

UNIVERSIDAD NACIONAL DEL COMAHUE

Maestría en Estadística Aplicada



**“ALTERNATIVAS DE ANÁLISIS PARA EL ESTUDIO DE LA
RELACIÓN ENTRE COMPOSICIÓN DE TAXONES Y VARIABLES
AMBIENTALES EN DATOS DE BIOMONITOREOS ACUÁTICOS”**

Trabajo de Tesis presentado por:

Lorena Noemí Laffitte

Licenciada en Ciencias Biológicas

(Universidad CAECE - 2001)

Neuquén, 2021

COMISIÓN ASESORA DE TESIS

Director

Sergio Jorge Bramardi

Título Profesional

Ingeniero Agrónomo UNCo 1989

Títulos Académicos

Magister Scientiae en Biometría UBA 1995

Doctor en Estadística UP Valencia 2000

Codirectora

Sara Guadalupe Beamud

Título Profesional

Licenciada en Ciencias Biológicas UNCo 2001

Título Académico

Doctora en Ciencias Biológicas UNCo 2009

JURADO DE TESIS

Dr. Pablo Demetrio

Dr. Pablo Macchi

Mg. María Susana Vitelleschi

Fecha de aprobación de la tesis: 3 de septiembre de 2021

SERGIO JORGE BRAMARDI

SARA GUADALUPE BEAMUD

CERTIFICAN,

que **Lorena Noemí Laffitte** ha realizado bajo nuestra dirección, la tesis “ALTERNATIVAS DE ANÁLISIS PARA EL ESTUDIO DE LA RELACIÓN ENTRE COMPOSICIÓN DE TAXONES Y VARIABLES AMBIENTALES EN DATOS DE BIOMONITOREOS ACUÁTICOS”, la cual se encuentra finalizada y en condiciones de ser presentada.

Y para que así conste a todos los efectos del interesado se expide el presente certificado en Neuquén a los 3 días del mes de septiembre del año 2021.

*Dedicada,
a las Mujeres de mi
familia y a Lu*

“Aquel día descubrí que mi único rival no eran más que mis propias debilidades, y que en éstas, está la única y mejor forma de superarnos. Aquel día dejé de temer a perder y empecé a temer a no ganar, descubrí que no era yo el mejor y que quizás nunca lo fui, me dejó de importar quién ganara o perdiera, ahora me importa simplemente saberme mejor que ayer. Aprendí que lo difícil no es llegar a la cima, sino jamás dejar de subir. Aquel día dejé de ser un reflejo de mis escasos triunfos pasados y empecé a ser mi propia tenue luz de este presente. Aprendí que de nada sirve ser luz si no vas a iluminar el camino de los demás. Aquel día decidí cambiar tantas cosas... Aquel día aprendí que los sueños son solamente para hacerse realidad, desde aquel día ya no duermo para descansar... ahora simplemente duermo para soñar.”

~ Walt Disney ~

AGRADECIMIENTOS

Agradezco en primer lugar a Sergio Bramardi, estupenda persona gracias a la cual me he acercado al fabuloso mundo de la estadística, quien con inteligencia, pasión, respeto y cariño supo guiarme desde hace muchos años y darme ánimo para andar este camino que tantas puertas me ha abierto.

A Guadalupe Beamud, a quien la sincronía de la vida la puso en mi camino en el momento justo y aceptó el desafío de acompañarme, siempre presente, atenta y dispuesta a avanzar conmigo compartiendo todo su conocimiento.

Agradezco muy afectuosamente a mi amigo y compañero de trabajos José Saravia, con quien compartimos entre otras cosas el gran esfuerzo de haber recorrido miles de kilómetros para realizar la maestría.

Al resto de mis compañeros y amigos, por el tiempo de estudio y esfuerzo, Natalia Rubio, Guillermo Sabino, Jenny Fuentealba, Candelaria Morelli y especialmente a Andrea Lavalle por su apoyo, tiempo y consejos.

Agradezco a la profesora Silvia Boche que supo tenerme extremada paciencia y quien siempre estuvo, esforzándose porque todo saliera bien y porque todos siguiéramos adelante. A los buenos profesores que inspiran, Julio Di Rienzo, María del Pilar Díaz, Nora Baccalá.

A la “Provincia del Neuquén”, mi querida provincia, donde trabajo hace 20 años y en la que puedo con mucha satisfacción ejercer mi profesión en la Subsecretaría de Ambiente, organismo que siempre ha alentado la capacitación y que en este caso resultó, además, fuente de los datos que se analizan.

Agradezco también a mis compañeros de la Dirección General de Biología Acuática, con quien comparto todas mis jornadas laborales, y a la Dirección Provincial de Fiscalización Hídrica, a todos por las tareas de campo y tomas de muestras.

Agradezco a Julieta Muñiz Saavedra, visionaria y ejecutiva, quien propició la metodología de biomonitoreos en la provincia. A Betina Laurenzano, Héctor Labollita y Roberto Caro por trabajar con esmero por la continuidad de los monitoreos. A Joe Flotemersch, por sus capacitaciones y fotografías.

A la Universidad Nacional del Comahue por brindarme la posibilidad de realizar mis estudios de posgrado y donde también tengo el honor de devolver a través de la docencia el conocimiento recibido, en el Asentamiento Universitario San Martín de los Andes (AUSMA).

A mis amigas, Mer Iummato, siempre alentando para que pudiera concluir con este trabajo, Vivi, Lore, Mariel, Imi, Joqui, Cecis.

Quiero agradecer sinceramente y de corazón a todos los mencionados y a los que seguro faltan, que hicieron posible que pudiera realizar la Maestría en Estadística y este trabajo final de Tesis.

Y, por último, pero no por eso menos importante, a mi querida familia, mis padres Lucía y Rodolfo, mi hermano Esteban, mi tío Pancho, mis abuelas Negra y Chiquita, mi compañero Erni y mis preciosos hijos Félix y Francine.

ÍNDICE GENERAL

	Página
Agradecimientos	VI
Resumen en español	XVIII
Resumen en inglés	XIX
<hr/>	
CAPÍTULO 1	
Introducción	1
1.1 Planteo del Problema	5
1.1.1 Problema biológico	5
1.1.2 Problema estadístico	6
1.2 Objetivos	10
1.3 Hipótesis	10
<hr/>	
CAPÍTULO 2	
Marco Teórico	
2.1 Conceptos generales de los métodos de ordenación	11
2.1.1 Análisis indirecto del gradiente (comparación indirecta)	13
2.1.2 Análisis directo del gradiente (comparación directa)	14
2.1.3 Pre-transformación de datos	15
2.1.4 Tipos de respuesta “lineal” y “unimodal”	19
2.1.5 Matrices de datos	21
2.2 Métodos de análisis indirecto del gradiente	22
2.2.1 Breve descripción de los métodos ACP, AC y DCA	22
2.2.1.1 Análisis de Componentes Principales (ACP)	22
2.2.1.2 Análisis Factorial de Correspondencias (AC)	26
2.2.1.3 Análisis de Correspondencias Destendenciado (DCA)	29
2.3 Métodos de análisis directo del gradiente	32
2.3.1 Análisis factorial múltiple (AFM)	32
2.3.1.1 Pasos	32
2.3.1.2 Gráficos	35
2.3.1.3 Elementos de interpretación	38

2.3.2	Análisis de Redundancia (RDA)	40
2.3.2.1	Pasos	41
2.3.2.2	Gráficos	46
2.3.3	Análisis Canónico de Correspondencias (ACC)	50
2.3.3.1	Pasos	53
2.3.3.2	Gráficos	55
2.3.3	Prueba de significancia del RDA y ACC	58
2.4	Esquemas de los métodos AFM, RDA y ACC	61
CAPÍTULO 3		
Materiales		
3.1	Biomonitoreo cuenca Lacar - Hua Hum	64
3.2	Monitoreo del alga exótica invasora <i>Didymosphenia geminata</i>	72
3.3	Software	77
CAPÍTULO 4		
Resultados y discusión		
4.1	Biomonitoreo cuenca Lacar – Hua Hum	79
4.1.1	Resultados	79
4.1.1.1	Análisis indirecto del gradiente: AFM	79
4.1.1.2	Análisis directo del gradiente: RDA	87
4.1.2	Discusión de aspectos biológicos	93
4.2	Monitoreo del alga exótica invasora <i>Didymosphenia geminata</i>	97
4.2.1	Resultados	97
4.2.1.1	Análisis indirecto del gradiente: AFM	97
4.2.1.2	Análisis directo del gradiente: ACC	104
4.2.2	Discusión de aspectos biológicos	110
4.3	Discusión general de aspectos estadísticos	112
4.3.1	AFM vs RDA	112
4.3.2	AFM vs ACC	115
CAPÍTULO 5		
Conclusiones		120

BIBLIOGRAFÍA	123
---------------------	-----

ANEXO

A1. Recursos Web	132
A2. Sentencias de R	132
A3. Tablas	141

ÍNDICE DE CUADROS

Tabla	Página
4.1 Inercia de los cuatro primeros ejes del AFM LHH.	80
4.2 Calidad de representación del AFM LHH por grupos.	80
4.3 Coeficientes Lg y RV entre grupos. AFM LHH.	81
4.4 Correlaciones significativas (p -valor $<0,05$) de variables ambientales y taxones con los dos primeros ejes del AFM LHH.	84
4.5 Resumen de los 4 primeros ejes canónicos del RDA LHH. Salidas de R y CANOCO.	88
4.6 Valores de factor de inflación de la varianza (VIF), correlación y coordenadas de las variables ambientales con los dos primeros ejes canónicos. RDA LHH.	89
4.7 Inercia de los cuatro primeros ejes del AFM DIDYMO.	98
4.8 Calidad de representación AFM DIDYMO por grupos.	98
4.9 Coeficientes Lg y RV entre grupos. AFM DIDYMO.	98
4.10 Correlaciones significativas (p -valor $< 0,05$) de variables ambientales y taxones con los dos primeros ejes del AFM DIDYMO.	101
4.11 Resumen de los 4 primeros ejes canónicos del ACC DIDYMO. Salidas de R y CANOCO.	105
4.12 Valores de factores de inflación de la varianza (VIF), de correlación y coordenadas de las variables ambientales con los dos primeros ejes canónicos. ACC DIDYMO.	106
I Detalle de los 21 sitios muestreados, años 2015 y 2016, incluyendo: subcuenca (TQ: Trabunco-Quitrahue, P: Pocahullo, L: Lacar-Hua Hum), sitio, nombre, código de muestra e índice BMPS según año de muestreo. Monitoreo LHH.	141

II	Escala de calidad de agua correspondiente al Índice BMPS (Miserendino y Pizzolón, 1999).	142
III	Listado completo de familias de macroinvertebrados incluyendo grupo, taxón, abreviatura utilizada en gráficos, puntuación del grado de sensibilidad a la contaminación (índice BMPS) y frecuencia absoluta. Monitoreo LHH.	142
IV	Estadísticos resumen de variables ambientales. E.E.: error estándar, Mín: mínimo, Máx.: máximo. Monitoreo LHH.	144
V	Valores observados de variables ambientales por sitio. Monitoreo LHH.	144
VI	Listado completo de sitios incluyendo código de sitio (utilizado en gráficos), cuenca y nombre. Monitoreo DIDYMO.	145
VII	Listado completo de especies de algas incluyendo grupo taxonómico, taxón, abreviatura utilizada en gráficos y frecuencia relativa porcentual. Monitoreo DIDYMO.	146
VIII	Estadísticos resumen de variables ambientales. E.E.: error estándar, Mín: mínimo, Máx.: máximo. Monitoreo DIDYMO.	148
IX	Estadísticos resumen de variables ambientales por cuencas. E.E.: error estándar, Mín: mínimo, Máx.: máximo. Monitoreo DIDYMO.	148
X	Valores observados de variables ambientales por sitio. Monitoreo DIDYMO.	150
XI	Planilla de campo para elaboración del índice cualitativo “Puntaje de evaluación del hábitat – PEH”. Adaptado de Barbour <i>et al.</i> , 1999.	151

