankton Distribution in a Marine Protected Area: The Berlengas Natural Reserve (Western Coast of Portugal). *Fresenius Environ. Bull.* **2011**, *20*, 496–505.
96. Mendes, S.; Fernández-Gómez, M.J.; Pereira, M.J.; Azeiteiro, U.M.; Galindo-Villardón, M.P. An empirical comparison of Canonical Correspondence Analysis and STATICO in the identification of spatio-temporal ecological relationships. *J. Appl. Stat.* **2012**, *39*, 979–994. [[CrossRef](#)]
97. Mendes, S.; Fernández-Gómez, M.J.; Resende, P.; Jorge Pereira, M.; Galindo-Villardón, M.P.; Azeiteiro, U.M. Spatio-temporal structure of diatom assemblages in a temperate estuary. A STATICO analysis. *Estuar. Coast. Shelf Sci.* **2009**, *84*, 637–644. [[CrossRef](#)]
98. Marques, S.C.; Primo, A.L.; Falcão, J.; Martinho, F.; Mendes, S.; Azeiteiro, U.M.; Pardal, M.A. The Impact of Conspicuous Environmental Changes on the Spatial and Temporal Dynamics of *Acartia Tonsa* and *Acartia Clausi*: A decadal Study in a Temperate Estuary (Mondego, Portugal). In *Trends in Copepod Studies: Distribution, Biology and Ecology*; Uttieri, M., Ed.; Nova Science Publishers Inc.: Hauppauge, NY, USA, 2018; pp. 145–171. ISBN 978-1-53612-593-1.
99. Slimani, N.; Guilbert, E.; Ayni, F.; El Jrad, A.; Boumaiza, M.; Thioulouse, J. The use of STATICO and COSTATIS, two exploratory three-ways analysis methods: An application to the ecology of aquatic heteroptera in the Medjerda watershed (Tunisia). *Environ. Ecol. Stat.* **2017**, *24*, 269–295. [[CrossRef](#)]
100. Feki-Sahnoun, W.; Hamza, A.; Béjaoui, B.; Mahfoudi, M.; Rebai, A.; Bel Hassen, M. Multi-table approach to assess the biogeography of phytoplankton: Ecological and management implications. *Hydrobiologia* **2018**, *815*, 229–251. [[CrossRef](#)]
101. Rodríguez-Martínez, C.C.; García-Sánchez, I.M.; Vicente-Galindo, P.; Galindo-Villardón, P. Exploring relationships between environmental performance, E-Government and corruption: A multivariate perspective. *Sustainability* **2019**, *11*, 6497. [[CrossRef](#)]
102. Amor-Esteban, V.; Galindo-Villardón, M.P.; García-Sánchez, I.M. Industry mimetic isomorphism and sustainable development based on the X-STATIS and Hj-biplot methods. *Environ. Sci. Pollut. Res.* **2018**, *25*, 26192–26208. [[CrossRef](#)]
103. Martínez-Córdoba, P.J.; Amor-Esteban, V.; Benito, B.; García-Sánchez, I.M. The commitment of Spanish local governments to sustainable development goal 11 from a multivariate perspective. *Sustainability* **2021**, *13*, 1222. [[CrossRef](#)]
104. Gudmundsson, L.; Tallaksen, L.M.; Stahl, K. Spatial cross-correlation patterns of European low, mean and high flows. *Hydrol. Process.* **2011**, *25*, 1034–1045. [[CrossRef](#)]
105. Fournier, M.; Motelay-Massei, A.; Massei, N.; Aubert, M.; Bakalowicz, M.; Dupont, J.P. Investigation of transport processes inside karst aquifer by means of STATIS. *Ground Water* **2009**, *47*, 391–400. [[CrossRef](#)]
106. González-Narváez, M. Distribución Espacio-Temporal Del Fitoplancton en el Pacífico Ecuatorial Oriental, Zona de la Región Niño 1 + 2, Aplicación del Método STATICO. Master's Thesis, Universidad de Salamanca, Salamanca, Spain, 2016.
107. Eaton, A.D.; Clesceri, L.S.; Rice, E.W.; Greenberg, A.E. *Standard Methods for the Examination of Water and Wastewater*, 21st ed.; American Public Health Association: Washington, DC, USA, 2005.
108. Strickland, J.D.H.; Parsons, T.R. *A Practical Handbook of Seawater Analysis*, 2nd ed.; Supply and Services Canada: Ottawa, ON, Canada, 1972; Volume 167.
109. Utermöhl, H. *Zur Vervollkommnung der Quantitativen Phytoplankton-Methodik*; Schweizerbart: Stuttgart, Germany, 1958.
110. Reguera, B.; Alonso, R.; Moreira, A.; Méndez, S. Guía para el diseño y puesta en marcha de un plan de seguimiento de microalgas productoras de toxinas. *Com. Ocean. Intergub.* **2011**, *59*, 1–65.
111. Torres, G. Evaluación del fitoplancton Como un Mecanismo Preventivo a la Ocurrencia de Bloom Algal Frente a las Costas de Esmeraldas, Manta, La Libertad y Puerto Bolívar en Ecuador 2013–2015. Ph.D. Thesis, Universidad Nacional Mayor de San Marcos, Perú, Lima, 2017.
112. Jiménez, R. Diatomeas y silicoflagelados del fitoplancton del Golfo de Guayaquil II Edición. *Acta Ocean. Pacífico* **1983**, *2*, 193–281.
113. Pesantes, F. Los dinoflagelados como indicadores de El Niño en el mar ecuatorial. *Acta Ocean. Pacífico* **1983**, *2*, 1–117.
114. Balech, E. *Los dinoflagelados del Atlántico Sudoccidental*; Ministerio de Agricultura Pesca y Alimentación; Secretaría General Técnica: Madrid, Spain, 1988; ISBN 84-7479-711-X.
115. Tomas, C.R. *Marine Phytoplankton: A Guide to Naked Flagellates and Coccolithophorids*; Academic Press: Cambridge, MA, USA, 1993; ISBN 978-0-12-693010-8.
116. Taylor, F.J.R.; Fukuyo, Y.; Larsen, J. *Taxonomy of Harmful Dinoflagellates in Manual on Harmful Marine Microalgae*; Hallegraeff, G.M., Anderson, D.M., Cembella, A.D., Eds.; IOC Manuals and Guides No 33; UNESCO: Paris, France, 1995; ISBN 2831708265.
117. Hasle, G.R.; Syvertsen, E.E.; Steidinger, K.A.; Tangen, K.; Tomas, C.R. *Identifying Marine Diatoms and Dinoflagellates*; Tomas, C.R., Ed.; Academic Press: San Diego, CA, USA, 1996; ISBN 978-0-12-693015-3.
118. Tomas, C.R. *Identifying Marine Phytoplankton*; Academic Press: San Diego, CA, USA, 1997; ISBN 9780080534428.
119. Young, J.; Geisen, M.; Cros, L.; Kleijne, A.; Sprengel, C.; Probert, I.; Østergaard, J. A guide to extant coccolithophore taxonomy. *J. Nannoplankt. Res. Spec.* **2003**, *1*, 1–132.
120. Jiménez, R. Diatomeas y silicoflagelados del fitoplancton del Golfo de Guayaquil III Edición. *Acta Ocean. Pacífico* **2014**, *19*, 1–89.
121. Kroonenberg, P.M. *Applied Multiway Data Analysis*; Wiley-Interscience: Hoboken, NJ, USA, 2008; ISBN 9780470164976.
122. Legendre, L.; Legendre, P. Le traitement multiple des données écologiques. *Ecol. Numer. Tom 1* **1979**, *66*, 775–776. [[CrossRef](#)]
123. Rao, C.R. The use and interpretation of Principal Component Analysis in applied research stable. *Indian J. Stat.* **1964**, *26*, 329–358.
124. Dolédec, S.; Chessel, D. Co-inertia analysis: An alternative method for studying species-environment relationships. *Freshw. Biol.* **1994**, *31*, 277–294. [[CrossRef](#)]
125. Gower, J.C. A general coefficient of similarity and some of its properties. *Biometrics* **1971**, *27*, 857–871. [[CrossRef](#)]

126. Escoufier, B.; Pagès, J. *Analyses Factorielles Simples et Multiples*; Dunod: Paris, France, 1998.
127. Escoufier, Y. Le Traitement des Variables Vectorielles. *Biometrics* **1973**, *29*, 751–760. [[CrossRef](#)]
128. Gower, J.C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **1966**, *53*, 325–338. [[CrossRef](#)]
129. INOCAR Capítulo I: Información general de la República del Ecuador Inocar 2012. Available online: https://www.inocar.mil.ec/docs/derrotero/derrotero_cap_1.pdf (accessed on 15 September 2019).
130. Santos, J.L. The impact of El Niño—Southern Oscillation Events on South America. *Adv. Geosci.* **2006**, *6*, 221–225. [[CrossRef](#)]
131. Schlitzer, R. Ocean Data View 2020. Available online: <https://odv.awi.de> (accessed on 23 May 2021).
132. Bastidas-Arteaga, E. Reliability of reinforced concrete structures subjected to corrosion-fatigue and climate change. *Int. J. Concr. Struct. Mater.* **2018**, *12*, 1–13. [[CrossRef](#)]
133. Basu, S.; Mackey, K.R.M. Phytoplankton as key mediators of the biological carbon pump: Their responses to a changing climate. *Sustainability* **2018**, *10*, 869. [[CrossRef](#)]
134. Zacchei, E.; Molina, J.L. Reviewing arch-dams' building risk reduction through a sustainability-safety management approach. *Sustainability* **2020**, *12*, 392. [[CrossRef](#)]
135. Anzecc, A. *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*; Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand: Canberra, Australia, 2000; pp. 1–103.
136. Wolanski, E.; Boorman, L.A.; Chicharo, L.; Langlois-Saliou, E.; Lara, R.; Plater, A.J.; Uncles, R.J.; Zalewski, M. Ecohydrology as a new tool for sustainable management of estuaries and coastal waters. *Wetl. Ecol. Manag.* **2004**, *12*, 235–276. [[CrossRef](#)]
137. Conde, A.; Hurtado, M.; Prado, M. Phytoplankton response to a weak El Niño event. *Ecol. Indic.* **2018**, *95*, 394–404. [[CrossRef](#)]
138. Torres, G.; Recalde, S.; Narea, R.; Rentería, W.; Troccoli, L. Variabilidad espacio-temporal del fitoplancton y variables oceanográficas en el Golfo de Guayaquil durante el 2013–15. *Rev. Inst. Investig. FIGMMG-UNMSM* **2017**, *20*, 70–79.
139. Torres, G.; Carnicer, O.; Canepa, A.; De La Fuente, P.; Recalde, S.; Narea, R.; Pinto, E.; Borbor-Córdova, M.J. Spatio-Temporal pattern of dinoflagellates along the Tropical Eastern Pacific coast (Ecuador). *Front. Mar. Sci.* **2019**, *6*, 1–20. [[CrossRef](#)]
140. Torres, G. Importancia ecológica del fitoplancton durante El Niño 1991–1993, en el Pacífico Ecuatorial (Ecuador). *Acta Ocean. Pacífico* **2005**, *13*, 35–49.
141. Torres, G.; Tapia, M.E. Distribución del fitoplancton en la región costera del mar ecuatorial, durante diciembre 2000. *Acta Ocean. Pacífico* **2002**, *11*, 62–72.