ÍNDICE DE FIGURAS

Figura		Página
2.1	Esquema de clasificación de los métodos multivariados de ordenación expuestos en esta tesis según y Borcard <i>et al.</i> (2011) y Legendre y Legendre (2012).	12
2.2	Respuesta lineal (a) versus unimodal (b) de aptitud/abundancia de taxones a lo largo de un gradiente ambiental.	19
2.3	Respuesta hipotética de taxones a lo largo un gradiente ambiental corto (a) y largo (b). La línea roja indica el segmento del gradiente realmente muestreado (parte de la gradiente que nos interesa), y la línea amarilla indica cómo se vería la respuesta de la especie si se ajustara mediante un modelo lineal. Lepš y Šmilauer (2003)	20
2.4	Matriz Y de frecuencias y X de variables continuas.	21
2.5	Ejemplo de efecto arco como resultado de la ordenación del AC.	30
2.6	Matriz de frecuencias F' y matriz descriptiva X est centrada por los pesos f_{i+} . Abdessemed y Escofier (1996).	33
2.7	Matriz yuxtapuesta del AFM.	35
2.8	Esquema general del AFM.	61
2.9	Esquema general de RDA (adaptado de Legendre y Legendre, 2012). Si la matriz Y es de presencia/ausencia o frecuencias, usar transformación de Hellinger.	62
2.10	Esquema general del ACC. Para esta tesis las matrices de entrada Y y X correspondieron a \bar{Q} y X est	63
3.1	Mapa de distribución de sitios de muestreo. Biomonitorio Lacar Hua Hum. Años 2015 y 2016.	69
3.2	Imágenes del biomonitorio Lacar – Hua Hum. Años 2015 y 2016.	70
3.3	Mapa de distribución de sitios de muestreo. Los símbolos de sitios están representados en colores según la cuenca de	69

	pertenencia. Monitoreo del alga exótica invasora <i>Didymosphenia geminata</i> . Provincia del Neuquén. Año 2014.	
3.4	Imágenes del monitoreo de <i>Didymosphenia geminata</i> en la Provincia del Neuquén.	74
4.1	Representación de los grupos conformados por las matrices X (Cuanti.lhh) e Y (Frec.hel.lhh). Plano principal AFM LHH.	82
4.2	Representación de sitios con baricentros y puntos parciales. Plano principal del AFM LHH.	83
4.3	Círculo de correlación de variables ambientales de la matriz X . AFM LHH.	84
4.4	Caracterización conjunta del AFM LHH en el plano principal. Representación de individuos (sitios) y frecuencias (familias de macroinvertebrados). Colores de etiquetas de sitios: subcuencas (violeta: Pocahullo, verde: Trabunco-Quitrahue, azul: Lacar). Colores de círculos: grado de sensibilidad de familias a la contaminación (azul: alto, verde: moderado, amarillo: bajo).	86
4.5	RDA Biplots, primeros dos ejes canónicos. Salida CANOCO. (a) Biplot de sitios y variables ambientales, escalado 1 (subcuencas: círculo verde, Trabunco-Quitrahue; rombo celeste, Lacar-Hua Hum; cuadrado rojo, Pocahullo). (b) Biplot de familias y variables ambientales, escalado 2.	90
4.6	Triplot del RDA LHH, primeros dos ejes canónicos. Salida del CANOCO. Escalado 2. Relación sitio-familia-ambiente. Los vectores rojos representan las variables predictoras ambientales, los vectores azules las frecuencias de familias de macroinvertebrados y los puntos los sitios, identificando con color las subcuencas (fucsia: Pocahullo, verde: Trabunco-Quitrahue, azul: Lacar).	92
4.7	Representación de los grupos conformados por las matrices X (Cuanti.didy) e Y (Frec.didy) en el plano principal del AFM DIDYMO.	99

4.8	Representación de sitios con baricentros y puntos parciales en el plano principal. AFM DIDYMO.	100
4.9	Círculo de correlación de variables ambientales de la matriz X . Plano principal del AFM DIDYMO.	102
4.10	Caracterización conjunta AFM DIDYMO. Representación conjunta de individuos (sitios) y frecuencias (especies de algas). Los colores de etiquetas de sitios identifican las cuencas (púrpura: río Colorado, verde: río Hua Hum, amarillo: río Neuquén, azul: río Limay).	104
4.11	ACC Biplots, primeros dos ejes canónicos. Salida CANOCO. (a) Biplot de sitios y variables ambientales, escalado 1. Los colores de símbolos identifican las cuencas (púrpura: río Colorado, verde: río Hua Hum, amarillo: río Neuquén, azul: río Limay). (b) Biplot de especies y variables ambientales, escalado 2.	107
4.12	Triplot del ACC DIDYMO, primeros dos ejes canónicos. Salida de CANOCO. Escalado 2. Los vectores rojos representan las variables predictoras ambientales, los nombres en azul representan las especies de algas. Los colores de los símbolos de sitios identifican las cuencas (púrpura: río Colorado, verde: río Hua Hum, amarillo: río Neuquén, azul: río Limay).	109

ABREVIATURAS

AC	Análisis de Correspondencias
ACC	Análisis de Correspondencias Canónico
ACoP	Análisis de Coordenadas Principales
ACorC	Análisis de Correlación Canónica
ACP	Análisis de Componentes Principales
ACol	Análisis de Co-Inercia
AD	Análisis Discriminante
AFM	Análisis Factorial Múltiple
ALC	Alcalinidad
ALT	Altitud
CE	Conductividad eléctrica
Db-RDA	Análisis de redundancia basado en la distancia
DCA	Análisis de correspondencia destendenciado o por segmentos
DE	Desvío estándar
DIDYMO	<i>Didymosphenia geminata</i>
DU	Dureza
DVS	Descomposición en valores singulares
EMNM	Escalamiento multidimensional no métrico
FE	Hierro
Lg	Coefficiente Lg
LHH	Lacar – Hua Hum
NT	Nitrógeno total
OD	Oxígeno disuelto
P	Pocahullo
PEH	Puntaje de evaluación del hábitat
PH	Nivel de acidez o alcalinidad
PRS	Fósforo reactivo soluble
PT	Fósforo total
RC	Río colorado
RDA	Análisis de redundancia
RL	Río Limay

RN	Río Neuquen
RV	Coficiente RV de Escofier
SI	Sílice
SRH-	
MESPyR N	Subsecretaría de Recursos Hídricos – Ministerio de Energía, Servicios Públicos y Recursos Naturales
SSA- MSTyA	Subsecretaría de Ambiente – Ministerio de Seguridad, Trabajo y Ambiente
SST	Sólidos en suspensión totales
TAG	Temperatura del agua
TQ	Trabunco-Quitrahue
US EPA	Agencia de Protección Ambiental de Estados Unidos
VIF	Factores de inflación de varianza

Como citar esta Tesis:

Laffitte, L.N. 2021. Alternativas de análisis para el estudio de la relación entre composición de taxones y variables ambientales en datos de biomonitoreos acuáticos. Tesis de Maestría en Estadística Aplicada. FAEA, Universidad Nacional del Comahue, Argentina.

RESUMEN

Los métodos de ordenación son bien conocidos por los ecólogos que desean representar y resumir sus observaciones en un espacio reducido a lo largo de uno, o más ejes. El abanico de métodos multivariados es muy amplio, en este trabajo se expusieron tres métodos de ordenación de análisis directo del gradiente (o canónicos): uno simétrico, análisis factorial múltiple y dos asimétricos, análisis de redundancia y análisis de correspondencias canónico. Mediante la utilización de estos métodos, se sometieron a análisis conjuntos de datos compuestos por matrices de frecuencias de composición de taxones y matrices de variables ambientales, para sitios de dos biomonitoreos realizados en la Provincia del Neuquén. El análisis factorial múltiple permitió cuantificar la concordancia entre las dos matrices de datos e identificar asociaciones entre sitios, taxones y variables ambientales. Los métodos de ordenación canónica permitieron ahondar en las relaciones, determinar gradientes ambientales y definir variables significativas en la ordenación de los sitios y taxones. Ambos biomonitoreos se asociaron a escalas espaciales sumamente distintas, no obstante, los resultados demostraron una clara importancia en la diferenciación existente entre sitios prístinos y más degradados, en el caso de la cuenca Lacar - Hua Hum debido a la transición entre zonas de cabeceras de cuenca y zonas antropizadas por asentamientos humanos; en el caso del biomonitoreo del alga exótica invasora con una escala provincial que evidenció la gran diferencia entre las principales cuencas provinciales y un gradiente de contaminación. El software CANOCO es de un entorno amigable, con excelentes gráficos e información ecológica intrínseca que guía al usuario. El software R con ilimitada potencia para realizar cualquier análisis requiere de conocimiento de base, estadístico y ecológico. Ambos softwares pudieron compararse y complementarse para la realización de este trabajo.

Palabras clave: ordenación, ordenación canónica, AFM, RDA, ACC, macroinvertebrados, hábitat, calidad del agua, Neuquén.

" ANALYSIS ALTERNATIVES FOR THE STUDY OF THE RELATIONSHIP BETWEEN TAXON COMPOSITION AND ENVIRONMENTAL VARIABLES IN AQUATIC BIOMONITORING DATA"

ABSTRACT

Ordination methods are well known by ecologists who want to represent and summarize their observations along one or more axes in a small space. Multivariate methods range is very wide, in this work, three direct gradient (or canonical) ordination methods were exposed: one symmetric, multiple factorial analysis; and two asymmetric, redundancy analysis and canonical correspondence analysis. Using these methods, data sets composed of taxa composition frequencies and environmental variables matrix were subjected to analysis of sites of two biomonitoring carried out in the Neuquen Province. Multiple factor analysis allowed us to quantify the agreement between both data matrix and to identify associations between sites, taxa, and environmental variables. Canonical ordering methods allowed deepening relationships, determining environmental gradients and defining significant variables in the ordering of sites and taxa. Both biomonitoring were associated with extremely different spatial scales, however, results demonstrated a clear importance in the differentiation between pristine and more degraded sites. In Lacar - Hua Hum basin, due to transition between headwaters zones and anthropized areas by human settlements; in the case of the invasive exotic algae biomonitoring, with a provincial scale that evidenced the great difference between main basins and a pollution gradient. CANOCO software is a friendly environment, with excellent graphics and intrinsic ecological information that guides the user. R software with unlimited power to perform any analysis requires basic, statistical and ecological knowledge. Both softwares could be compared and complemented to carry out this work.

Key words: ordination, canonical ordination, MFA, RDA, CCA, macroinvertebrates, habitat, water quality, Neuquen.

Esta página es intencionalmete dejada en blanco

This page is intentionally left in blank

CAPÍTULO 1. Introducción

Los ecosistemas acuáticos cambian constantemente debido a factores estresantes que pueden ser naturales y/o antropogénicos. Considerando a dichos ecosistemas como 'objetivos móviles', uno de los mayores desafíos para las evaluaciones ambientales que enfrentan los científicos, gerentes de los recursos y formuladores de políticas, es utilizar escalas de tiempo apropiadas. Al respecto, la mayoría de los sistemas acuáticos carecen de datos de monitoreo y, si existe un programa de seguimiento, rara vez se han recopilado datos durante más de unos pocos años. Por lo tanto, a menudo es difícil o imposible determinar la naturaleza y el momento de los cambios en el ecosistema con base en estos conjuntos de datos a corto plazo. Además, como las evaluaciones ambientales generalmente se realizan después de identificar un problema, los datos críticos sobre la línea de base previa a la perturbación (o de referencia) rara vez están disponibles (Smol *et al.*, 2001).

El uso de protocolos estandarizados ya sea en el campo o en el laboratorio, es una característica clave en las evaluaciones ambientales. Esto permite que los datos obtenidos sean comparables y puedan combinarse o tratarse para producir evaluaciones consistentes.

Los ríos, arroyos, lagos y lagunas de la Provincia del Neuquén representan un patrimonio natural de valor excepcional, tanto por su biodiversidad y belleza paisajística, como por su valor económico para la provisión de agua segura y por su uso recreativo con actividades a nivel internacional. Algunos de los estudios que se desarrollan sobre estos recursos son a nivel de cuencas hidrográficas, que son unidades espacialmente explícitas en el paisaje, y contienen una serie de atributos físicos, ecológicos y sociales que interactúan entre sí. Son sistemas socio-ecológicos que proporcionan una gama de servicios ecosistémicos valorados por la sociedad. Su capacidad de proporcionar estos servicios depende, en parte, de la medida en que se vean afectados por la actividad humana.

En los últimos años, la integridad biológica de los cuerpos de agua ha sido el foco de atención de profesionales y ciudadanos y en la Provincia del Neuquén se ha consolidado la metodología de trabajo en evaluación biológica y clasificación de

hábitats. De acuerdo con Barbour *et al.* (1999), la “bioevaluación” se refiere al proceso de evaluación de la condición biológica de un cuerpo de agua mediante mediciones directas de la biota residente, o sea de aquellos organismos que viven en las aguas superficiales, incluidos peces, insectos, algas, plantas y otros.

Los protocolos de bioevaluación rápida recomiendan una evaluación integrada, comparando el hábitat (por ej., estructura física, régimen de flujo), calidad del agua y medidas biológicas con condiciones de referencia empíricamente definidas (a través de sitios de referencia reales, datos históricos y/o modelos o extrapolaciones). Las condiciones de referencia se establecen mejor a través del monitoreo sistemático de sitios reales que representan el rango natural de variación en la química del agua “mínimamente perturbada”, el hábitat y las condiciones biológicas (Gibson *et al.*, 1996).

El monitoreo puede realizarse para varios procesos o sistemas biológicos con el objetivo de observar los cambios temporales y espaciales en el estado de salud, evaluar los impactos de un ambiente específico o estresores antropogénicos y evaluar la viabilidad de las medidas antropogénicas (por ejemplo, recuperación, remediación y/o reintroducción). La diversidad de especies se utiliza como un aspecto principal en el monitoreo biológico, que se considera un parámetro valioso para determinar la salud del ambiente. El biomonitoreo es uno de los componentes esenciales para evaluar la calidad del agua y se ha convertido en un elemento integral para realizar estudios sobre la contaminación del agua (Parmar *et al.*, 2016).

Los protocolos que se utilizan en la Provincia del Neuquén están adaptados de los Protocolos Rápidos de Evaluación de Ríos y Arroyos propuestos por la Agencia de Protección Ambiental de Estados Unidos (US EPA), para la evaluación rápida y completa del hábitat estudiado y la caracterización del cuerpo de agua (Barbour *et al.*, 1999; Flotemersch *et al.*, 2006).

Estos protocolos se basan en la aplicación de bioindicadores. En 1969 Margalef hablaba de indicadores biológicos sin dar una definición exacta del término. La expresión “bioindicador” se utiliza como un término agregado que se refiere a todas las fuentes de reacciones bióticas y abióticas a los cambios ecológicos. En

lugar de simplemente trabajar como indicadores del cambio natural, los taxones se utilizan para mostrar los impactos de los cambios naturales circundantes o el cambio ambiental, así como para indicar impactos negativos o positivos. También pueden detectar cambios en el ambiente debido a la presencia de contaminantes que pueden afectar la biodiversidad, así como las especies presentes en él (Gerhardt, 1999; Parmar *et al.*, 2016).

Tradicionalmente, la evaluación y el monitoreo de los ecosistemas acuáticos, se ha basado en el estudio de la calidad del agua, a partir de la medición de variables físico-químicas. Sin embargo, estos no son suficientes para evaluar su integridad ecológica. Como consecuencia surge la necesidad de incorporar otros componentes que reflejen el estado ambiental de los cuerpos de agua que consideren las alteraciones del caudal o la degradación del hábitat, que a su vez afectan a las comunidades biológicas y los servicios ecosistémicos. El uso de biomonitores en las evaluaciones ha sido señalado como un indicador más preciso de las condiciones ambientales que las físico-químicas, ya que la biota está expuesta a la totalidad de los estresores integrando además su efecto en el tiempo. Por lo tanto, el biomonitoreo permite evaluar los efectos de la sinergia producida por diversos factores de estrés (Domínguez *et al.*, 2020).

Instalar el uso de bioindicadores como evaluadores de calidad ambiental en cuencas neuquinas permitiría actuaciones rápidas de parte de la administración provincial. No obstante, el diseño de un sistema de biomonitoreo que permita determinar la calidad del ecosistema, requiere establecer el valor indicador de las especies que ahí habitan, usando la presencia y abundancia de organismos y el monitoreo físico y químico del cuerpo de agua. El monitoreo biológico da una visión histórica de la salud del ecosistema y los datos físicos y químicos son necesarios para establecer las causas concretas de cambios naturales cuando los hubiere, o bien para el seguimiento y cuantificación de impactos antrópicos en el ecosistema (Friberg *et al.*, 2011)

La ecología de comunidades de enfoque cuantitativo es una de las ramas más desafiantes de la ecología moderna. Los ecólogos comunitarios generalmente necesitan analizar los efectos de múltiples factores ambientales en docenas (si no en cientos) de especies simultáneamente y los errores estadísticos (tanto de

medición como estructurales) tienden a ser muy grandes y de mal comportamiento. Por lo tanto, no es sorprendente que los ecólogos hayan empleado una variedad de enfoques multivariados para el análisis de datos de comunidades (Palmer, 1993).

En virtud de lo planteado y dada la naturaleza de la información ambiental a analizar, esta tesis propone el estudio estadístico de dos grupos de datos multivariados mediante los métodos detallados a continuación, haciendo énfasis en el estudio de la naturaleza matemática de cada método y la pertinencia en función de las características de los datos:

- Análisis Factorial Múltiple – AFM
- Análisis de Redundancia – RDA (siglas expresadas en inglés)
- Análisis de Correspondencias Canónico - ACC

Los análisis de datos multivariados son utilizados en todos los campos de la ciencia. El Análisis Factorial Múltiple (AFM; Escofier y Pagès, 1984, 1990) es utilizado en distintos ámbitos, mientras que el Análisis de Redundancia (RDA; Rao, 1964; Wollenberg, 1977) y el Análisis de Correspondencias Canónico (ACC; ter Braak, 1986, 1987) son ampliamente utilizados en ecología.

Se trabajará con dos bases de datos de origen observacional, conformadas por dos matrices de datos, una matriz de frecuencias (absolutas y relativas) de taxones de algas y macroinvertebrados y una matriz de variables ambientales compuesta por registros físico-químicos de medición *in situ*, resultados de análisis de laboratorio y caracterización del hábitat.

Base de datos biomonitoreo Lacar Hua-Hum (LHH): está compuesta por 21 sitios de muestreo que constituyen los objetos de estudio o individuos, a los que se asocian 13 variables ambientales y 46 variables que corresponden a las frecuencias absolutas de familias de macroinvertebrados.

Base de datos biomonitoreo *Didymosphenia geminata* (Lyngbye) M. Schmidt (*D. geminata*) (DIDYMO): compuesta por 44 sitios de muestreo de ríos y arroyos, 9 variables ambientales y frecuencias relativas de 44 especies de algas.

1.1. Planteo del problema

1.1.1. Problema biológico

El gran problema que se genera en las administraciones públicas es que muchas veces se obtienen los recursos necesarios para realizar monitoreos y generar datos, pero que carecen de evaluación posterior, por lo que el potencial de lo logrado en campo no se traduce en la gestión de los recursos. La evaluación estadística entonces, pasa a ser fundamental y es indispensable que se faciliten los procesos de evaluación, proponiendo los análisis adecuados para responder a los objetivos que puedan ser planteados.

Asimismo, los análisis reportados en la literatura frecuentemente se llevan a cabo utilizando técnicas que no están totalmente adaptadas a los datos en estudio, lo que lleva a conclusiones que no son óptimas con respecto a las observaciones de campo (Legendre y Legendre, 2012).

Los problemas para resolver en los estudios ecológicos son siempre complejos porque los datos ecológicos están sujetos a la no linealidad y presentan una gran complejidad debido a muchas variables y a sus interacciones. Por lo tanto, se recomienda estudiar los problemas ecológicos utilizando métodos apropiados que puedan manejar dicha complejidad. Además, en los estudios ecológicos modernos, los requisitos para el conocimiento de las interacciones entre los ecosistemas y las propiedades ecológicas han aumentado considerablemente. Durante las últimas décadas, la investigación intensiva ha empleado varias técnicas para dilucidar problemas ecológicos de lo más variado. Junto con el rápido desarrollo de la informática y las ciencias de la información, se ha desarrollado una gran cantidad de técnicas y algoritmos, que se han utilizado para la gestión del ecosistema, así como para estudios ecológicos (Park *et al.*, 2015).

A menudo, los estudios ecológicos enfrentan la dificultad de abordar el análisis de grandes bases de datos con variables de diferentes características. Estos estudios se basan en muestreos donde en cada muestra obtenida (tiempos, sitios, etc.) se miden diferentes variables ambientales y/o la composición de especies (provenientes de distintas comunidades de organismos). Los objetivos de los

análisis suelen ser diferentes según la necesidad, por lo tanto, el abordaje estadístico puede ser diferente con un mismo conjunto de datos. Por consiguiente, la cantidad y naturaleza de las variables hace que los análisis multivariados sean los mejores a la hora de extraer la primera información de los datos, a través de un análisis exploratorio.

La estadística aplicada a datos ecológicos se ha convertido en una disciplina científica propia y el buen uso de los métodos de análisis de datos multivariados por parte de los ecólogos implica un conocimiento adecuado de las propiedades matemáticas de estos métodos (Thioulouse *et al.*, 2018).

1.1.2. Problema estadístico

Los desarrollos teóricos en análisis multivariados que sustentan casi la totalidad de numerosas aplicaciones modernas se produjeron en el siglo XX con un fuerte impacto de las escuelas francesa, hindú y norteamericana. Los orígenes del “Análisis Multivariado” se remontan a las primeras generalizaciones de la correlación y regresión. Uno de los primeros trabajos es el de Pearson (1901), donde se establecen las primeras ideas de Componentes Principales. Otro de los trabajos pioneros se debe a Spearman (1904). Durante el primer tercio de siglo se produjeron importantes avances que posteriormente posibilitaron el desarrollo del análisis multivariado. Cabe destacar los aportes de Fisher en el análisis de la varianza, la razón de verosimilitud de Neyman y Pearson y otros avances en Chi-cuadrado, correlación y proximidades. El establecimiento definitivo de la mayoría de los análisis multivariados se produce alrededor de los años treinta, con las contribuciones de Hotelling (1931, 1933); Wilks (1932, 1935); Fisher (1935, 1936); Mahalanobis (1936) y Bartlett (1939). En principio el análisis multivariado quedó reducido al campo de la especulación teórica. Posteriormente se estudiaron las posibles aplicaciones (Rao, 1952).

El análisis de datos multivariados comprende el estudio estadístico de varias variables medidas en elementos de una población con los siguientes objetivos (Peña, 2002):

1. Resumir los datos mediante un pequeño conjunto de nuevas variables, construidas como transformaciones de las originales, con la mínima pérdida de información.
2. Encontrar grupos en los datos, si existen.
3. Clasificar nuevas observaciones en grupos definidos.
4. Relacionar dos conjuntos de variables.

El análisis multivariado de datos proporciona métodos objetivos para conocer cuántas variables indicadoras, que a veces se denominan factores, son necesarias para describir adecuadamente una realidad compleja y determinar las propiedades de estas variables.