Manuscritos en proceso de ser sometidos

Título	Técnicas utilizadas	Tipo
Applied multivariate statistical analysis (STATICO method) for marine phytoplankton species assemblages off Ecuador: Geographic patterns and temporal variability	STATICO	Aplicación
Environment impact on pacific phytoplankton in Niño 1+2 region (Ecuador)	STATICO HJ-Biplot	Aplicación

Presentaciones en eventos científicos internacionales

Título	Año	Evento	Técnicas utilizadas	Tipo
Identificación de la variabilidad espacio temporal del Fitoplancton, en el Pacífico Ecuatorial Oriental, durante las condiciones del evento cálido 2015: un estudio basado en el modelo TUCKER3	2016	XXVI Simposio de Estadística 2016	Tucker 3	Comunicación Oral
Spatio-temporal variability of biological and environmental parameters in the presence of warm conditions in the Eastern Equatorial Pacific, Niño 1+2.	2018	International Meeting on Marine Research 2018	PTA STATICO	Póster
Behavior phytoplankton in the Eastern Equatorial Pacific, in relation to environmental variations caused by the presence of warm events	2018	IV International Conference on ENSO	STATICO HJ-Biplot	Comunicación Oral Póster
Temporal stability analysis of the species-environment relationships in the region of continental Ecuador, using the STATICO multi-way method	2019	IV International Workshop on Proximity Data Multivariate Analysis and Classification	STATICO	Póster
Comportamiento espacial y temporal del fitoplancton ante las variaciones ambientales en la región de Ecuador	2020	I Foro Internacional de Charlas Multidisciplinarias de Análisis de Datos	STATICO	Comunicación Oral

REFERENCIAS

*El día que clamé
me respondiste;
me fortaleciste con vigor
en mi alma
Salmos 138:3*

REFERENCIAS BIBLIOGRÁFICAS

- Abdi, H., Valentin, D., Chollet, S., & Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference*, 18(4), 627–640. <https://bit.ly/3ejgPTG>
- 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. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2), 124–167. <https://bit.ly/3hDMBN2>
- Atique, U., Kwon, S., & An, K.-G. (2020). Linking weir imprints with riverine water chemistry, microhabitat alterations, fish assemblages, chlorophyll-nutrient dynamics, and ecological health assessments. *Ecological Indicators*, 117, 106652.
- Bénasséni, J., & Bennani Dosse, M. (2012). Analyzing multiset data by the Power STATIS-ACT method. *Advances in Data Analysis and Classification*, 6(1), 49–65. <https://bit.ly/3ehz4bS>
- Benzécri, J. P. (1973). *L'analyse des données: 2 L'analyse des correspondances* (J. P. Benzécri (ed.); 22nd ed.). Dunod. <https://bit.ly/3reGvWD>
- Bonada, N., Rieradevall, M., Prat, N., & Resh, V. H. (2006). Benthic macroinvertebrate assemblages and macrohabitat connectivity in Mediterranean-climate streams of northern California. *Journal of the North American Benthological Society*, 25(1), 32–43.
- Borbor-Cordova, M. J., Torres, G., Mantilla-Saltos, G., Casierra-Tomala, A., Bermúdez, J. R., Renteria, W., & Bayot, B. (2019). Oceanography of Harmful Algal Blooms on the Ecuadorian Coast (1997–2017): Integrating Remote Sensing and Biological Data. *Frontiers in Marine Science*, 6, 1–19. <https://bit.ly/3wJ1x0G>
- Bouroche, J. M. (1975). *Analyse des données ternaires: la double analyse en composantes principales. These de 3-eme cycle*. Paris VI, France.
- Carassou, L., & Ponton, D. (2007). Spatio-temporal structure of pelagic larval

- and juvenile fish assemblages in coastal areas of New Caledonia, southwest Pacific. *Marine Biology*, 150(4), 697–711. <https://bit.ly/2U5CI1Y>
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3), 283–319. <https://bit.ly/3elFcQx>
- Cassis, D., Muñoz, P., & Avaria, S. (2002). Variación temporal del fitoplancton entre 1993 y 1998 en una estación fija del seno Aysén, Chile (45°26’S 73°00’W). *Revista de Biología Marina y Oceanografía*, 37(1), 43–65. <https://bit.ly/3xSSHic>
- Castillo-Escrivà, A., Valls, L., Rochera, C., Camacho, A., & Mesquita-Joanes, F. (2017). Metacommunity dynamics of Ostracoda in temporary lakes: Overall strong niche effects except at the onset of the flooding period. *Limnologica*, 62, 104–110. <https://bit.ly/3ibBABN>
- Çelekli, A., Kayhan, S., & Çetin, T. (2020). First assessment of lakes’ water quality in Aras River catchment (Turkey); Application of phytoplankton metrics and multivariate approach. *Ecological Indicators*, 117, 106706.
- Cruz, M., Tapia, M. E., & Naranjo, C. (2011). Especies planctónicas marinas como bioindicadoras de masas de agua frente a la costa ecuatoriana durante el evento El Niño 2009 - 2010. *Acta Oceanográfica Del Pacífico*, 16(1), 2010–2011.
- Cuadras, C. M. (2019). *Nuevos métodos multivariantes* (C. Ediciones (ed.)). <https://bit.ly/3wArzDv>
- Darwiche-Criado, N., Jiménez, J. J., Comin, F. A., Sorando, R., & Sánchez-Pérez, J. M. (2015). Identifying spatial and seasonal patterns of river water quality in a semiarid irrigated agricultural Mediterranean basin. *Environmental Science and Pollution Research*, 22(23), 18626–18636.
- De la Hoz Aristizábal, M. V. (2004). Dinámica del fitoplancton de la Ciénega Grande de Santa Marta, caribe colombiano. *Boletín de Investigaciones Marinas y Costeras.*, 33(864), 159–179.
- Diaz, S., Hodgson, J. G., Thompson, K., Cabido, M., Cornelissen, J. H. C., Jalili, A., Montserrat-Marti, G., Grime, J. P., Zarrinkamar, F., Asri, Y., & others.

- (2004). The plant traits that drive ecosystems: evidence from three continents. *Journal of Vegetation Science*, *15*(3), 295–304.
- Dolédéc, S., & Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology*, *31*, 277–294. <https://bit.ly/3hDmWUF>
- Dray, S., Chessel, D., & Thioulouse, J. (2003). CO-Inertia analysis and the linking of ecological data tables. *Ecology*, *84*(11), 3078–3089. <https://bit.ly/2U4l7aI>
- Dufrêne, M., & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs*, *67*(3), 345–366. <https://bit.ly/36zscmb>
- Escofier, B., & Pagès, J. (1984). L'analyse factorielle multiple. *Cahiers Du Bureau Universitaire de Recherche Opérationnelle. Série Recherche*, *42*, 3–68.
- Escofier, B., & Pagès, J. (1990). Analyses factorielles simples et multiples: Objectifs. *Méthodes et Interpretation*, *1*, 284.
- Escofier, B., & Pagès, J. (1994). Multiple factor analysis (AFMULT package). *Computational Statistics and Data Analysis*, *18*(1), 121–140. <https://bit.ly/3B2faLZ>
- Escofier, B., & Pagès, J. (1998). Analyses factorielles simples et multiples. *Dunod, Paris*.
- Escoufier, Y. (1987). The duality diagram: a means for better practical applications. In *Developments in Numerical Ecology* (pp. 139–156). Springer.
- Escoufier, Yves. (1973). Le Traitement des Variables Vectorielles. *Biometrics*, *29*(4), 751–760.
- Feki-Sahnoun, W., Hamza, A., Béjaoui, B., Mahfoudi, M., Rebai, A., & Bel Hassen, M. (2018). Multi-table approach to assess the biogeography of phytoplankton: ecological and management implications. *Hydrobiologia*, *815*(1), 229–251. <https://bit.ly/36IrWB7>
- Fowler, N. (1990). The 10 most common statistical errors. *Bulletin of the Ecological Society of America*, *71*(3), 161–164.

- Franquet, E., Dolédec, S., & Chessel, D. (1995). Using multivariate analyses for separating spatial and temporal effects within species-environment relationships. *Hydrobiologia*, 300–301(1), 425–431. <https://bit.ly/3irmicr>
- Gabriel, K. R. (1971). The biplot graphic display of matrices with applications to principal component analysis. *Biometrika*, 58(3), 453–467. <https://bit.ly/2UN8KQ0>
- Gaertner, J. C., Chessel, D., & Bertrand, J. (1998). Stability of spatial structures of demersal assemblages: a multitable approach. *Aquatic Living Resources*, 11(2), 75–85.
- Galindo-Villardón, Ma. Purificación. (1986). Una Alternativa de Representación Simultánea: HJ-Biplot. *Questiio*, 10, 13–23.
- Galindo-Villardón, Maria Purificación, & Cuadras, C. M. (1986). Una extensión del método Biplot y su relación con otras técnicas. *Publicaciones de Bioestadística y Biomatemática*, 17.
- Gonçalves, A. M. M., Pardal, M. A., Marques, S. C., Mendes, S., Fernández-Gómez, M. J., Galindo-Villardón, M. P., & Azeiteiro, U. M. (2012a). Diel vertical behavior of copepoda community (naupliar, copepodites and adults) at the boundary of a temperate estuary and coastal waters. *Estuarine, Coastal and Shelf Science*, 98, 16–30. <https://bit.ly/3wHp81Z>
- Gonçalves, A. M. M., Pardal, M. A., Marques, S. C., Mendes, S., Fernández-Gómez, M. J., Galindo-Villardón, M. P., & Azeiteiro, U. M. (2012b). Responses of copepoda life-history stages to climatic variability in a southern-European temperate estuary. *Zoological Studies*, 51(3), 321–335.
- Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3), 325–338.
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27(4), 857–871. <https://bit.ly/3xFKWMH>
- Greenacre, M. (2010). Correspondence analysis of raw data. *Ecology*, 91(4), 958–963.
- Gualancañay, E., Tapia, M. E., & Naranjo, C. (2004). Composición y variación estacional del fitoplancton, zooplancton y microbentos en el estuario interior