Podemos clasificar a los métodos multivariados en dos grandes grupos:

- 1) Clasificación: análisis que categorizan o clasifican los objetos de investigación.
- 2) Ordenación: análisis que reducen los datos a variables sintéticas (ordenamientos).

Los métodos de ordenación se refieren, literalmente, a organizar los ítems a lo largo de una escala (eje) o de varios ejes. Se usan estos métodos como una manera de resumir gráficamente relaciones complejas, extrayendo uno o unos pocos patrones dominantes de un gran número de posibles patrones. El proceso de extraer esos ejes se llama “ordenación” porque es el resultado de colocar objetos a lo largo de un eje o dimensión.

Los ordenamientos se usan principalmente en ecología para buscar y describir patrones, que ayudan a:

- ✓ seleccionar los factores más importantes de un conjunto de factores hipotéticos.
- ✓ separar patrones fuertes de patrones débiles.
- ✓ revelar patrones y sugerir posibles procesos.

Guo *et al.* (2015), describen a la ordenación como un método multivariado de análisis de gradiente y reducción de datos en el que la distribución de las muestras se organiza en unas pocas dimensiones basadas en el análisis propio o la similitud

entre las muestras (es decir, un parecido, correlación, o matriz de covarianza). Se utiliza popularmente en la ecología comunitaria cuantitativa. Se definen dos grupos principales: la ordenación sin restricciones (una matriz de datos) y la ordenación restringida (u ordenación canónica) (dos o más conjuntos de datos).

La ordenación para datos de dos vías, es decir individuos por variables, revela su estructura principal en un gráfico construido a partir de un conjunto reducido de ejes ortogonales. Por lo tanto, es una forma pasiva de análisis, y el usuario interpreta los resultados de la ordenación. Los siguientes métodos se incluyen en esta categoría: Análisis de Componentes Principales (ACP), Análisis de Correspondencias (AC), Análisis de Coordinadas Principales (ACoP), Análisis de Correspondencias Destendenciado (DCA, siglas en inglés) y escalamiento multidimensional no métrico (EMNM). Estos métodos son conocidos en el ámbito de la Ecología como *ordenación sin restricciones* (Borcard *et al.*, 2011).

Cuando se tiene, como en este caso, un conjunto de individuos (o sitios) a los cuales se les miden distintos conjuntos de variables, estamos ante la presencia de los denominados Datos de Conjuntos Múltiples (Kroonenberg, 2008). Aquí se pueden seguir dos tipos de análisis. Uno de ellos consiste en estudiar de forma simultánea las diversas matrices de variables (rol simétrico), para obtener resultados, especialmente gráficos, que permitan analizar la relación entre las observaciones, las variables y las matrices, como es el caso de las técnicas de Análisis Factorial Múltiple (AFM) o Análisis de Procrustes Generalizado (APG). La otra situación que se puede presentar es que una de las matrices de variables dependa de la otra (rol asimétrico) y el objetivo sea extraer gradientes de variación en variables dependientes explicadas por variables independientes, aquí encontramos los métodos de ordenación canónica como el Análisis Canónico de Redundancia (RDA, siglas en inglés), el Análisis de Redundancia basado en la distancia (db-RDA) y el Análisis de Correspondencia Canónica (ACC), técnicas conocidas en estudios de ecología como *ordenación restringida* (Borcard *et al.*, 2011).

En consecuencia, si se desea extraer estructuras de un conjunto de datos que están relacionadas con estructuras en otros conjuntos de datos, y/o probar

formalmente hipótesis estadísticas sobre la importancia de estas relaciones, la ordenación canónica podría ser la más adecuada.

Hoy en día, con el rápido desarrollo de la ecología comunitaria cuantitativa, los requisitos son crecientes entre los ecólogos, que generalmente necesitan emplear varios enfoques multivariados para analizar los efectos de múltiples factores ambientales simultáneamente, en estudios comunitarios con docenas de especies. Por lo tanto, los métodos comunes de ordenación como RDA, ACP, ACC y EMNM se han utilizado con frecuencia para relacionar especies (Guo *et al.*, 2015).

En esta tesis se busca comparar para el análisis de un mismo conjunto de datos, métodos como el AFM que trata de manera independiente a todas las matrices (rol simétrico) y dos métodos como el ACC y el RDA que asumen que una de las matrices representa variables de respuesta que dependen de otra matriz con variables explicativas (rol asimétrico).

Para utilizar de manera eficiente las herramientas numéricas disponibles, es esencial comprender claramente los principios que subyacen a los métodos numéricos y sus límites.

Asimismo, resulta de interés comparar la aplicación de dos *softwares*. El CANOCO (ter Braak, 1985), que es pago y resulta más amigable en su uso para sus usuarios, desarrollado específicamente para análisis multivariados de datos ecológicos, pero donde aún faltan algunos análisis multivariados del gran abanico existente. Y el R (R Core Team, 2019), que es un software de código libre y gratuito, que permite la implementación de una gran cantidad de funciones relacionadas al análisis estadístico de datos bajo una plataforma que requiere tener conocimientos básicos de programación.

Ambos softwares, son una herramienta potente y versátil que, a través del análisis de datos multivariados, brinda al investigador la capacidad de analizar aspectos ecológicos de la dinámica de poblaciones, de comunidades y de ecosistemas.

1.2. Objetivos

Objetivo general

Comparar distintos métodos multivariados empleando dos programas estadísticos en matrices de datos ecológicos, evaluando su aplicabilidad para ambientes acuáticos.

Objetivos específicos

1. Comparar los fundamentos matemáticos del AFM, ACC y RDA analizando ventajas y desventajas de los tres métodos.
2. Comparar los algoritmos que utilizan R y CANOCO para abordar estos métodos.
3. Evaluar la aplicación de estos métodos en biomonitoreos de ambientes acuáticos, estableciendo relaciones entre la abundancia de las especies/familias y las condiciones ambientales de los sitios de muestreo, comparando los resultados de cada método.

1.3. Hipótesis

Las técnicas disponibles para relacionar una matriz de frecuencias y un grupo de variables continuas que describen un mismo conjunto de individuos (AFM, RDA y ACC) no dan los mismos resultados, dependiendo esto de las características particulares de la base de datos, habiendo una más adecuada para cada situación.

CAPÍTULO 2. Marco Teórico

2.1. Conceptos generales de los métodos de ordenación

El término “ordenación” se emplea principalmente en ecología y se refiere a todas aquellas técnicas tendientes a ordenar grupos de objetos o individuos en un número relativamente bajo de dimensiones que refleje algún patrón de respuesta de ese conjunto de objetos. El objetivo de la ordenación es ayudar a generar hipótesis acerca de la relación entre individuos en función de una serie de factores que se observan simultáneamente a través de una reducción de la dimensionalidad de los datos con un mínimo de pérdida de información.

La ordenación busca un patrón continuo en los datos multivariados. El uso de la ordenación en ecología fue iniciado por el botánico y ecólogo australiano de origen inglés David Goodall, cuyo primer artículo que usa la ordenación (ACP) se publicó en 1954 (Goodall, 1954).

Legendre y Legendre (1998) mencionan que Gower (1984) señaló que el término ordenación, utilizado en estadística multivariada, en realidad proviene de la ecología, donde se refiere a la representación de objetos (sitios, estaciones, etc.) como puntos a lo largo de uno o varios ejes de referencia.

La suposición principal de la ordenación es que los datos analizados son redundantes, es decir, contienen más variables de las necesarias para describir la información detrás y se puede reducir el número de estas variables (y dimensiones) sin perder demasiada información. Por ejemplo, en el caso de los datos de composición de especies, a menudo algunas de las especies son ecológicamente similares (por ejemplo, especies que prefieren crecer en hábitat húmedo que seco), lo que significa que el conjunto de datos contiene varias variables redundantes (especies) que cuentan la misma historia. O, para explicar la redundancia de otra manera, a partir de la ocurrencia (o ausencia) de una especie, a menudo se puede predecir la ocurrencia (o ausencia) de varias otras especies (por ejemplo, si la muestra incluye especies de hábitats húmedos, se espera que las especies que prefieren hábitats secos no estén presentes).

El Análisis de Componentes Principales (ACP), el Análisis de Correspondencias (AC), el Análisis Discriminante (AD), el Escalado Multidimensional No Métrico (EMNM), se pueden utilizar para analizar datos sin variables explicativas, mientras que el Análisis de Correspondencias Canónico (ACC) y el Análisis Canónico de Redundancia (RDA) utilizan variables respuestas y explicativas (Zuur *et al.*, 2007).

En la siguiente figura se sintetizan los métodos de ordenación a abordar en este trabajo de tesis para el análisis de las matrices de datos:

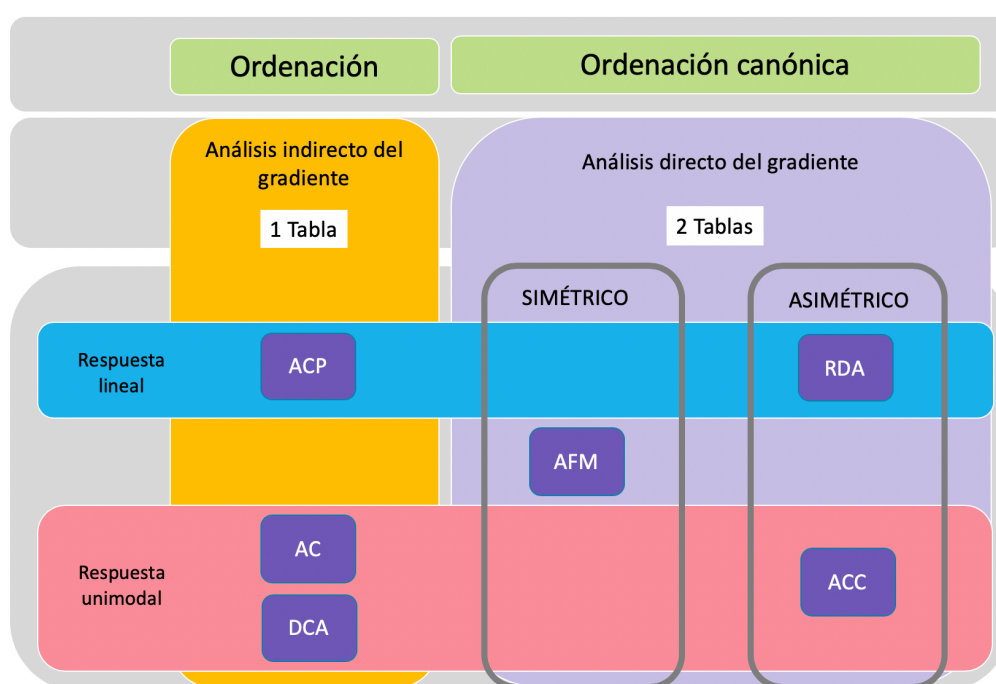


Figura 2.1. Esquema de clasificación de los métodos multivariados de ordenación expuestos en esta tesis según Borcard *et al.* (2011) y Legendre y Legendre (2012).

El análisis canónico es el análisis simultáneo de dos, o posiblemente varias matrices de datos. Los análisis canónicos permiten a los ecólogos realizar comparaciones directas de dos matrices de datos (también llamado "análisis de gradiente directo" (Legendre y Legendre, 2012). El AFM es un análisis simétrico directo del gradiente (Borcard *et al.*, 2011), mientras que el RDA, así como el ACC son dos formas de análisis canónico asimétrico directo del gradiente, ampliamente utilizado por ecólogos y paleoecólogos.

2.1.1. Análisis indirecto del gradiente (comparación indirecta)

El modo en que influyen las variables ambientales en la composición biótica de los ecosistemas es a menudo explorada de la siguiente forma: se utilizan técnicas de ordenación (como por ejemplo el Análisis de Componentes Principales (ACP), Análisis Factorial de Correspondencias (AC), métodos Biplot, etc.) para resumir y representar los datos de la composición de la comunidad en un diagrama de ordenación, el cual es posteriormente interpretado según la información ambiental de que se dispone mediante regresión múltiple de las variables ambientales sobre los ejes de ordenación. Esta aproximación en dos pasos es lo que se conoce como **análisis indirecto del gradiente**, en el sentido utilizado por Whittaker (1967), ya que la interpretación ambiental es por lo tanto indirecta (Fernández Gómez *et al.*, 1996; ter Braak, 1987).

En la ordenación sin restricciones (análisis indirecto del gradiente), los ejes de ordenación no están limitados por factores ambientales. El método tiene como objetivo descubrir los gradientes principales (direcciones de cambios) en los datos de composición de especies, y devuelve ejes de ordenación sin restricciones, que corresponden a las direcciones de mayor variabilidad dentro del conjunto de datos.

Opcionalmente, estos gradientes pueden ser *post hoc* (después del análisis) interpretados por variables ambientales (si están disponibles). Las variables ambientales no ingresan al algoritmo de ordenación. La ordenación sin restricciones es principalmente un método analítico exploratorio, utilizado para explorar el patrón en datos multivariados, genera hipótesis, pero no las prueba.

El análisis indirecto de gradiente utiliza solamente la matriz de abundancia de taxones (proporción de individuos de cada especie o taxón, en relación con el total de individuos que conforman la comunidad). Si hay alguna información sobre el entorno, se utiliza después del análisis de gradiente indirecto, como herramienta interpretativa. Es posible que los gradientes más importantes para una especie dada sean aquellos para los cuales no se tienen datos ambientales (por ejemplo, intensidad de perturbaciones pasadas), sin embargo, el análisis indirecto aprovechará la redundancia en el conjunto de datos y mostrará dichos gradientes.

2.1.2. Análisis directo del gradiente (comparación directa)

En 1948, Rao describió un nuevo método, Análisis de Variación Canónica. Esta expresión se refiere a la forma canónica de una matriz. La forma canónica es la forma más simple y comprensiva en la cual ciertas funciones relacionan expresiones que pueden ser reducidas sin pérdida de la generalidad. Por ejemplo, una forma canónica de la matriz de covarianza es la matriz de autovalores calculada de ella. La descomposición en valores singulares es un análisis canónico (Legendre y Legendre, 2012).

Dentro del **análisis directo de gradiente** (ordenación restringida o canónica), los ejes de ordenación están limitados por factores ambientales, esto significa que las variables ambientales (en una matriz **X**) participan en forma activa en la ordenación obtenida. Son métodos orientados a detectar patrones de variación en las especies (en una matriz **Y**), que pueden ser explicados en forma óptima por las variables ambientales consideradas. El efecto de estas variables externas sobre las especies se canaliza a través de los ejes factoriales (Legendre y Legendre, 2012). Los métodos canónicos permiten a los ecólogos realizar comparaciones de dos matrices sobre los mismos individuos u objetos (filas) (Legendre y Legendre, 2012)

Hay dos familias principales de métodos de ordenación canónica: simétrica y asimétrica.

a. Métodos simétricos

Los métodos simétricos se usan en los casos en que los dos conjuntos de datos juegan el mismo papel en el estudio. Estos métodos incluyen por ejemplo el Análisis de Correlación Canónica (ACorC), Análisis de Co-inercia (ACoI), Análisis de Procrustes (APG) y algunos otros (Legendre y Legendre, 2012).

Borcard *et al.* (2011) incorporan el AFM a estos métodos. En el AFM las dos o más matrices juegan el mismo rol en el análisis, no hay dependencia. En este trabajo, se utilizó la nomenclatura de **X** e **Y** para representar a cada una de las matrices, en lugar de **Y₁** e **Y₂** como sugiere la bibliografía.

b. Métodos asimétricos

En las formas asimétricas de análisis, hay un conjunto de datos de respuesta y un conjunto de datos explicativos. Métodos asimétricos son el Análisis de Redundancia (RDA) y el Análisis de Correspondencia Canónica (CCA) (Legendre y Legendre, 2012).

Las matrices tienen un rol **asimétrico**, o sea no juegan el mismo rol. Como en regresión lineal, hay una matriz de respuesta **Y** y una matriz explicativa **X**. La matriz **Y**, a menudo contiene datos de composición de la comunidad y la matriz **X**, datos de variables ambientales (Legendre *et al.*, 2011).

Desde su introducción (ter Braak, 1986), el Análisis de Correspondencias Canónico y el Análisis de Redundancia han sido los métodos de ordenación más populares en la ecología comunitaria (Oksanen *et al.*, 2019). El método generalmente se usa como análisis confirmatorio, es decir, es capaz de probar las hipótesis sobre la relación entre los factores ambientales en la composición de las especies. Permiten probar directamente las hipótesis ecológicas a priori (1) sacando a relucir toda la varianza de Y que está relacionada con X y (2) permitiendo que se realicen pruebas formales de estas hipótesis.

2.1.3. Pre-transformación de los datos

Datos ambientales

Diversos autores identifican la necesidad de pre-transformar los datos antes de los análisis.

Las variables ambientales de tipo físicas, químicas o geológicas a menudo se usan como variables explicativas en estudios ecológicos o de calidad de aguas. También se pueden usar directamente para obtener ordenaciones de los objetos o sitios sobre la base de estas variables.

Los datos deben examinarse inicialmente utilizando métodos univariados, como calcular estadísticos de asimetría o realizar histogramas de frecuencia. Dependiendo del tipo de asimetría encontrada, se pueden aplicar varias

transformaciones, como la raíz cuadrada, la raíz cuadrada doble o la transformación logarítmica. Los métodos generales, como las transformaciones de Box-Cox, están disponibles para encontrar automáticamente la transformación de normalización más eficiente (Sokal y Rohlf, 1995; Legendre y Legendre, 2012). Estas a menudo se denominan transformaciones normalizadoras, porque eliminar la asimetría es un paso importante para obtener datos distribuidos normalmente. Sin embargo, enfatizamos que el objetivo antes de la ordenación no es obtener una distribución multinormal de los datos, sino simplemente reducir la asimetría de las distribuciones. Las pruebas de normalidad pueden ser útiles para examinar los datos e identificar las variables cuyas distribuciones deben estudiarse más de cerca para encontrar, si es posible, una transformación que reduzca la asimetría.

Por lo general, los usuarios de los métodos de ordenación solo tienen criterios estadísticos, como el sesgo de las distribuciones, para decidir sobre la transformación de las variables físicas.

En la mayoría de los casos, las variables ambientales no se expresan en las mismas unidades físicas de medición, por ello generalmente se estandarizan por desvío estándar y rango.

Datos de composición de especies

La distancia Euclídea puede conducir a la siguiente paradoja cuando se usa como una medida de semejanza entre sitios basados en la abundancia de especies: los sitios sin ninguna especie en común pueden estar a una distancia menor que otros sitios que comparten especies. Esto sería incorrecto desde el punto de vista ecológico (Legendre y Legendre, 2012). La principal dificultad en ecología con respecto a la distancia Euclídea surge del hecho de que los métodos de ordenación utilizados con frecuencia, es decir, ACP y RDA, ordenan objetos en el espacio multidimensional de descriptores usando esta distancia.

No obstante, los datos de composición de ensamblaje (abundancia de especies) para gradientes ambientales cortos, que contienen relativamente pocos ceros, pueden ser ordenados por PCA o RDA, en ese caso, la distancia Euclídea es una medida significativa de la distancia ecológica entre las observaciones. Sin

embargo, estas variables pueden tener distribuciones asimétricas porque las especies tienden a tener crecimiento exponencial cuando las condiciones son favorables (He y Legendre, 1996, 2002).

Los datos de composición comunitaria muestreados a lo largo de gradientes ecológicos o ambientales largos generalmente contienen muchos valores cero porque se sabe que las especies tienen respuestas unimodales a lo largo de estos gradientes (ter Braak y Prentice, 1988). Esto se debe a que las especies tienen condiciones de nicho óptimas, donde se encuentran en mayor abundancia. Debido a que los métodos de ordenación utilizan una función de distancia como su métrica para posicionar los objetos uno con respecto al otro en el espacio de ordenación, es importante asegurarse de que la distancia elegida sea significativa para los objetos en estudio. Elegir una medida de distancia adecuada significa tratar de modelar las relaciones entre los sitios de manera adecuada para los datos de composición de ensamblaje disponibles. La elección de una medida de distancia es una decisión ecológica, no estadística (Legendre y Legendre, 2012).

Se pueden obtener biplots de especies en ordenación no restringida o restringida sin recurrir a la distancia Euclídea (utilizada en el ACP y el RDA) o la distancia Chi-cuadrado (preservada en el AC y ACC) que no siempre son apropiados para el análisis de los datos de composición de la comunidad.

Para lograr este objetivo, se proponen transformaciones para las matrices de datos de especies. Implican el uso de los datos de especies transformados para probar las relaciones con variables explicativas (es decir, variables ambientales o factores de un modelo de análisis de varianza multifactorial). Se pueden graficar biplots que muestren las relaciones de las especies con las variables explicativas (Legendre y Gallagher, 2001).

Legendre y Gallagher (2001) propusieron cinco transformaciones previas de los datos de especies. Estas transformaciones son: 1) Distancia de la Cuerda, 2) Métrica Chi-cuadrado, 3) Distancia Chi-cuadrado, 4) Distancia entre perfiles especie, 5) Distancia Hellinger. Cuatro de ellas están disponibles en **vegan** en la función *decostand* ().

Mientras que CANOCO presenta cinco opciones: 1) Logarítmica, 2) Raíz cuadrada, 3) Arcoseno (%), 4) Arcoseno (fracción), 5) Binarizar. No obstante, Leps y Smilauer (1999) sugieren que se importen los datos transformados a CANOCO si se requiere usar alguna transformación no disponible en el software.

De las transformaciones mencionadas, especialmente la distancia de Hellinger (Hellinger, 1909) es una medida recomendada para agrupar u ordenar datos de abundancia de especies (Rao, 1995; Legendre y Gallagher, 2001). Se define como:

$$D_{Hellinger}(x_1, x_2) = \sqrt{\sum_{j=1}^p \left[\sqrt{\frac{y_{1j}}{y_{1+}}} - \sqrt{\frac{y_{2j}}{y_{2+}}} \right]^2}$$

Si los datos y_{ij} son transformados en y'_{ij} de la siguiente manera: $y'_{ij} = \sqrt{\frac{y_{ij}}{y_{i+}}}$, entonces la distancia Euclídea entre los vectores fila de los datos transformados es idéntica a la distancia de Hellinger entre los vectores fila originales de la abundancia de especies. Esta distancia es asimétrica (insensible a dobles ceros) y su límite superior es $\sqrt{2}$.

La distancia de Hellinger es en realidad la distancia de la Cuerda calculada sobre frecuencias con transformación raíz cuadrada (Legendre y Legendre, 2012). Ambas transformaciones dan poco peso a las especies raras (Legendre y Gallagher, 2001).

Según Rao (1995), las ventajas de usar la distancia de Hellinger son:

- * la medida sólo depende de los perfiles del par en cuestión. No se altera cuando un conjunto externo de perfiles es considerado.
- * la medida no depende del tamaño de muestra sobre el cual los perfiles son estimados.

No obstante, Legendre y Gallagher (2001) afirman que no existe una demostración teórica que indique cual es la mejor transformación para utilizar.

2.1.4. Tipos de respuesta “lineal” y “unimodal”

Para los métodos basados en el análisis de matrices de especies con datos de abundancia o presencia/ausencia, se debe identificar el tipo de respuesta a lo largo del gradiente ambiental. Dentro de estos métodos, se reconocen tradicionalmente dos categorías:

Modelo Lineal: supone que la respuesta de la especie es lineal a lo largo del gradiente ambiental, lo que podría ser cierto para datos ecológicos bastante homogéneos, donde los gradientes ecológicos son cortos (Figura 2.2 (a) y Figura 2.3 (a)).