- del Golfo de Guayaquil, durante el año 2003. *Acta Oceanográfica Del Pacífico*, 12(1), 103–128.
- Gudmundsson, L., Tallaksen, L. M., & Stahl, K. (2011). Spatial cross-correlation patterns of European low, mean and high flows. *Hydrological Processes*, 25(7), 1034–1045. <https://bit.ly/3kfMZmN>
- Gurevitch, J., & Chester, S. T. (1986). Analysis of repeated measures experiments. *Ecology*, 67(1), 251–255.
- Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16(10), 1–84. <https://bit.ly/3iffRsR>
- Haury, L. R., McGowan, J. A., & Wiebe, P. H. (1978). Patterns and Processes in the Time-Space Scales of Plankton Distributions. In J. H. Steele (Ed.), *Spatial Pattern in Plankton Communities* (pp. 277–327). Springer US. <https://bit.ly/3kfJ3T7>
- Hill, M. O. (1973). Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology*, 61(1), 237–249. <https://bit.ly/3r8luNk>
- Hotelling, H. (1933). Analysis of complex statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441.
- Hotelling, H. (1936). Simplified calculation of principal components. *Psychometrika*, 1(1), 27–35. <https://bit.ly/2UNtpn0>
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54(2), 187–211.
- Hutchinson, G. E. (1965). The niche: an abstractly inhabited hypervolume. *The Ecological Theatre and the Evolutionary Play*, 26–78.
- Hutchinson, G. E. (1957). The multivariate niche. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415–421.
- INOCAR. (2012). *Capítulo I: Información general de la República del Ecuador Inocar 2012*. Instituto Oceanográfico de La Armada. <https://bit.ly/36B1SIj>
- Jaffrenou, P. A. (1978). *Sur l'analyse des familles finies de variables vectorielles: bases algébriques et application à la description statistique. Thèse de 3e cycle* [Université Lyon 1, France]. <https://bit.ly/3wAs40l>

- Jiménez, R. (2014). Diatomeas y silicoflagelados del fitoplancton del Golfo de Guayaquil III Edición. *Acta Oceanográfica Del Pacífico*, 19(1), 1–89.
- Jiménez, R., & Bonilla, D. (1980). Composición y distribución de la biomasa del plancton en el Frente Ecuatorial. *Acta Oceanográfica Del Pacífico*, 1, 19–64.
- Kendal, E., Karaman, M., Tekdal, S., & Doğukan, S. (2019). Analysis of promising barley (*Hordeum vulgare* L.) lines performance by AMMI and GGE biplot in multiple traits and environment. *Applied Ecology and Environmental Research*, 17(2), 5219–5233.
- Koutsidi, M., Moukas, C., & Tzanatos, E. (2020). Trait-based life strategies, ecological niches, and niche overlap in the nekton of the data-poor Mediterranean Sea. *Ecology and Evolution*, 10(14), 7129–7144.
- Kriegsman, M. A. (2018). *Discriminant DiSTATIS: A Multi-way discriminant analysis for distance matrices, illustrations with the sorting task*. Ph.D. Thesis. University of Texas at Dallas, EEUU.
- Kroonenberg, P. M. (2008). Applied Multiway Data Analysis. In Wiley-Interscience (Ed.), *A John Wiley & Sons, INC.*, (Issue 9).
<https://bit.ly/36ApNHN>
- L'Hermier Des Plantes, H. (1976). *Structuration des tableaux a trois indices de la statistique: Théorie et application d'une methode d'analyse conjointe*. Thèse de 3e cycle. [Université des Sciences et Techniques du Languedoc].
<https://bit.ly/3hKQoZj>
- Lavit, C. (1988). *Analyse conjointe de tableaux quantitatifs* (Dunod (ed.)). Masson. <https://bit.ly/3re1DfU>
- Lavit, C., Escoufier, Y., Sabatier, R., & Traissac, P. (1994). The ACT (STATIS method). *Computational Statistics and Data Analysis*, 18(1), 97–119.
<https://bit.ly/3B4HoFN>
- Legendre, L., & Legendre, P. (1979). Le traitement multiple des données écologiques. *Écologie Numérique. Tom 1.*, 66(5), 775–776.
<https://bit.ly/3wCSdLP>
- Legendre, P., & Legendre, L. (1998). Numerical ecology. In *Ecology* (2nd ed.).

- Elsevier. <https://bit.ly/3B7QySa>
- Lekve, K., Stenseth, N. C., Gjøsæter, J., & Dolédec, S. (2002). Species richness and environmental conditions of fish along the Norwegian Skagerrak coast. *ICES Journal of Marine Science*, 59(4), 757–769.
- Li, Y., Fang, J., Tian, L. Z., Wang, F., & Chen, Y. S. (2020). DIATOM ASSEMBLAGES IN SURFACE SEDIMENTS AND THEIR REFLECTION ON THE WATER MIGRATION PATHWAY FROM THE INSHORE WATERS OF WESTERN BOHAI BAY, CHINA. *APPLIED ECOLOGY AND ENVIRONMENTAL RESEARCH*, 18(2), 2179–2196.
- Licandro, P., & Ibanez, F. (2000). Erratum: Changes of zooplankton communities in the Gulf of Tigullio (Ligurian Sea, Western Mediterranean) from 1985 to 1995. Influence of hydroclimatic factors. (Journal of Plankton Research (2000) 22 (2225-2253)). *Journal of Plankton Research*, 22(12), 2225–2253. <https://bit.ly/2UcpyQv>
- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *Food Quality and Preference*, 72(September 2018), 31–39. <https://bit.ly/3yWZZ4P>
- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2020). Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTATIS methods. Application to sensometrics. *Food Quality and Preference*, 79(May 2018), 1–9. <https://bit.ly/3i4mZYQ>
- Lobry, J., Lepage, M., & Rochard, E. (2006). From seasonal patterns to a reference situation in an estuarine environment: Example of the small fish and shrimp fauna of the Gironde estuary (SW France). *Estuarine, Coastal and Shelf Science*, 70(1–2), 239–250. <https://bit.ly/2UL2Tej>
- Magnusson, W. E. (1997). Teaching experimental design in ecology, or how to do statistics without a bikini. *Bulletin of the Ecological Society of America*, 78(3), 205–209.
- Marcondes Filho, D., Fogliatto, F. S., & de Oliveira, L. P. L. (2011). Multivariate control charts for monitoring non-linear batch processes. *Producao*, 21(1),

- 132–148. <https://bit.ly/3yPSYmm>
- Márquez, J. L., Díaz de Pascual, A., & Defives, G. (1992). Aplicación del método Statist: Factores fisico-químicos del agua del embalse Uribante. *Economía*, 17(7), 35–58.
- Martínez, G., Contreras, M., & Vila, I. (2003). Modelos conceptuales de abundancia de fitoplancton asociados a la heterogeneidad espacial en el Embalse Rapel (Chile central). *Revista Chilena de Historia Natural*, 76(2), 255–266. <https://bit.ly/36Bs8SN>
- Mendes, S., Fernández-Gómez, M. J., Galindo-Villardón, M. P., Morgado, F., Maranhão, P., Azeiteiro, U. M., & Bacelar-Nicolau, P. (2009). The study of bacterioplankton dynamics in the Berlengas Archipelago (West coast of Portugal) by applying the HJ-biplot method. *ARQUIPÉLAGO-Life and Marine Sciences*, 26(January), 25–35.
- Mendes, S., Fernández-Gómez, M. J., Pereira, M. J., Azeiteiro, U. M., & Galindo-Villardón, M. P. (2012). An empirical comparison of Canonical Correspondence Analysis and STATICO in the identification of spatio-temporal ecological relationships. *Journal of Applied Statistics*, 39, 979–994. <https://bit.ly/3wHNJU6>
- Mendes, S., Fernández-Gómez, M. J., Resende, P., Jorge Pereira, M., Galindo-Villardón, M. P., & Azeiteiro, U. M. (2009). Spatio-temporal structure of diatom assemblages in a temperate estuary. A STATICO analysis. *Estuarine, Coastal and Shelf Science*, 84(4), 637–644. <https://bit.ly/3wFjFbO>
- Mendes, S., Marques, S. C., Azeiteiro, U. M., Fernández-Gómez, M. J., Galindo-Villardón, M. P., Maranhão, P., Morgado, F., & Leandro, S. M. (2011). Zooplankton Distribution in a Marine Protected Area: the Berlengas Natural Reserve (Western Coast of Portugal). *Fresenius Environmental Bulletin*, 20(2), 496–505. <https://bit.ly/3B4HZY3>
- Mérigot, B., Gaertner, J. C., Brind'amour, A., Carbonara, P., Esteban, A., Garcia-Ruiz, C., Gristina, M. C., Imzilen, T., Jadaud, A., Joksimovic, A., Kavadas, S., Kolutari, J., Maina, I., Maiorano, P., Manfredi, C., Micallef, R.,

- Peristeraki, P., Relini, G., Sbrana, M., ... Vrgoc, N. (2019). Stability of the relationships among demersal fish assemblages and environmental-trawling drivers at large spatio-temporal scales in the northern mediterranean sea. *Scientia Marina*, 83(S1), 153–163. <https://bit.ly/3wPVpEd>
- Noriega, C. E., Muniz, K., Flores-Montes, M. J., Macêdo, S. J., Araujo, M., Feitosa, F. A., & Lacerda, S. R. (2009). Series temporales de variables hidrobiológicas en un estuario tropical (Brasil). *Revista de Biología Marina y Oceanografía*, 44(1), 93–108. <https://bit.ly/3icqHiZ>
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11), 559–572. <https://bit.ly/3egJed6>
- Perugachi-Salamea, C., González-Narváez, M., Pambabay-Calero, J., García-Arévalo, I., Vargas-Ayala, F., Shigla-Cuji, G., & Nath-Nieto, J. (2014). Caracterización oceanográfica del perfil costero ecuatoriano a través de métodos estadísticos aplicados a las estaciones fijas de monitoreo del INOCAR, serie a partir de 1981. *Acta Oceanográfica Del Pacífico*, 19(2), 7–21.
- Picazo, F., Millan, A., & Doledec, S. (2012). Are patterns in the taxonomic, biological and ecological traits of water beetles congruent in Mediterranean ecosystems? *Freshwater Biology*, 57(11), 2192–2210.
- Pierce, S., Negreiros, D., Cerabolini, B. E. L., Kattge, J., Diaz, S., Kleyer, M., Shipley, B., Wright, S. J., Soudzilovskaia, N. A., Onipchenko, V. G., & others. (2017). A global method for calculating plant CSR ecological strategies applied across biomes world-wide. *Functional Ecology*, 31(2), 444–457.
- R Core Team and contributors worldwide. (2018). *The R stats package* (3.5.2).
- Rao, C. R. (1964). The use and interpretation of Principal Component Analysis in applied research stable. *The Indian Journal of Statistics*, 26(4), 329–358. <https://bit.ly/3B45HE0>
- Ravindra, K., Sokhi, R., & Van Grieken, R. (2008). Atmospheric polycyclic aromatic hydrocarbons: source attribution, emission factors and regulation. *Atmospheric Environment*, 42(13), 2895–2921.