Modelo Unimodal: supone que la respuesta de la especie es unimodal a lo largo del gradiente, teniendo su óptimo en cierta posición del gradiente. Este modelo está más cerca de la realidad de los datos ecológicos y es más adecuado para conjuntos de datos heterogéneos (estructurados por un gradiente ecológico fuerte o largo, con alta rotación o cambio de especies y muchos ceros en la matriz de especies) con gradientes ambientales amplios (Figura 2.2 (b) y Figura 2.3 (b)).

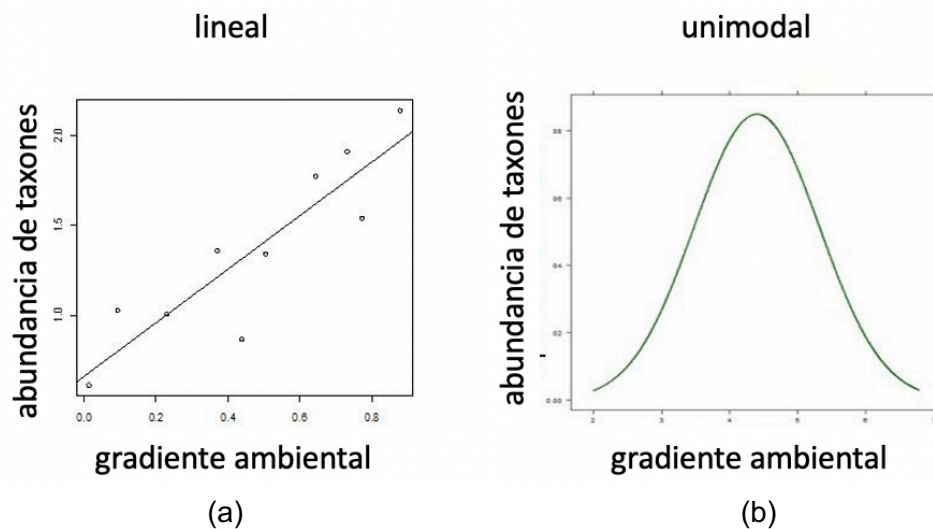


Figura 2.2. Respuesta lineal (a) versus unimodal (b) de aptitud/abundancia de taxones a lo largo de un gradiente ambiental.

La Figura 2.3 representa una comparación de la idoneidad de la aproximación lineal de una curva de respuesta unimodal. En el caso de que se muestree una

fracción más corta del gradiente ambiental (Figura 2.3a), se puede suponer que la respuesta de la especie (aunque fundamentalmente unimodal) puede modelarse como lineal. En el caso del gradiente largo, modelar la respuesta de la especie como lineal sería incorrecto (Figura 2.3b).

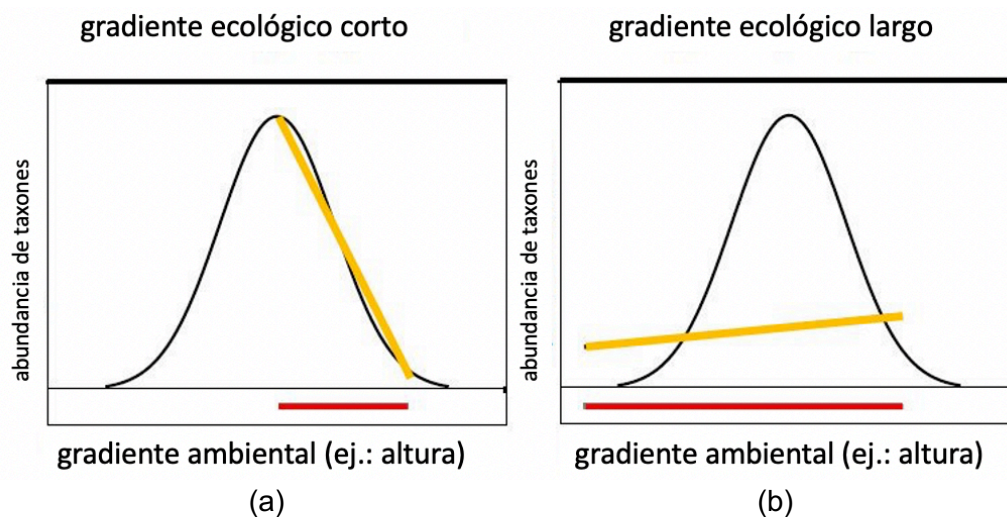


Figura 2.3. Respuesta hipotética de taxones a lo largo de un gradiente ambiental corto (a) y largo (b). La línea roja indica el segmento del gradiente realmente muestreado (parte del gradiente que nos interesa), y la línea amarilla indica cómo se vería la respuesta de la especie si se ajustara mediante un modelo lineal (Lepš y Šmilauer, 2003)

La diversidad de especies expresa la riqueza o el número de especies diferentes que están presentes en determinados ecosistemas o regiones. Whittaker (1967, 1969) relacionó el modelo unimodal con los niveles de diversidad, tres de los cuales se han convertido en centrales para la ecología comunitaria: la diversidad alfa es la diversidad (ya sea en términos de un índice de diversidad sintética o riqueza de especies) de una comunidad, y la diversidad gamma es la diversidad de una región o un paisaje. La **diversidad beta** puede verse como “la rotación de especies” (intercambio direccional de especies entre un par de muestras o a lo largo de un gradiente espacial, temporal o ambiental) o como una “variación” en la composición de especies (descripción no direccional de la heterogeneidad en la composición de especies dentro del conjunto de datos) (Anderson *et al.*, 2011).

Un gradiente con alta diversidad beta se considera un gradiente "largo" porque hay muchos cambios en la composición de las especies.

2.1.5. Matrices de datos

Las matrices para analizar se simbolizan de acuerdo con la nomenclatura de Legendre y Legendre (2012), como **Y** ($n \times p$) y **X** ($n \times m$), donde **Y** es una matriz o tabla de frecuencias, cuyas celdas expresan frecuencias absolutas o relativas (porcentajes), y **X** es una matriz de descriptores o variables continuas en la que las celdas son datos cuantitativos (Figura 2.4).

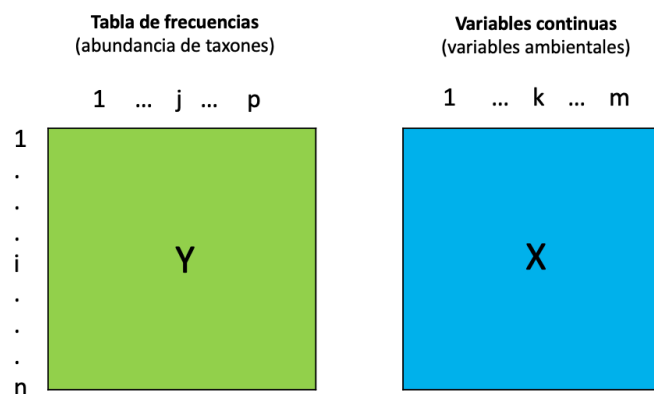


Figura 2.4. Matriz **Y** de frecuencias y **X** de variables continuas.

Las matrices **X** e **Y** tienen en común la información de los individuos en las filas, donde:

- individuos: $i = 1, 2, \dots, n$
- frecuencias: $j = 1, 2, \dots, p$
- variables continuas: $k = 1, 2, \dots, m$

La matriz de frecuencias **Y** es de dimensión $n \times p$ y de término general $[f_{ij}]$.

Las frecuencias absolutas están representadas por f_{ij} y las frecuencias relativas (probabilidades o proporciones) por p_{ij} . La matriz de frecuencias relativas asociada a la matriz **Y** se denota **Q**. La matriz de frecuencias relativas estandarizadas se denota \bar{Q} ($n \times p$) (Zuur *et al.*, 2007; Legendre y Legendre, 2012).

$$\bar{Q} = [\bar{q}_{ij}] = \left[\frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} \right] = \frac{(O_{ij} - E_{ij})/\sqrt{E_{ij}}}{\sqrt{f_{++}}}$$

Donde f_{++} es la suma de todas las frecuencias absolutas, y $p_{ij} = f_{ij}/f_{++}$ las frecuencias relativas. Con frecuencias relativas marginales de las filas $p_{i+} = f_{i+}/f_{++}$, y frecuencias relativas marginales de las columnas $p_{+j} = f_{+j}/f_{++}$.

Y el estadístico $(O_{ij} - E_{ij})/\sqrt{E_{ij}}$ en el numerador es llamado “componentes de Chi- cuadrado” en el análisis de tablas de contingencia.

La matriz de variables continuas \mathbf{X} es de dimensión $n \times m$ y de término general $[x_{ik}]$.

La nomenclatura de matrices y fórmulas de todo el documento de tesis es coincidente para todos los métodos con el fin de poder realizar comparaciones y deducciones.

2.2. Métodos de análisis indirecto del gradiente

2.2.1. Breve descripción de los métodos ACP, AC y DCA

2.2.1.1. ACP–Análisis de componentes principales (Hotelling, 1933; Pearson, 1901)

La técnica de componentes principales es debida a Hotelling (1933), aunque sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados introducidos por Pearson (1901), Peña (2002).

Dadas n observaciones de m variables, el análisis de componentes principales tiene como objetivo, analizar si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales (Peña, 2002). Estas nuevas variables se denominan componentes principales, son no correlacionadas entre si, son combinaciones lineales de las variables primitivas, y explican el máximo de la inercia o variabilidad total.

Con esta reducción de dimensión se simplifica el análisis de la dispersión de las observaciones poniéndose en evidencia posibles agrupamientos, permitiendo detectar las principales variables responsables de dicha dispersión y analizar las relaciones existentes entre ellas.

La distancia preservada en el ACP es la distancia Euclídea y las relaciones detectadas son lineales (Borcard et al., 2011).

Pasos de un ACP

a . Transformación de la matriz de datos

Sea $\mathbf{X}_{(n \times m)}$ una matriz de datos cuantitativos que recoge información de n individuos (objetos o unidades experimentales) y m variables (descriptores) (Figura 2.4).

La transformación de la matriz \mathbf{X} puede ser el centrado, esto es restarle la media, o la estandarización, es decir restar la media (centrado) y dividir por la desviación estándar (reducido).

Cuando las variables originales están en distintas unidades conviene aplicar el análisis de la matriz de correlaciones o análisis normado. Cuando las variables tienen las mismas unidades, ambas alternativas son posibles (usar la matriz de varianzas o la de correlaciones). Si las diferencias entre las varianzas de las variables son informativas y queremos tenerlas en cuenta en el análisis no debemos estandarizar las variables, por el contrario, si las diferencias de variabilidad no son relevantes se eliminan con el análisis normado (Cuadras, 1996; Peña, 2002).

b. Obtención de autovectores, autovalores y componentes principales

Las componentes principales son nuevas variables especificadas por los ejes de una rotación rígida del sistema original de coordenadas y correspondientes a las sucesivas direcciones de máxima varianza de la dispersión de puntos. Las componentes principales dan las posiciones de los objetos en el nuevo sistema de coordenadas (Legendre y Legendre, 2012).

Algebraicamente, las componentes principales serán una combinación lineal de las m variables aleatorias. Geométricamente, estas combinaciones lineales representaran un nuevo sistema de coordenadas obtenido por rotación del sistema original.

Las componentes principales son no correlacionadas y tienen varianzas iguales a los valores propios o autovalores de la matriz de covarianzas \mathbf{S} (para el caso del ACP en el que se utilice la matriz de varianzas).

A la varianza total se llama generalmente **inercia total** y es igual a la suma de los autovalores de la matriz de covarianzas.

El análisis de componentes principales también puede realizarse a partir de la matriz \mathbf{X} estandarizada, esto equivale a trabajar sobre la matriz de correlaciones \mathbf{R} en lugar de la matriz de covarianzas y se denomina Análisis de Componentes Principales Normado. Las componentes principales no coinciden en ambos casos, ni es posible pasar de uno a otro.

c. Relaciones de transición y proyección de la matriz original sobre el nuevo espacio de dimensión reducida

Las componentes principales son los autovectores normalizados de la **matriz de correlación entre individuos**. Tanto esa matriz, como la **matriz de correlación entre variables** son simétricas y tienen igual rango. Existen relaciones, entre los autovectores correspondientes a ambas matrices que permiten obtener los autovectores de una en función de los de la otra.

Estas relaciones se denominan **relaciones de transición**, de las cuales se deduce que las coordenadas de las variables centradas y estandarizadas son las correlaciones de las variables con las componentes principales.

Cuando se quieren considerar simultáneamente las proyecciones de los puntos individuos y variables sobre el mismo plano factorial, se suelen transformar las coordenadas de los individuos para que sobre cada eje se tenga una dispersión respecto al origen comparable para las observaciones y para variables.

Luego de obtener las coordenadas de los individuos dadas por los autovectores, se proyectan los individuos y las variables en el nuevo espacio. La nube de puntos-individuos está centrada en el origen, puesto que se han centrado los datos iniciales. No ocurre lo mismo con la nube de variables.

La posición relativa de los individuos en el espacio rotado m -dimensional es la misma que en el espacio m -dimensional original. Esto significa que la distancia Euclídea entre individuos se ha preservado con la rotación de los ejes.

El biplot es una herramienta de visualización para presentar los resultados del análisis de componentes principales. Hay dos opciones principales, el biplot de distancia y el de correlación (Zuur *et al.*, 2007; Legendre y Legendre, 2012).

d. Elementos de interpretación y algunas consideraciones del ACP

Habiendo obtenido las coordenadas de los vectores variables, su proyección sobre el plano factorial ayudará a interpretar las componentes principales y a analizar las relaciones entre variables a través del ángulo que estas forman, ya que estas variables centradas y estandarizadas serán vectores cuyos extremos se ubican sobre la esfera unidad.

Un plano factorial es una aproximación de la nube de puntos y como tal tendrá puntos bien representados, pero podrá contener puntos con mala calidad de proyección. Se utilizan índices complementarios que ayudan a la interpretación de estos puntos (filas y columnas) (Escofier y Pagès, 1992).

Los autovalores asociados a los ejes principales representan la cantidad de varianza del eje dado. Se puede calcular la proporción de varianza explicada por un eje dado como el autovalor del eje dividido por la varianza total. Si pocos ejes principales explican la mayor parte de la varianza, la ordenación fue exitosa (la información multidimensional se redujo con éxito a unas pocas dimensiones principales).

Las cargas de factores, también conocidas como cargas de las componentes son las correlaciones de las variables (especies o, en general, descriptores) con los ejes individuales del ACP. Si está estandarizado, las cargas factoriales se pueden

comparar entre variables y ayudan a interpretar qué descriptores están asociados principalmente con cada eje.

2.2.1.2. AC–Análisis factorial de Correspondencias (Benzécri, 1969)

Los cálculos en el Análisis de Componentes Principales (ACP) se realizan en una métrica Euclídea. A principios de la década de 1970, el francés Benzécri y sus compañeros de trabajo desarrollaron un método similar utilizando pesos o ponderaciones en una métrica de Chi-cuadrado (simbolizado como χ^2). Este método se llamó “*Analyses des correspondences*”, traducido por Hill (1974) en Análisis de Correspondencias. Una buena visión general, con aplicaciones, de este análisis ponderado de componentes principales se da en Greenacre (1983) y Zuur et al. (2007).

El objetivo del AC es encontrar la mejor representación simultánea de dos conjuntos constituidos por las filas y las columnas de una tabla de contingencia. Estudia las relaciones existentes entre dos o más variables cualitativas.

El AC, es un método de ordenación unimodal sin restricciones en el que la matriz de entrada corresponde a frecuencias. Una propiedad interesante, que atrajo a los ecólogos a este método, es el hecho de que se puede calcular y mostrar correspondencia entre individuos (muestras, sitios, etc.) y especies en el mismo espacio de ordenación. En el espacio de todos los ejes de ordenación, las distancias entre los individuos (y también entre las especies) están representadas por la métrica de distancia de Chi-cuadrado.

El AC es similar al ACP en el sentido que preserva en el espacio de los ejes principales (después de la rotación), la distancia Euclídea entre los perfiles de las probabilidades condicionales ponderadas. Esto es equivalente a preservar la distancia Chi-cuadrado entre las filas y las columnas de una tabla de contingencia.

En general, el Análisis de Correspondencias puede ser aplicado a matrices de datos que son dimensionalmente homogéneas y sólo contienen números positivos o ceros (Legendre y Legendre, 2012). Los datos deben ser enteros no negativos o presencias-ausencias.

El Análisis de Correspondencias puede ser también realizado sobre tablas de contingencia que comparan dos grupos de variables. Por ejemplo, las filas de la matriz podrían ser diferentes especies divididas en clases de abundancia y las columnas diferentes variables ambientales. No obstante, una mejor manera de comparar las especies con los datos ambientales es el Análisis de Correspondencias Canónico (ACC).

Pasos de un AC

El Análisis de Correspondencias puede describirse en tres pasos (Legendre y Legendre, 2012):

- La tabla de contingencia se transforma en una matriz de contribuciones al estadístico Chi-cuadrado de Pearson, después de ajustar el modelo nulo de una tabla de contingencia.
- La descomposición en valores singulares se aplica a esa matriz y los autovalores y autovectores son calculados como en el ACP.
- Se realizan otras operaciones matriciales para producir las matrices necesarias para graficar los diagramas.

a. Descomposición en valores singulares

La descomposición en valores singulares (DVS) es uno de los posibles algoritmos para realizar los cálculos en el ACP y AC (Legendre y Legendre, 2012).

La DVS es aplicada a la matriz $\bar{\mathbf{Q}}$ con el siguiente resultado:

$$\bar{\mathbf{Q}}_{(n \times p)} = \hat{\mathbf{U}}_{(n \times n)} \mathbf{W}_{(p \times p)} \mathbf{U}'_{(p \times p)}$$

Donde \mathbf{U} y $\hat{\mathbf{U}}$ son las matrices que contienen los vectores columna que están normalizados y son ortogonales entre ellos. \mathbf{W} es una matriz diagonal $\mathbf{D}_{(w_i)}$, siendo los valores w_i en \mathbf{W} , no negativos y los valores singulares de $\bar{\mathbf{Q}}$.

Previo a la descomposición de autovalores debe ser calculada una matriz de suma de cuadrados y productos cruzados $\bar{\mathbf{Q}}'\bar{\mathbf{Q}}$, que puede ser llamada matriz de

covarianza en un sentido general. Los autovalores y autovectores de $\bar{\mathbf{Q}}'\bar{\mathbf{Q}}$ son calculados mediante la ecuación:

$$(\bar{\mathbf{Q}}'\bar{\mathbf{Q}} - \lambda_k \mathbf{I})\mathbf{u}_k = \mathbf{0}$$

$\mathbf{W}'\mathbf{W} = \mathbf{\Lambda}_{(p \times p)}$ es la matriz diagonal de autovalores de $\bar{\mathbf{Q}}'\bar{\mathbf{Q}}$.

$\mathbf{U}_{(p \times p)}$ es la matriz de autovectores de $\bar{\mathbf{Q}}'\bar{\mathbf{Q}}_{(p \times p)}$, conteniendo las cargas (“loadings” en inglés) de las columnas de la tabla de contingencia.

$\hat{\mathbf{U}}_{(n \times p)} = \bar{\mathbf{Q}}\mathbf{U}\mathbf{\Lambda}^{-1/2}$ es la matriz de autovectores de $\bar{\mathbf{Q}}\bar{\mathbf{Q}}'_{(n \times n)}$

b. Inercia

La inercia total (o varianza total) en AC se define como el estadístico Chi-cuadrado de la matriz de sitio por especie, dividida por el número total de observaciones. Los puntos alejados del origen en cada diagrama son los más interesantes, porque estos puntos hacen una contribución relativamente mayor al estadístico de Chi-cuadrado que los puntos más cercanos al origen. Entonces, cuanto más lejos del origen se traza un sitio, más diferente es del sitio promedio (Zuur *et al.*, 2007).

c. Gráficos

En AC, la posición de una especie representa el valor óptimo en términos del modelo de respuesta Gaussiana (nicho) a lo largo del primer y segundo eje. Por esta razón, la mayoría de los paquetes de software ecológico presentan los puntajes de las especies como un punto o etiqueta, y no como una línea (Zuur *et al.*, 2007).

Las matrices \mathbf{U} y $\hat{\mathbf{U}}$ pueden ser usadas para graficar las posiciones de los vectores filas y columnas en dos diagramas de ordenación separados.

Las matrices \mathbf{U} y $\hat{\mathbf{U}}$ pueden ser ponderadas por la inversa de la raíz cuadrada de las coordenadas (“scores” en inglés) de la filas y columnas obteniéndose así las matrices diagonales de pesos:

$$\mathbf{D}(p_{+j})^{-1/2} \quad (p \times p) \quad \mathbf{D}(p_{i+})^{-1/2} \quad (n \times n)$$

Las relaciones de transición expresan los resultados de un análisis factorial, por ejemplo, los del espacio fila en función del espacio columna y recíprocamente, los del espacio columna en función del espacio fila. Una vez que se han realizado las transiciones, las interpretaciones de los planos factoriales que representan deben hacerse conjuntamente.

2.2.1.3. DCA – Análisis de Correspondencias Destendenciado o por segmentos

El Análisis de Correspondencias (AC) puede presentar el problema de crear efecto de arco en los diagramas de ordenación, lo cual es causado por una correlación no lineal en el plano principal (Figura 2.5).

Los gradientes ambientales o temporales a menudo apoyan la sucesión espacial o temporal de especies. Dado que las especies que están controladas por factores ambientales (en comparación con la dinámica de la población, eventos históricos, etc.) generalmente tienen distribuciones unimodales a lo largo de los gradientes, el efecto de los gradientes en las relaciones de distancia entre los sitios, calculada sobre la presencia de especies (datos de presencia/ausencia o abundancia), es necesariamente no lineal (Legendre y Legendre, 2012).

El Análisis de Correspondencias (AC) es un método más robusto para la ordenación comunitaria que el Análisis de Componentes Principales (ACP). Sin embargo, con largos gradientes ecológicos sufre algunos inconvenientes o “fallas”.

El AC adolece de dos problemas principales: el efecto de arco (causado por las curvas de respuesta de especies unimodales) y la compresión de los extremos del gradiente. Debido al primer problema, el segundo eje AC no se puede interpretar fácilmente. Debido al segundo problema, la separación de muestras (y especies) a lo largo del primer eje no está necesariamente relacionada con la cantidad de cambio (o diversidad beta) a lo largo del gradiente primario.

Hill y Gauch (1980) desarrollaron un nuevo método de ordenación, el Análisis de Correspondencias Destendenciado (DCA), para corregir los principales problemas

del AC. Estos problemas se corrigen en dos fases: por reducción de tendencia y por reescalamiento.

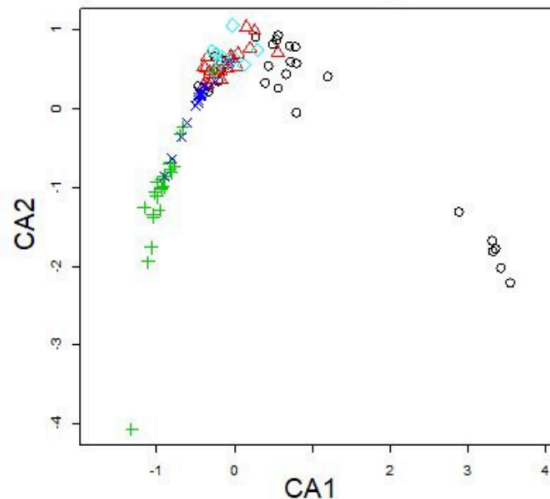


Figura 2.5. Ejemplo de efecto de arco como resultado de la ordenación de AC. Las muestras del mismo color y símbolo pertenecen al mismo tipo de comunidad (tipo de vegetación en el caso de este conjunto de datos). Fuente:

<https://www.davidzeleny.net/anadat-r/doku.php/en:start>

El DCA es un método de ordenación de aproximación matricial (aproxima lo mejor posible a la matriz de datos original) para detectar la estructura latente de la matriz de frecuencias \mathbf{Y} . Estima la heterogeneidad en la composición de la comunidad, solo basándose en la información provista por las especies. La información ambiental se anexa con posterioridad a la ordenación a través de correlaciones entre variables ambientales y los ejes de ordenación, lo que lo convierte en un análisis indirecto del gradiente.