- Robert, P., & Escoufier, Y. (1976). A Unifying Tool for Linear Multivariate Statistical Methods : The RV- Coefficient. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 25(3), 257–265. <https://bit.ly/3hEOf0S>
- Rodríguez-Martínez, C. C. (2020). *Contribuciones a los métodos STASIS basados en técnicas de aprendizaje no supervisado* [Universidad de Salamanca]. <https://bit.ly/36CM8Vi>
- Rodríguez, O., & Corrales, D. (2014). Interstatis: the STATIS method for interval valued data. *Revista de Matematica: Teoria y Aplicaciones*, 21(1), 73–83.
- Sabatier, R., & Vivien, M. (2008). A new linear method for analyzing four-way multiblocks tables: STATIS-4. *Journal of Chemometrics*, 22(6), 399–407. <https://bit.ly/3wA3RqU>
- Sabatier, R., Vivien, M., Reynès, C., Myrtille, V., & Christelle, R. (2013). Une nouvelle proposition, l'Analyse Discriminante Multitableaux : STATIS-LDA. *Journal de La Société Française de Statistique & Revue de Statistique Appliquée*, 154(3), 31–43.
- Santos, J. L. (2006). The impact of El Niño - Southern Oscillation Events on South America. *Advances in Geosciences*, 6, 221–225. <https://bit.ly/3hCYJOj>
- Sautron, V., Chavent, M., Viguier, N., & Villa-Vialaneix, N. (2016). Multiway-SIR for longitudinal multi-table data integration. *22nd International Conference on Computational Statistics (COMPSTAT)*, 1.
- Sauzay, L., Hanafi, M., Qannari, E. M., & Schlich, P. (2006). *Analyse de K+ 1 tableaux à l'aide de la méthode STATIS: application en évaluation sensorielle, 9ieme Journées Européennes Agro-industrie et Méthodes Statistiques* (Société Française de Statistique (SFdS) (ed.); pp. 157–160). Montpellier (France).
- Scott, E. R., & Crone, E. E. (2021). Using the right tool for the job: the difference between unsupervised and supervised analyses of multivariate ecological data. *Oecologia*, 1–13. <https://bit.ly/3hEZSoP>
- Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using

- multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, 22(4), 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(1), 31–46.
- Slimani, N., Guilbert, E., Ayni, F. El, Jrad, A., Boumaiza, M., & Thioulouse, J. (2017). The use of STATICO and COSTATIS, two exploratory three-ways analysis methods: an application to the ecology of aquatic heteroptera in the Medjerda watershed (Tunisia). *Environmental and Ecological Statistics*, 24(2), 269–295. <https://bit.ly/36ztjCn>
- Slimani, N., Jiménez, J. J., Guilbert, E., Boumaiza, M., & Thioulouse, J. (2020). Surface water quality assessment in a semiarid Mediterranean region (Medjerda, Northern Tunisia) using partial triadic analysis. *Environmental Science and Pollution Research International*, 27(24), 30190–30198.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101. <https://bit.ly/3ieAHs4>
- Stanimirova, I., Walczak, B., Massart, D. L., Simeonov, V., Saby, C. A., & Di Crescenzo, E. (2004). STATIS, a three-way method for data analysis. Application to environmental data. *Chemometrics and Intelligent Laboratory Systems*, 73(2), 219–233. <https://bit.ly/3ra9Wcs>
- Steel, A. E., Kennedy, M. C., Cunningham, P. G., & Stanovick, J. S. (2013). Applied statistics in ecology: common pitfalls and simple solutions. *Ecosphere*, 4(9), 1–13. <https://bit.ly/2T91XQg>
- Tapia, M. E., & Naranjo, C. (2004). Diversidad del plancton en el estuario interior de Puerto Bolívar, Ecuador durante julio 2002. *Acta Oceanográfica Del Pacífico*, 12(1), 29–49.
- Taybi, A. F., Mabrouki, Y., Dakki, M., Berrahou, A., & Millán, A. (2020). Longitudinal distribution of macroinvertebrate in a very wet North African Basin: Oued Melloulou (Morocco). *Annales de Limnologie-International Journal of Limnology*, 56, 17.

- Tenenhaus, M., & Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, *50*(1), 91–119. <https://bit.ly/3eknNHY>
- ter Braak, C. J. F. (1985). Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Response Model. *Biometrics*, *41*(4), 859. <https://bit.ly/3i88Tpn>
- ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, *67*(1), 1167–1179. <https://bit.ly/3hEMDnZ>
- ter Braak, C. J. F. (1987). The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, *69*(3), 69–77. <https://bit.ly/3B45Kzy>
- 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. <https://bit.ly/3B3j7Qv>
- Thioulouse, J., & Chessel, D. (1987). Les analyses multi-tableaux en écologie factorielle. I: 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*(1), 272–283.
- Thurstone, L. L. (1947). Multiple factor analysis. In *Multiple factor analysis*. University of Chicago Press.
- Torres, G. (2005). Importancia ecológica del fitoplancton durante El Niño 1991-1993, en el Pacífico Ecuatorial (Ecuador). *Acta Oceanográfica Del Pacífico*, *13*(1), 35–49.
- Torres, G., Calderón, T., Mero, M., & Franco, V. (2004). Procesos planctónicos en el Golfo de Guayaquil (Campo Amistad) Julio-Agosto/2001. *Acta Oceanográfica Del Pacífico*, *12*(1), 81–91. <https://bit.ly/2VIP8ND>
- Torres, G., Mero, M., Calderón, T., Franco, V., & Salazar, E. (2004). Relación fitoplancton-zooplancton en el Pacífico Ecuatorial (Ecuador), durante

- Septiembre 2001. *Acta Oceanográfica Del Pacífico*, 12(1), 51–61.
- Torres, G., Recalde, S., Narea, R., Renteria, W., & Troccoli, L. (2017). Variabilidad espacio-temporal del fitoplancton y variables oceanográficas en el Golfo de Guayaquil durante el 2013-15. *Revista Del Instituto de Investigación FIGMMG-UNMSM*, 20(40), 70–79.
- Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3), 279–311. <https://bit.ly/3kfPkhz>
- 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. *Computational Statistics and Data Analysis*, 51(9), 4193–4205. <https://bit.ly/3zcxYXh>
- Vega, M., Pardo, R., Barrado, E., & Debán, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, 32(12), 3581–3592.
- Vicente-Villardón, J. L. (2016). *MULTBILOT: A package for multivariate analysis using Biplots*. (18.1101). Departamento de Estadística. Universidad de Salamanca. <https://bit.ly/3r8mOzM>
- Vivien, M., & Sabatier, R. (2004). A generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables. *Computational Statistics and Data Analysis*, 46(1), 155–171. <https://bit.ly/3xLsiD6>
- Yan, W., Hunt, L. A., Sheng, Q., & Szlavnic, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science*, 40(3), 597–605.
- Yi, Y., Yang, Z., & Zhang, S. (2011). Ecological risk assessment of heavy metals in sediment and human health risk assessment of heavy metals in fishes in the middle and lower reaches of the Yangtze River basin. *Environmental Pollution*, 159(10), 2575–2585.
- Zalocar De Domitrovic, Y. (2002). Structure and variation of the Paraguay River phytoplankton in two periods of its hydrological cycle. *Hydrobiologia*, 472, 177–196. <https://bit.ly/3i9ZopN>

ANEXOS

*No te abandonaré,
hasta cumplir con todo
lo que te he
prometido*
Génesis 28:15

ANEXOS

A1 Resumen del manuscrito en el que se aplica el método STATICO.

Applied multivariate statistical analysis (STATICO method) for marine phytoplankton species assemblages off Ecuador: Geographic patterns and temporal variability

Mariela González-Narváez^{1,2¶*}, María José Fernández-Gómez^{2,3,4&}, Susana Mendes^{5¶}, Jorge I. Sonnenholzner^{6¶}, Sonia Recalde⁷, Purificación Galindo-Villardón^{2,4}

Abstract

The coastal marine environment of Ecuador is characterized by two biogeographic regions that form the basis of its fishery productivity. However, its biological richness is limited by the occurrence of natural processes such as El Niño/Southern Oscillation (ENSO) events, which cause variations in sea surface temperature. To understand biological and environmental variability in the spatio-temporal scale, we used descriptive analysis, classification methods and permutation multivariate analysis of variance (PERMANOVA). We used the STATICO method to characterize the stability of the relationship between phytoplanktonic richness and environmental variations. We were able to classify two coastal profiles with different species-environment relationship patterns. We located the strongest relationship in the upper 20 m of the ocean. This relationship was found mainly in the southern sector of the Gulf of Guayaquil (an estuarine system) but was weakened in warmer conditions. The diatom *Nitzschia longissima* was characteristic of the entire region, and it maintained its outstanding abundance during normal and warm conditions. The presence of warmer conditions in 2015 influenced the increase in *Nitzschia* sp and *Thalassiosira* sp to their highest abundances during the study period, consequently decreasing the abundance of *Stauroneis membranous*. During this warm event, a stronger relationship was

demonstrated in estuarine zones during the rainy season and oceanic zones in the dry season in the northern section, while the southern section had occurrences during both seasons. Warm conditions generated less variability in environmental parameter concentrations. Our study demonstrated the importance of using multivariate three-way statistical techniques to study ecosystem spatio-temporal patterns.

A2 Resumen del manuscrito en el que se aplican los métodos STATICO y HJ-Biplot

Environment impact on pacific phytoplankton in Niño 1+2 region (Ecuador)

Mariela González-Narváez^{1,2}, María José Fernández-Gómez^{2,4}, Susana Mendes³,
Omar Ruiz-Barzola^{1,2}, Julia Nieto¹, Purificación Galindo-Villardón^{2,4}

Abstract

The aim of this study was to determine, through the application of multivariate methods, the effect caused by environmental variations on phytoplankton assemblages in the eastern equatorial Pacific region during different oceanographic conditions (normal, warm and extreme warm). Four sampling stations located 10 miles offshore of continental Ecuador were studied, three of them located in the Niño 1+2 region, an area known for its sensitivity to detect variations in sea surface temperature. There are measurements per water column (seven standard depths), at two climatic stations, in the period 2013–2015 (normal, warm and extremely warm oceanographic conditions). To achieve the objective, a statistical multivariate analysis was applied in order to identify and compare the spatial stability of the environment-species relationships that occurred in each year. The STATICO method allowed us to achieve this objective, and in this investigation, we introduce a modification in this method, instead of using a Principal Component Analysis (PCA) to investigate the stable part, named compromise, we use the HJ-Biplot, to provide the best quality of representation for both the biotic and abiotic factors. The analysis was complemented with the heatmap obtained from a compromise matrix. The study showed how the phytoplankton species respond to the environmental variations that were evident in the region as result of the oceanographic conditions that occurred in the eastern equatorial Pacific. The species abundance of the genera *Proboscia*, *Dactyliosolen* (*Dactant*) and *Hemiaulus* (*centricae* diatoms) correlated positively with

temperature and oxygen, whereas, *Thalassionema* and *Navicula* (diatoms pennate) correlated positively with nutrients and salinity during warm conditions. This behaviour pattern is dynamic and changes according to the oceanographic conditions registered in the region. The species with low quality of representation were those that showed lower average abundance and higher variability with respect to the other species.

A3 Datos del Conjunto de Matrices ambientales: HJ-STATICO

T	S	OD	NO ₃	NO ₂	PO ₄	SiO ₄
24.16	33.86	4.57	1.41	0.03	0.34	4.85
23.54	33.89	4.06	7.57	0.13	0.53	6.43
26.41	34.29	4.10	0.75	0.12	0.51	3.46
27.51	34.50	4.44	0.37	0.12	0.06	2.92
27.96	34.18	4.56	0.47	0.06	0.44	3.98
28.08	33.46	4.46	0.04	0.02	0.12	1.89
27.78	33.17	3.92	0.25	0.02	0.36	2.76
27.21	32.84	4.59	0.10	0.02	0.12	0.97
27.33	32.76	4.78	0.10	0.02	0.22	0.47
27.62	32.18	4.59	0.16	0.04	0.27	1.63
27.78	32.06	4.49	0.06	0.02	0.10	1.97
21.27	34.72	4.07	9.36	0.40	0.74	10.76
22.61	34.76	4.46	4.83	0.18	0.43	3.59
23.94	34.68	4.11	3.04	0.11	0.44	6.50
25.98	34.76	4.72	0.72	0.17	0.34	2.76
25.05	34.87	4.13	6.74	0.66	0.88	4.95
25.66	34.77	4.37	0.54	0.11	0.32	3.23
23.37	34.97	4.12	4.51	0.41	0.92	3.09
24.28	34.16	4.76	0.22	0.04	0.31	1.13
24.00	34.41	4.76	2.47	0.23	0.25	2.49
24.68	34.06	4.82	0.07	0.06	0.24	2.66
26.17	33.27	4.65	0.10	0.02	0.08	1.40
21.92	34.23	4.37	5.60	0.11	0.35	13.45
24.03	33.97	4.27	6.58	0.29	1.04	11.00
24.91	34.56	3.63	6.80	0.20	0.88	8.32
25.73	34.86	4.42	2.57	0.14	0.40	2.33
26.86	34.85	4.20	5.00	0.16	0.70	5.45
27.17	34.53	4.48	0.21	0.02	0.18	2.05
24.81	34.62	3.54	0.65	0.02	0.80	5.98
25.19	33.93	4.56	0.96	0.04	0.27	1.63
26.05	33.53	4.69	0.36	0.03	0.05	1.21
25.73	33.57	4.70	0.05	0.07	0.39	2.46
26.59	33.04	4.59	0.01	0.04	0.30	1.07
21.45	34.77	3.88	11.57	0.47	1.37	11.66
23.78	34.35	3.77	8.64	0.28	1.07	10.51
24.76	33.61	4.20	4.21	0.13	0.61	8.46
26.13	33.57	4.10	4.40	0.12	0.37	9.01
25.48	34.17	4.10	6.21	0.12	0.59	8.43
25.64	34.84	4.67	0.36	0.04	0.25	2.42
23.14	34.88	3.69	5.28	0.14	1.05	5.22
23.89	34.33	4.21	5.80	0.28	0.56	5.66
24.65	34.18	4.51	0.63	0.58	0.58	2.46
24.37	34.35	4.80	0.53	0.04	0.34	1.73
26.00	33.83	4.67	0.17	0.03	0.15	1.30

A4 Datos del Conjunto de Matrices de Especies: HJ-STATICO

e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13	e14	e15	e16	e17	e18	e19	e20	e21	e22	e23
3.25	0.00	2.57	4.56	4.91	2.45	4.88	5.68	4.95	4.30	1.28	1.44	1.57	4.09	5.09	3.43	1.53	2.81	0.00	2.61	2.69	1.28	0.00
3.33	3.15	0.00	3.03	3.01	1.48	4.40	3.67	3.01	1.70	4.05	3.11	4.37	2.51	4.52	4.76	2.65	1.32	0.00	3.31	0.00	0.00	0.00
0.00	0.00	0.00	0.00	2.61	0.00	5.88	0.00	3.03	0.00	0.00	0.00	0.00	4.17	4.52	4.98	0.00	0.00	0.00	0.00	3.83	2.85	4.38
3.61	5.76	1.28	3.11	5.16	5.44	4.57	5.52	2.69	1.53	5.07	1.41	5.19	4.15	4.34	4.99	1.28	1.22	0.00	4.59	1.32	3.85	1.28
2.65	0.00	0.00	0.00	2.75	1.22	4.23	4.51	4.45	1.28	1.41	0.00	1.36	4.11	4.52	5.16	1.28	0.00	0.00	0.00	4.01	1.12	1.41
4.06	2.94	0.00	2.61	2.51	1.12	3.76	2.67	1.22	2.25	2.35	4.53	2.35	4.51	2.25	2.48	0.00	0.00	0.00	0.00	4.35	3.77	1.12
1.22	0.00	0.00	0.00	0.00	0.00	1.22	2.51	2.35	2.25	1.12	0.00	0.00	4.19	2.25	0.00	1.12	0.00	0.00	0.00	4.22	4.04	3.53
4.09	2.84	0.00	2.51	1.32	0.00	0.00	3.86	3.79	3.37	2.51	0.00	0.00	4.17	2.35	1.12	0.00	0.00	0.00	0.00	4.56	4.12	3.79
4.54	4.90	0.00	4.52	4.71	0.00	4.87	4.76	0.00	2.35	4.05	4.60	2.35	4.37	1.12	3.99	0.00	1.22	0.00	0.00	4.19	3.97	3.57
2.61	0.00	0.00	2.35	3.83	0.00	3.91	3.77	0.00	2.25	1.12	2.71	0.00	3.91	3.53	2.64	0.00	0.00	0.00	1.22	4.07	4.09	2.45
3.91	2.61	0.00	2.73	4.21	0.00	4.21	4.09	2.35	2.25	1.12	1.46	1.12	2.35	2.25	3.93	0.00	0.00	0.00	1.22	3.91	4.22	3.72
2.73	0.00	0.00	0.00	1.58	0.00	0.00	4.63	1.43	1.74	2.80	0.00	2.96	3.83	2.48	3.77	1.22	4.02	0.00	2.71	2.45	1.28	0.00
4.76	2.92	0.00	2.73	4.88	0.00	3.68	2.75	0.00	3.43	1.28	0.00	1.38	4.31	5.13	4.88	3.44	2.98	0.00	1.47	1.53	1.41	1.48
4.37	1.50	0.00	4.40	5.59	0.00	6.25	5.15	4.70	3.89	4.90	1.81	4.88	4.55	4.34	5.52	2.65	0.00	2.91	0.00	1.22	3.67	4.14
3.34	1.60	0.00	5.26	5.11	0.00	5.82	6.56	4.70	2.82	2.99	0.00	4.37	4.22	5.50	5.96	5.74	4.46	0.00	0.00	1.22	2.45	0.00
3.28	1.54	0.00	3.47	3.59	0.00	1.54	3.34	4.84	1.36	4.18	3.05	3.64	5.02	5.67	5.81	5.67	0.00	0.00	5.15	3.82	1.22	4.11
2.67	2.53	1.06	3.93	4.25	0.62	4.95	5.16	2.96	1.40	3.82	1.67	2.50	4.56	4.29	4.68	1.59	1.25	1.00	1.19	3.93	4.05	3.76
2.78	0.00	1.32	2.94	2.77	0.00	3.08	2.71	3.17	1.28	1.22	0.00	0.00	4.68	0.00	4.78	0.00	0.00	0.00	0.00	4.09	3.95	1.28
0.00	0.00	0.00	0.00	0.00	0.00	3.63	2.45	3.57	1.12	1.22	0.00	2.35	4.15	0.00	3.83	0.00	0.00	1.22	0.00	4.59	4.41	4.11
0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.32	0.00	0.00	0.00	0.00	1.12	3.79	0.00	3.85	0.00	0.00	0.00	0.00	4.04	4.39	4.28
2.55	0.00	0.00	1.28	0.00	0.00	1.12	2.45	1.12	0.00	0.00	0.00	0.00	3.85	0.00	3.85	0.00	0.00	0.00	0.00	4.17	4.40	4.10
0.00	0.00	0.00	0.00	1.32	0.00	2.48	0.00	0.00	0.00	0.00	0.00	0.00	3.79	3.89	4.17	0.00	0.00	0.00	0.00	3.96	2.55	2.35
4.88	2.98	1.53	3.49	3.46	0.00	3.26	3.91	3.44	1.32	2.87	1.58	2.92	4.87	4.44	5.34	3.05	3.57	2.91	2.81	1.44	3.94	0.00
3.26	3.15	0.00	1.70	2.61	0.00	5.13	0.00	4.76	4.15	4.70	0.00	5.15	4.19	4.74	4.92	3.46	2.73	1.38	0.00	1.41	1.22	0.00
1.58	0.00	0.00	0.00	3.38	0.00	3.36	1.63	4.84	1.41	0.00	0.00	1.38	4.12	3.99	4.87	1.32	2.48	0.00	0.00	2.73	4.17	4.33
2.98	0.00	0.00	3.31	5.30	0.00	6.41	5.75	5.89	3.04	4.53	0.00	4.40	3.83	2.97	4.89	2.85	3.85	1.57	0.00	2.57	2.69	1.22
1.28	1.41	0.00	2.61	3.03	0.00	4.88	5.56	3.16	0.00	5.05	0.00	1.32	4.43	4.38	3.24	1.43	0.00	1.60	0.00	1.32	2.63	2.96
1.32	0.00	0.00	0.00	1.22	0.00	2.55	3.63	2.64	3.67	1.12	0.00	0.00	2.73	2.35	2.45	0.00	0.00	0.00	0.00	3.99	3.81	3.57
0.00	0.00	0.00	3.67	2.45	1.12	3.85	4.07	2.63	0.00	1.12	4.35	3.47	4.39	4.22	4.02	0.00	0.00	1.22	0.00	4.47	4.42	4.14
1.28	3.27	0.00	1.22	2.51	0.00	4.09	2.91	1.12	2.35	1.28	1.52	1.12	4.45	2.41	3.85	0.00	0.00	0.00	0.00	2.51	3.89	2.51
2.65	0.00	0.00	0.00	4.09	0.00	3.73	3.79	2.45	1.12	2.41	0.00	1.12	4.17	2.45	4.32	0.00	0.00	0.00	0.00	4.22	3.77	3.57
0.00	1.43	0.00	0.00	0.00	0.00	1.22	1.22	0.00	1.12	0.00	0.00	0.00	3.83	3.67	3.76	0.00	0.00	0.00	0.00	3.91	4.20	4.03
0.00	0.00	0.00	1.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.63	1.12	2.51	0.00	0.00	0.00	0.00	4.11	4.24	3.37
3.04	0.00	0.00	2.64	5.00	1.56	4.40	3.30	3.51	5.00	3.04	1.92	3.63	4.57	3.00	2.88	3.07	4.51	0.00	1.41	1.44	2.51	0.00
1.32	2.96	1.12	2.61	4.64	0.00	3.06	2.51	3.03	1.41	0.00	0.00	1.28	4.27	3.01	4.62	0.00	0.00	4.92	0.00	2.55	0.00	3.12
4.58	3.78	3.57	3.19	3.33	3.09	2.98	1.82	1.77	1.32	2.77	2.00	1.69	4.69	4.27	4.30	3.46	3.95	2.71	3.52	2.58	1.28	0.00
4.63	5.29	1.28	1.62	4.48	4.94	4.40	4.81	4.39	1.52	0.00	6.08	2.81	3.09	4.61	4.58	3.18	4.14	0.00	1.54	2.45	1.28	0.00
4.57	4.59	0.00	3.04	2.78	3.20	4.38	2.83	3.18	2.78	0.00	5.66	4.71	4.47	4.47	4.67	2.99	1.43	0.00	4.67	3.79	1.28	4.09
0.00	0.00	0.00	0.00	0.00	0.00	2.85	1.12	0.00	0.00	2.69	0.00	2.35	4.69	2.57	1.12	0.00	0.00	0.00	0.00	4.11	3.86	3.71
2.71	1.74	0.00	2.25	2.82	0.00	2.51	3.79	1.22	3.37	2.35	0.00	0.00	4.63	2.25	2.35	0.00	0.00	0.00	0.00	4.19	2.45	1.32
1.36	1.63	0.00	0.00	2.86	1.12	2.45	0.00	1.12	0.00	1.46	2.48	4.27	1.12	1.22	0.00	0.00	1.46	0.00	0.00	4.14	3.86	4.73
0.00	3.25	0.00	2.63	3.38	0.00	3.14	3.27	2.45	2.35	2.61	2.96	2.35	4.46	3.63	3.71	0.00	0.00	0.00	0.00	4.27	4.25	4.34
1.22	0.00	0.00	1.12	1.38	0.00	3.77	2.45	1.12	0.00	0.00	0.00	1.12	4.75	1.12	2.51	0.00	0.00	0.00	0.00	4.16	3.81	3.47
0.00	0.00	0.00	0.00	1.28	0.00	2.64	3.67	2.25	0.00	0.00	0.00	0.00	1.22	3.47	3.77	0.00	0.00	0.00	0.00	4.29	3.95	4.09

A5 Datos de las Etiquetas para el Conjunto de Matrices: HJ-STATICO

ID	space	month
Es_Feb	Esmeraldas	Feb
Es_Mar	Esmeraldas	Mar
Es_Abr	Esmeraldas	Abr
Es_May	Esmeraldas	May
Es_Jun	Esmeraldas	Jun
Es_Jul	Esmeraldas	Jul
Es_Ago	Esmeraldas	Ago
Es_Sep	Esmeraldas	Sep
Es_Oct	Esmeraldas	Oct
Es_Nov	Esmeraldas	Nov
Es_Dic	Esmeraldas	Dic
Li_Feb	La Libertad	Feb
Li_Mar	La Libertad	Mar
Li_Abr	La Libertad	Abr
Li_May	La Libertad	May
Li_Jun	La Libertad	Jun
Li_Jul	La Libertad	Jul
Li_Ago	La Libertad	Ago
Li_Sep	La Libertad	Sep
Li_Oct	La Libertad	Oct
Li_Nov	La Libertad	Nov
Li_Dic	La Libertad	Dic
Ma_Feb	Manta	Feb
Ma_Mar	Manta	Mar
Ma_Abr	Manta	Abr
Ma_May	Manta	May
Ma_Jun	Manta	Jun
Ma_Jul	Manta	Jul
Ma_Ago	Manta	Ago
Ma_Sep	Manta	Sep
Ma_Oct	Manta	Oct
Ma_Nov	Manta	Nov
Ma_Dic	Manta	Dic
Pb_Feb	Pto. Bolivar	Feb
Pb_Mar	Pto. Bolivar	Mar
Pb_Abr	Pto. Bolivar	Abr
Pb_May	Pto. Bolivar	May
Pb_Jun	Pto. Bolivar	Jun
Pb_Jul	Pto. Bolivar	Jul
Pb_Ago	Pto. Bolivar	Ago
Pb_Sep	Pto. Bolivar	Sep
Pb_Oct	Pto. Bolivar	Oct
Pb_Nov	Pto. Bolivar	Nov
Pb_Dic	Pto. Bolivar	Dic

A6 Valores Promedio para datos ambientales y especies según lugares de muestreo HJ-STATICO.

	T	S	OD	NO ₃	NO ₂	PO ₄	SiO ₄
Esmeraldas	26.85	33.38	4.42	1.02	0.06	0.28	2.85
Manta	25.36	34.15	4.31	2.62	0.10	0.49	5.00
La Libertad	24.27	34.49	4.45	2.96	0.22	0.45	3.87
Pto. Bolívar	24.48	34.26	4.23	4.34	0.20	0.63	6.08

	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13	e14	e15	e16	e17	e18	e19	e20	e21	e22	e23
Esmeraldas	3.02	2.02	0.35	2.31	3.18	1.07	3.81	3.73	2.53	2.14	2.19	1.75	1.67	3.87	3.34	3.41	0.72	0.60	0.00	1.18	3.38	3.03	2.30
Manta	1.75	1.11	0.14	1.57	2.55	0.10	3.50	2.95	2.81	1.65	2.10	0.68	1.90	4.06	3.34	4.02	1.10	1.15	0.79	0.26	2.97	3.54	2.70
La Libertad	2.41	0.92	0.22	2.18	2.64	0.06	2.96	3.32	2.41	1.55	2.04	0.59	2.11	4.25	2.85	4.65	1.85	1.16	0.47	0.96	3.18	3.07	2.69
Pto. Bolívar	2.13	2.11	0.54	1.74	2.91	1.27	3.32	2.69	2.19	1.61	1.22	1.83	2.04	4.10	3.05	3.25	1.15	1.28	0.83	1.01	3.45	2.60	2.62

A7 Indicadores HJ-Biplot para la matriz compromiso

>resultado\$HJB

\$CRTi #presenta la contribución relativa total de los elementos fila (Q)

[1]	0.046	0.017	0.004	0.026	0.061	0.008	0.01	0.012	0.09	0.023
[11]	0.04	0.021	0.091	0.011	0.045	0.023	0.1	0.1	0.03	0.047
[21]	0.06	0.07	0.066							

\$CREiF1 #presenta la contribución relativa del elemento fila (Q) al factor

	[,1]	[,2]
e1	0.046	0.087
e2	0.01	0.077
e3	0.003	0.01
e4	0.026	0.037
e5	0.065	0.034
e6	0.003	0.001
e7	0.004	0.107
e8	0.001	0.14
e9	0.088	0.162
e10	0.025	0.001
e11	0.038	0.062
e12	0.01	0.001
e13	0.101	0.013
e14	0.011	0.007
e15	0.044	0.025
e16	0.024	0.002
e17	0.109	0.002
e18	0.107	0.064
e19	0.03	0.002
e20	0.049	0.035
e21	0.065	0.01
e22	0.076	0.005
e23	0.066	0.118

\$CRFIEi #presenta la contribución relativa del factor al elemento fila (Q)

	[,1]	[,2]
e1	0.877	0.094
e2	0.501	0.22
e3	0.708	0.137
e4	0.896	0.071
e5	0.948	0.028
e6	0.323	0.009
e7	0.235	0.403

e8	0.042	0.59
e9	0.895	0.092
e10	0.958	0.003
e11	0.845	0.077
e12	0.401	0.002
e13	0.983	0.007
e14	0.851	0.031
e15	0.858	0.027
e16	0.914	0.003
e17	0.968	0.001
e18	0.948	0.032
e19	0.889	0.003
e20	0.92	0.037
e21	0.968	0.008
e22	0.962	0.003
e23	0.891	0.089

\$CRTj #presenta la contribución relativa total de los elementos columna (P)

[1]	0.074	0.096	0.106	0.214	0.105	0.101	0.3
-----	-------	-------	-------	-------	-------	-------	-----

\$CREjFl #presenta la contribución relativa del elemento columna (P) al factor

	[,1]	[,2]
T	0.06	0.176
S	0.067	0.518
OD	0.096	0.134
NO3	0.236	0.066
NO2	0.103	0.006
PO4	0.106	0.029
SiO4	0.331	0.071

\$CRFIEj #presenta la contribución relativa del factor al elemento columna (P)

	[,1]	[,2]
T	0.717	0.118
S	0.622	0.269
OD	0.801	0.063
NO3	0.974	0.015
NO2	0.869	0.003
PO4	0.926	0.014
SiO4	0.965	0.012

\$QLREi #presenta la calidad relativa de los elementos fila (Q)

e1	e2	e3	e4	e5	e6	e7	e8	e9	e10
0.971	0.721	0.844	0.968	0.976	0.332	0.638	0.63	0.987	0.96

e11	e12	e13	e14	e15	e16	e17	e18	e19	e20
0.921	0.402	0.99	0.882	0.886	0.918	0.968	0.98	0.892	0.96
e21	e22	e23							
0.976	0.966	0.981							

\$QLREj #presenta la calidad relativa de los elementos columna(P)

T	S	OD	NO3	NO2	PO4	SiO4
0.836	0.891	0.864	0.989	0.872	0.94	0.976

A8 Indicadores HJ-Biplot para las k-matrices en la Intra-estructura

>resultado\$HJBK

\$CRTi #presenta la contribución relativa total de los elementos fila (KQ)

[1]	0.103	0.052	0.005	0.043	0.023	0.02	0.003	0.052	0.039	0.038
[11]	0.009	0.054	0.071	0.006	0.031	0.09	0.09	0.055	0.022	0.013
[21]	0.082	0.078	0.021	0.074	0.012	0.002	0.041	0.052	0.009	0.08
[31]	0.05	0.048	0.023	0.086	0.028	0.038	0.007	0.038	0.031	0.104
[41]	0.029	0.007	0.144	0.015	0.054	0.029	0.058	0.035	0.005	0.033
[51]	0.032	0.001	0.055	0.031	0.129	0.008	0.047	0.019	0.103	0.007
[61]	0.074	0.032	0.078	0.083	0.033	0.016	0.052	0.015	0.056	0.064
[71]	0.057	0.01	0.035	0.132	0.053	0.004	0.004	0.023	0.112	0.006
[81]	0.052	0.024	0.015	0.01	0.014	0.038	0.079	0.045	0.027	0.029
[91]	0.099	0.067								

\$CREiFl #presenta la contribución relativa del elemento fila (KQ) al factor

	[,1]	[,2]
e1	0.119	0.152
e2	0.056	0.016
e3	0.001	0
e4	0.029	0.216
e5	0.019	0.047
e6	0	0.049
e7	0	0.002
e8	0.056	0.093
e9	0.033	0.039
e10	0.04	0.104
e11	0.003	0.003
e12	0.052	0.138
e13	0.065	0.019
e14	0	0.03
e15	0.032	0.028
e16	0.086	0.039
e17	0.113	0.013
e18	0.051	0
e19	0.025	0.004
e20	0.008	0.003
e21	0.101	0.002
e22	0.095	0.003
e23	0.016	0.001
e1	0.105	0.039
e2	0	0.012
e3	0	0.005
e4	0.014	0.188

e5	0.048	0.114
e6	0.002	0
e7	0.055	0.065
e8	0.06	0.064
e9	0.027	0.138
e10	0.028	0.01
e11	0.113	0.082
e12	0.01	0.019
e13	0.051	0.032
e14	0.006	0.011
e15	0.021	0.041
e16	0.013	0.043
e17	0.088	0.119
e18	0.039	0.001
e19	0.001	0.001
e20	0.18	0.014
e21	0.019	0
e22	0.076	0
e23	0.044	0.001
e1	0.043	0.137
e2	0.02	0.159
e3	0.003	0.018
e4	0.035	0.004
e5	0.033	0.013
e6	0	0.002
e7	0.057	0.08
e8	0	0.189
e9	0.151	0.074
e10	0.002	0.004
e11	0.048	0.002
e12	0.006	0.009
e13	0.122	0
e14	0.008	0.006
e15	0.087	0.002
e16	0.036	0.015
e17	0.092	0.027
e18	0.098	0.027
e19	0.036	0.003
e20	0.008	0.069
e21	0.061	0.011
e22	0.012	0.002
e23	0.042	0.147
e1	0.045	0.081

e2	0.044	0.036
e3	0.005	0.038
e4	0.039	0.004
e5	0.163	0.036
e6	0.042	0.121
e7	0.003	0.012
e8	0.002	0
e9	0.018	0.042
e10	0.138	0.002
e11	0	0.019
e12	0.041	0.084
e13	0.03	0.018
e14	0.018	0.006
e15	0	0.019
e16	0.001	0.04
e17	0.032	0.105
e18	0.09	0.155
e19	0.055	0.016
e20	0.017	0.065
e21	0.034	0.031
e22	0.11	0.025
e23	0.071	0.043

\$CRFIEi #presenta la contribución relativa del factor al elemento fila (KQ)

	[,1]	[,2]
e1	0.706	0.156
e2	0.66	0.033
e3	0.112	0.001
e4	0.415	0.531
e5	0.504	0.219
e6	0.002	0.256
e7	0.019	0.061
e8	0.667	0.191
e9	0.532	0.107
e10	0.647	0.291
e11	0.194	0.033
e12	0.583	0.27
e13	0.559	0.028
e14	0.001	0.561
e15	0.627	0.096
e16	0.583	0.046
e17	0.767	0.016
e18	0.568	0

e19	0.699	0.02
e20	0.372	0.025
e21	0.753	0.003
e22	0.749	0.004
e23	0.464	0.006
e1	0.878	0.084
e2	0.001	0.166
e3	0.114	0.505
e4	0.217	0.744
e5	0.565	0.355
e6	0.112	0.003
e7	0.428	0.132
e8	0.753	0.21
e9	0.35	0.47
e10	0.772	0.07
e11	0.81	0.154
e12	0.231	0.113
e13	0.834	0.136
e14	0.466	0.254
e15	0.331	0.175
e16	0.258	0.223
e17	0.522	0.186
e18	0.826	0.004
e19	0.057	0.024
e20	0.773	0.016
e21	0.78	0.005
e22	0.867	0
e23	0.955	0.006
e1	0.599	0.208
e2	0.452	0.398
e3	0.426	0.336
e4	0.868	0.01
e5	0.83	0.035
e6	0.088	0.134
e7	0.832	0.128
e8	0.003	0.535
e9	0.938	0.05
e10	0.231	0.044
e11	0.818	0.004
e12	0.25	0.041
e13	0.952	0
e14	0.845	0.077
e15	0.947	0.002

e16	0.904	0.039
e17	0.941	0.031
e18	0.942	0.028
e19	0.885	0.009
e20	0.393	0.373
e21	0.939	0.019
e22	0.622	0.009
e23	0.607	0.231
e1	0.498	0.054
e2	0.541	0.026
e3	0.368	0.157
e4	0.801	0.005
e5	0.865	0.011
e6	0.557	0.096
e7	0.642	0.14
e8	0.321	0.001
e9	0.564	0.078
e10	0.869	0.001
e11	0.003	0.14
e12	0.551	0.068
e13	0.868	0.032
e14	0.819	0.018
e15	0.013	0.08
e16	0.04	0.117
e17	0.583	0.115
e18	0.794	0.083
e19	0.851	0.015
e20	0.447	0.101
e21	0.838	0.045
e22	0.779	0.011
e23	0.751	0.027

\$CRTj #presenta la contribución relativa total de los elementos columna (KP)

[1]	0.25	0.321	0.078	0.125	0.038	0.1	0.088	0.084	0.183	0.154
[11]	0.133	0.267	0.09	0.09	0.13	0.124	0.08	0.181	0.075	0.096
[21]	0.314	0.029	0.038	0.177	0.226	0.041	0.127	0.361		

\$CREjFI #presenta la contribución relativa del elemento columna (KP) al factor

	[,1]	[,2]
T	0.332	0
S	0.346	0.473
OD	0.018	0.225
NO3	0.182	0.194

NO2	0.017	0
PO4	0	0.073
SiO4	0.104	0.034
T	0.001	0.405
S	0.213	0.262
OD	0.199	0.119
NO3	0.135	0.195
NO2	0.214	0
PO4	0.113	0.013
SiO4	0.125	0.006
T	0.142	0.046
S	0.08	0.49
OD	0.026	0.365
NO3	0.214	0.015
NO2	0.087	0
PO4	0.085	0.081
SiO4	0.366	0.002
T	0.006	0.016
S	0.023	0.008
OD	0.187	0.199
NO3	0.24	0.232
NO2	0.03	0.058
PO4	0.127	0.055
SiO4	0.387	0.432

\$CRFIEj #presenta la contribución relativa del factor al elemento columna (KP)

	[,1]	[,2]
T	0.688	0
S	0.56	0.013
OD	0.123	0.026
NO3	0.759	0.014
NO2	0.224	0
PO4	0.001	0.007
SiO4	0.612	0.003
T	0.004	0.29
S	0.699	0.086
OD	0.779	0.046
NO3	0.61	0.088
NO2	0.482	0
PO4	0.756	0.009
SiO4	0.833	0.004
T	0.85	0.035
S	0.503	0.388

OD	0.255	0.445
NO3	0.919	0.008
NO2	0.911	0
PO4	0.694	0.082
SiO4	0.907	0.001
T	0.15	0.029
S	0.456	0.011
OD	0.786	0.06
NO3	0.793	0.055
NO2	0.544	0.074
PO4	0.74	0.023
SiO4	0.801	0.064

\$QLREi #presenta la calidad relativa de los elementos fila (KQ)

e1	e2	e3	e4	e5	e6	e7	e8	e9	e10
0.862	0.693	0.113	0.946	0.723	0.258	0.08	0.858	0.639	0.938
e11	e12	e13	e14	e15	e16	e17	e18	e19	e20
0.228	0.854	0.586	0.562	0.723	0.629	0.782	0.568	0.718	0.397
e21	e22	e23	e1	e2	e3	e4	e5	e6	e7
0.755	0.753	0.469	0.963	0.167	0.619	0.961	0.92	0.114	0.56
e8	e9	e10	e11	e12	e13	e14	e15	e16	e17
0.964	0.82	0.842	0.964	0.344	0.969	0.72	0.506	0.481	0.708
e18	e19	e20	e21	e22	e23	e1	e2	e3	e4
0.83	0.081	0.789	0.785	0.867	0.96	0.807	0.85	0.761	0.878
e5	e6	e7	e8	e9	e10	e11	e12	e13	e14
0.865	0.222	0.96	0.538	0.988	0.275	0.822	0.291	0.952	0.921
e15	e16	e17	e18	e19	e20	e21	e22	e23	e1
0.95	0.943	0.971	0.97	0.894	0.766	0.957	0.631	0.838	0.552
e2	e3	e4	e5	e6	e7	e8	e9	e10	e11
0.568	0.525	0.806	0.877	0.653	0.782	0.322	0.642	0.87	0.142
e12	e13	e14	e15	e16	e17	e18	e19	e20	e21
0.619	0.9	0.836	0.093	0.157	0.698	0.876	0.866	0.548	0.883
e22	e23								
0.79	0.778								

\$QLREj #presenta la calidad relativa de los elementos columna(KP)

T	S	OD	NO3	NO2	PO4	SiO4
0.688	0.574	0.149	0.773	0.225	0.008	0.616
T	S	OD	NO3	NO2	PO4	SiO4
0.294	0.785	0.825	0.698	0.482	0.765	0.837
T	S	OD	NO3	NO2	PO4	SiO4
0.885	0.891	0.7	0.927	0.911	0.777	0.908
T	S	OD	NO3	NO2	PO4	SiO4
0.18	0.467	0.846	0.848	0.618	0.763	0.865

A9 Datos del Conjunto de Matrices de medidas Ambientales analizadas mediante el MixSTATICO

Temperatura (T)	Salinidad (S)	Oxígeno Disuelto (DO)	Nitrato (NO ₃)	Nitrito (NO ₂)	Fosfato (PO ₄)	Silicato (SiO ₄)
26.74	32.65	5.29	0.77	0.08	0.71	10.42
25.25	33.11	5.38	0.73	0.28	0.58	11.88
24.02	33.24	4.35	6.91	0.16	1.62	5.12
22.23	33.47	4.35	8.76	0.22	1.31	7.52
21.92	34.11	4.28	5.95	0.39	2.00	14.36
23.58	34.29	4.62	3.64	0.30	1.33	5.24
23.04	33.69	4.37	2.73	0.43	1.75	13.73
21.17	31.56	3.35	9.36	0.29	0.97	12.87
27.45	33.24	4.45	0.50	0.06	0.71	11.06
22.79	34.47	4.46	4.96	0.09	1.26	10.08
24.47	34.15	4.36	1.55	0.28	0.38	8.07
23.12	33.56	4.16	3.86	0.73	1.02	8.14
26.11	33.21	4.60	1.80	0.19	0.70	3.22
24.65	33.64	4.12	1.70	0.56	1.17	19.10
26.07	33.40	4.35	3.85	0.15	0.75	2.10
24.95	33.81	4.53	4.92	0.02	0.51	10.66
26.33	33.17	4.60	2.02	0.20	0.38	10.11
20.45	34.41	3.20	7.75	0.41	0.95	9.55
24.31	33.95	4.32	0.65	0.34	0.31	1.63
22.18	34.22	5.12	0.36	0.11	0.33	2.65
25.50	32.91	4.73	1.13	0.03	1.02	3.66
22.68	33.78	4.82	0.39	0.04	0.83	9.34
23.90	33.56	4.62	0.99	0.10	0.84	3.16
22.80	33.99	5.01	1.47	0.24	1.25	8.87
25.49	32.89	4.62	0.24	0.01	0.25	1.92
19.09	34.38	3.89	4.04	0.23	1.00	12.21
22.73	33.59	4.47	2.27	0.05	0.71	6.73
21.60	33.93	3.90	3.19	0.04	0.56	12.46
26.33	32.68	4.32	0.72	0.04	0.26	1.16
21.85	33.96	4.00	2.43	0.09	0.58	5.01
21.35	34.20	4.53	1.09	0.01	0.38	8.84
22.54	33.84	4.41	3.36	0.25	0.50	4.66
26.09	32.20	4.23	0.58	0.18	0.10	3.73
23.18	33.72	4.51	0.29	0.13	0.33	6.87
24.31	33.43	3.81	2.59	0.23	0.18	5.11
22.60	33.81	4.17	1.54	0.38	0.46	11.88
26.06	32.41	4.39	0.12	0.08	0.11	2.23
20.30	34.38	3.27	5.03	0.26	0.61	9.11
24.93	32.99	4.56	0.30	0.05	0.13	2.82
21.21	34.24	3.58	6.08	0.31	0.63	10.89
25.45	32.65	3.96	0.26	0.03	0.21	6.24
21.64	34.23	4.25	3.52	0.29	0.69	7.41
24.38	33.09	4.48	1.03	0.03	0.37	1.27
21.56	34.23	3.13	3.81	0.37	0.09	8.48

A10 Datos del Conjunto de Matrices de Especies analizadas mediante el MixSTATICO

Space	Time	<i>Dactyliosolen fragilissimus</i> (e7)	<i>Guinardia striata</i> (e8)	<i>Nitzschia longissima</i> (e16)	<i>Gymnodinium sp.</i> (e21)	<i>Gyrodinium sp.</i> (e22)	<i>Stauroneis membranacea</i> (e19)	<i>Mesodinium rubrum</i> (e23)	<i>Mesodinium rubrum</i> (e24)	<i>Stauroneis membranacea</i> (e25)	Grupo
Esmeraldas	Feb	0.00E+00	4.16E+04	2.28E+04	8.64E+03	2.75E+04	0.00E+00	5.50E+03	2	0	Céntricas
La Libertad	Feb	0.00E+00	2.91E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.36E+03	1	0	Céntricas
Manta	Feb	0.00E+00	2.36E+03	3.93E+03	7.85E+02	0.00E+00	0.00E+00	0.00E+00	0	0	Pennadas
Pto. Bolivar	Feb	0.00E+00	2.36E+03	6.99E+04	6.28E+03	0.00E+00	0.00E+00	3.53E+04	3	0	Pennadas
Esmeraldas	Mar	2.12E+04	1.49E+04	4.16E+04	6.28E+03	0.00E+00	1.34E+04	3.14E+03	1	3	Pennadas
La Libertad	Mar	2.22E+05	3.30E+04	3.30E+04	1.65E+04	3.14E+03	3.14E+03	4.71E+03	2	2	Céntricas
Manta	Mar	1.02E+04	1.88E+04	3.84E+05	1.65E+04	1.73E+04	1.26E+04	6.28E+03	2	3	Pennadas
Pto. Bolivar	Mar	6.28E+03	1.81E+04	1.14E+05	2.91E+04	0.00E+00	5.50E+03	7.85E+02	0	2	Céntricas
Esmeraldas	Abr	0.00E+00	0.00E+00	1.26E+05	3.14E+03	2.36E+03	0.00E+00	1.02E+04	3	0	Pennadas
La Libertad	Abr	0.00E+00	0.00E+00	3.06E+04	2.91E+04	0.00E+00	0.00E+00	4.71E+04	3	0	Ciliados
Manta	Abr	4.95E+04	3.30E+04	3.24E+05	6.28E+03	7.85E+02	2.04E+04	2.36E+03	1	3	Pennadas
Pto. Bolivar	Abr	0.00E+00	0.00E+00	1.49E+04	0.00E+00	7.85E+02	2.36E+03	1.57E+03	0	2	Pennadas
Esmeraldas	May	0.00E+00	3.22E+04	2.83E+04	0.00E+00	1.57E+03	0.00E+00	4.71E+03	2	0	Céntricas
La Libertad	May	6.28E+03	6.28E+03	0.00E+00	7.85E+02	0.00E+00	1.57E+03	7.85E+02	0	1	Céntricas
Manta	May	1.02E+04	1.65E+04	8.64E+03	0.00E+00	7.85E+02	0.00E+00	3.93E+03	2	0	Pennadas
Pto. Bolivar	May	0.00E+00	0.00E+00	1.34E+04	0.00E+00	1.57E+03	0.00E+00	3.38E+04	3	0	Ciliados
Esmeraldas	Jun	8.80E+04	1.79E+05	5.50E+03	0.00E+00	0.00E+00	1.10E+04	2.36E+03	1	3	Céntricas
La Libertad	Jun	4.63E+04	1.67E+05	4.08E+04	6.28E+03	0.00E+00	1.02E+04	7.85E+02	0	3	Céntricas
Manta	Jun	5.73E+04	1.07E+05	2.98E+04	7.85E+02	0.00E+00	0.00E+00	1.57E+03	0	0	Céntricas
Pto. Bolivar	Jun	1.03E+05	1.43E+06	4.08E+04	4.71E+03	0.00E+00	1.26E+04	1.02E+04	3	3	Céntricas
Esmeraldas	Jul	7.30E+04	1.37E+05	2.51E+04	7.85E+03	0.00E+00	1.57E+03	0.00E+00	0	1	Céntricas
La Libertad	Jul	3.38E+04	2.15E+05	4.71E+03	3.93E+03	0.00E+00	4.71E+03	7.85E+02	0	2	Céntricas
Manta	Jul	7.85E+02	0.00E+00	1.18E+04	1.10E+04	0.00E+00	1.34E+04	1.57E+03	0	3	Céntricas
Pto. Bolivar	Jul	2.20E+04	1.45E+05	6.28E+03	1.57E+03	0.00E+00	0.00E+00	1.57E+03	0	0	Céntricas
Esmeraldas	Ago	2.43E+04	1.11E+05	1.88E+04	1.02E+04	0.00E+00	4.71E+03	1.57E+03	0	2	Céntricas
La Libertad	Ago	1.18E+04	8.80E+04	8.64E+03	2.36E+04	1.57E+03	0.00E+00	6.28E+03	2	0	Céntricas
Manta	Ago	7.23E+04	1.67E+05	4.40E+04	9.42E+03	0.00E+00	2.83E+04	0.00E+00	0	3	Céntricas
Pto. Bolivar	Ago	0.00E+00	3.14E+03	7.07E+03	1.10E+04	0.00E+00	0.00E+00	1.65E+04	3	0	Céntricas
Esmeraldas	Sep	2.36E+03	8.56E+04	1.18E+04	1.10E+04	0.00E+00	6.28E+03	2.36E+03	1	2	Céntricas
La Libertad	Sep	0.00E+00	2.12E+04	2.59E+04	1.49E+04	0.00E+00	0.00E+00	2.36E+03	1	0	Céntricas
Manta	Sep	3.38E+04	1.04E+05	8.25E+04	7.85E+03	0.00E+00	4.71E+03	7.85E+03	3	2	Céntricas
Pto. Bolivar	Sep	0.00E+00	2.36E+03	3.85E+04	2.91E+04	0.00E+00	2.51E+04	7.85E+03	3	3	Pennadas
Esmeraldas	Oct	0.00E+00	7.07E+03	2.36E+03	1.10E+04	0.00E+00	0.00E+00	3.93E+03	2	0	Céntricas
La Libertad	Oct	0.00E+00	0.00E+00	1.35E+05	2.75E+04	0.00E+00	0.00E+00	3.14E+03	1	0	Pennadas
Manta	Oct	1.26E+04	4.95E+04	5.50E+03	3.22E+04	0.00E+00	3.14E+03	5.50E+03	2	2	Céntricas
Pto. Bolivar	Oct	0.00E+00	0.00E+00	7.07E+03	3.22E+04	0.00E+00	0.00E+00	2.36E+03	1	0	Dinoflagelados
Esmeraldas	Nov	2.91E+04	1.08E+05	6.28E+03	2.12E+04	0.00E+00	2.04E+04	7.85E+03	3	3	Céntricas
La Libertad	Nov	2.83E+04	2.03E+05	1.30E+05	2.04E+04	0.00E+00	0.00E+00	8.64E+03	3	0	Céntricas
Manta	Nov	1.65E+04	1.21E+05	4.71E+04	1.41E+04	0.00E+00	7.07E+03	5.50E+03	2	2	Céntricas
Pto. Bolivar	Nov	6.28E+03	0.00E+00	7.46E+04	2.36E+03	0.00E+00	0.00E+00	1.57E+03	0	0	Pennadas
Esmeraldas	Dic	1.73E+04	4.32E+04	0.00E+00	9.42E+03	0.00E+00	1.81E+04	7.07E+03	2	3	Céntricas
La Libertad	Dic	8.09E+04	1.23E+05	2.23E+05	5.50E+03	0.00E+00	7.07E+03	0.00E+00	0	2	Pennadas
Manta	Dic	4.48E+04	1.17E+05	1.41E+04	1.34E+04	0.00E+00	0.00E+00	3.93E+03	2	0	Céntricas
Pto. Bolivar	Dic	2.51E+04	2.66E+05	1.80E+05	8.64E+03	0.00E+00	0.00E+00	7.85E+03	3	0	Céntricas

Las abreviaturas de las variables están detalladas entre paréntesis. Variables discretas-cuantitativas (e19 y e23), cualitativas (ordinal: e24 y e25; nominal: Grupo).

A11 Datos de las Etiquetas para el Conjunto de Matrices analizadas mediante el MixSTATICO

Space	Time
Esmeraldas	Feb
La Libertad	Feb
Manta	Feb
Pto Bolivar	Feb
Esmeraldas	Mar
La Libertad	Mar
Manta	Mar
Pto Bolivar	Mar
Esmeraldas	Abr
La Libertad	Abr
Manta	Abr
Pto Bolivar	Abr
Esmeraldas	May
La Libertad	May
Manta	May
Pto Bolivar	May
Esmeraldas	Jun
La Libertad	Jun
Manta	Jun
Pto Bolivar	Jun
Esmeraldas	Jul
La Libertad	Jul
Manta	Jul
Pto Bolivar	Jul
Esmeraldas	Ago
La Libertad	Ago
Manta	Ago
Pto Bolivar	Ago
Esmeraldas	Sep
La Libertad	Sep
Manta	Sep
Pto Bolivar	Sep
Esmeraldas	Oct
La Libertad	Oct
Manta	Oct
Pto Bolivar	Oct
Esmeraldas	Nov
La Libertad	Nov
Manta	Nov
Pto Bolivar	Nov
Esmeraldas	Dic
La Libertad	Dic
Manta	Dic
Pto Bolivar	Dic