La reducción de tendencia es el proceso de eliminar el efecto de arco (ter Braak y Šmilauer, 1998). DCA hace esto dividiendo el primer eje en segmentos y luego centrando el segundo eje en cero. El reescalado es el proceso de cambiar las posiciones de las muestras a lo largo de los ejes de ordenación para que la diversidad beta sea constante. Es necesario reescalar para que una distancia dada en el espacio de ordenación signifique lo mismo en diferentes partes del diagrama de ordenación, o en diferentes diagramas de ordenación.

Después de la reducción de tendencia por segmentos y el reescalado no lineal de los ejes, la ordenación DCA tiene la interesante propiedad de que los ejes se escalan en unidades de la desviación estándar promedio (DE) de la rotación de especies (“*turnover*” en inglés, o tasa de extinción y aparición de especies) (Gauch, 1982). A lo largo de un gradiente regular, aparece una especie, se eleva a su valor modal y desaparece en una distancia de aproximadamente 4 DE; de manera similar, se produce una renovación completa en la composición de especies, sobre los sitios, en aproximadamente 4 unidades DE. Es decir, hay un cambio medio en la composición de especies dentro de aproximadamente 1 a 1,4 unidades DE. Por lo tanto, la longitud del primer eje DCA es una medida aproximada de la longitud del gradiente ecológico, medida en unidades de rotación de especies. Es por esto, que el DCA con reescalado no lineal de los ejes es un método útil para estimar las longitudes de los gradientes ecológicos (Legendre y Legendre, 2012).

Lepš y Šmilauer (2003) propusieron la utilización del DCA para decidir si aplica el método de ordenación lineal o unimodal en los datos. Primero se calcula el DCA sobre la matriz **Y**, verificando la longitud de los primeros ejes DCA (escalado en unidades de desviación estándar, DE). La longitud del primer eje DCA > 4 DE, indica un conjunto de datos heterogéneo en el que se deben utilizar métodos unimodales (ACC), mientras que la longitud < 3 DE indica un conjunto de datos homogéneo para el cual los métodos lineales son adecuados (RDA). En la zona entre 3 y 4 DE, los métodos lineales y unimodales pueden ser utilizados sin distinción.

Si bien los métodos lineales no deberían usarse para datos heterogéneos, los métodos unimodales pueden ser utilizados para datos homogéneos, pero los métodos lineales, en este caso, son más potentes y deberían preferirse. No obstante, el método es criticado y algunos de los investigadores no lo recomiendan (Borcard *et al.*, 2011; Legendre y Legendre, 2012), mientras que otros lo defienden (ter Braak y Šmilauer, 2015).

El DCA es utilizado en esta tesis para decidir sobre el uso del método RDA o ACC.

2.3. Métodos de análisis directo del gradiente

2.3.1. AFM – Análisis Factorial Múltiple (Escofier y Pagès, 1984, 1990)

El Análisis Factorial Múltiple (AFM), desarrollado por Escofier y Pagès (1984,1990), en el seno de la Escuela Francesa de Análisis de Datos, es un método factorial adaptado al tratamiento de matrices de datos en las que un mismo conjunto de individuos se describe a través de varios grupos de variables.

Los grupos de variables pueden ser más de dos y de diferente naturaleza, cuantitativas y cualitativas, o del manejo de un mismo conjunto de variables medidas en distintos periodos de tiempo. La única restricción que impone el AFM, es que las variables sean del mismo tipo matemático (cuantitativa o cualitativa) dentro de cada matriz.

El AFM realiza un análisis simétrico de varias matrices de datos resumiendo y visualizando toda esa información en un único gráfico donde se tiene en cuenta la contribución de todos los grupos activos de variables para definir la distancia entre los individuos. Es un método sumamente útil para explorar relaciones entre varios grupos de variables ecológicamente significativas sin importar su número o tipo.

Abdessemed y Escofier (1996) presentan el AFM para el caso de una tabla de contingencia y un grupo de variables numéricas. Consiste en la transformación de la matriz de frecuencias y la introducción de métricas en filas y columnas de los dos grupos.

Para este trabajo, la información de partida del AFM está constituida por las matrices **X** e **Y** (Figura 2.4) y se desarrolla, en forma general, en dos etapas:

2.3.1.1. Pasos de un AFM

a. Primera etapa. Análisis parcial

En una primera etapa se analiza cada matriz por separado, que está centrada o estandarizada.

Se realiza un Análisis de Correspondencias Simple (AC) para el grupo de frecuencias \mathbf{Y} , y un Análisis de Componentes Principales (ACP) para el grupo de variables continuas \mathbf{X} .

En el AC, las filas y columnas se tratan con simetría completa y se representan por sus perfiles. Para estudiar la matriz de frecuencias y la matriz de variables continuas juntas, es deseable mantener el punto de vista del AC para la matriz de frecuencias y, por lo tanto, mantener las distancias Chi-cuadrado entre filas o columnas, para poder restaurar la estructura.

Como el AFM es un análisis de componentes principales (ACP) particular, la técnica de generalización se basa en la equivalencia entre el AC de la matriz de frecuencias \mathbf{Y} y el ACP no normado de una matriz transformada \mathbf{F}' , sus filas y columnas están provistas de métricas adecuadas. En efecto, es equivalente a realizar el AC de la matriz de términos general $[f_{ij}]$ o el ACP no normado de la matriz transformada de término general $[f_{ij}/f_{i+}f_{+j}]$, los individuos están provistos de los pesos f_{i+} y las variables de los pesos f_{+j} (Abdessemed y Escofier 1996).

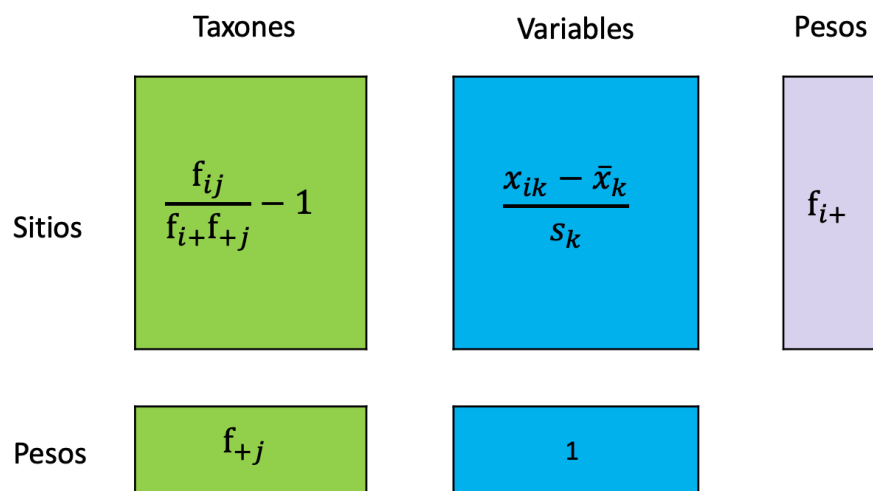


Figura 2.6. Matriz de frecuencias \mathbf{F}' y matriz descriptiva \mathbf{X} centrada por los pesos f_{i+} . Abdessemed y Escofier (1996).

Por lo tanto, esta equivalencia permite aplicar el AFM a la matriz de la Figura 2.6 compuesta por dos grupos de variables numéricas definidas por \mathbf{F}' y \mathbf{X}_{est} . La matriz \mathbf{F}' está centrada y tiene el término general $\left[\frac{f_{ij}}{f_{i+}f_{+j}} - 1 \right]$ y la matriz \mathbf{X} está centrada y reducida por los pesos f_{i+} .

Se denota como \bar{x}_k y s_k a la media y a la desviación estándar de la variable k calculada con los pesos f_{i+} . El término general de la matriz \mathbf{X}_{est} es $[(x_{ik} - \bar{x}_k)/s_k]$. Los sitios i tienen pesos f_{i+} y las especies tienen peso f_{+j} .

Se denota $\mathbf{D} = \text{diag}(f_{i+})$ y $\mathbf{M} = \text{diag}(f_{+j})$ a las métricas inducidas por los pesos f_{i+} y f_{+j} en los espacios en los que están ubicados respectivamente las variables y los sitios. Las dos nubes de sitios se han centrado para eliminar la influencia de la posición de los dos baricentros y solo tienen en cuenta la forma de las nubes. Esto hace posible interpretar el coseno entre dos vectores que representan las variables de \mathbf{X} como coeficiente de correlación, con la métrica \mathbf{D} . Las métricas \mathbf{M} y \mathbf{D} tienen el efecto de preservar la estructura AC de la matriz de frecuencias inicial \mathbf{Y} .

La matriz de variables descriptivas solo está influenciada por la métrica \mathbf{D} , que da la distancia desde un sitio hasta el origen, una importancia proporcional a su tamaño expresado en \mathbf{Y} , y en particular influye en la conexión entre dos variables, ya que esto se expresa a través de la correlación.

Se denota λ_1^p el primer autovalor asociado al AC de la matriz \mathbf{Y} o del ACP no normado de \mathbf{F}' , y λ_1^m al primer autovalor asociado al ACP de la matriz \mathbf{X} . Cada matriz centrada se pondera dividiendo las variables por el primer autovalor obtenido del ACP o AC, para darles el mismo peso en el Análisis global, teniendo en cuenta las diferentes variaciones entre los grupos.

Como todos los grupos de variables están definidos sobre el mismo conjunto de individuos, las matrices pueden ser concatenadas una al lado de la otra (yuxtapuestas horizontalmente) y formar una única matriz de individuos por variables.

b. Segunda Etapa. Análisis global

En la segunda etapa se concatenan las columnas de las matrices ponderadas y la matriz yuxtapuesta es sometida a un ACP global.

El AFM es como un ACP ponderado de la matriz yuxtapuesta:

$$\begin{array}{c}
 1 \\
 \vdots \\
 i \\
 \vdots \\
 n
 \end{array}
 \begin{array}{c}
 \mathbf{Y} \\
 \frac{\lambda_1^p}{\lambda_1^m}
 \end{array}
 \dots
 \begin{array}{c}
 \mathbf{X} \\
 \frac{\lambda_1^m}{\lambda_1^p}
 \end{array}$$

Figura 2.7. Matriz yuxtapuesta del AFM.

Los diferentes subconjuntos de variables se proyectan juntas en el plano de ordenación global. Las estructuras comunes y las diferencias se evalúan para los individuos y variables.

Los dos grupos de variables se representan en el espacio de los individuos por su coordenada sobre el eje factorial del AFM, que es la inercia de la proyección de cada grupo sobre el factor principal del AFM (Morison, 2010).

Las inercias de las dos nubes en el AFM son iguales y su valor es la suma de la inercia de cada grupo, dividido por su primer autovalor.

Si el primer valor propio del AFM es cercano al máximo ($2 =$ número de grupos en este caso), indica una estructura común a los grupos. Si la coordenada de cada grupo de variables en cada eje factorial es cercana al máximo (1), la estructura del grupo es más fuerte y su influencia será determinante en la construcción del primer factor del AFM (Escofier y Pagès, 1984; Morison 2010).

2.3.1.2. Gráficos del AFM

En el AFM, se analizan tres tipos de objetos: individuos, variables y grupos de variables.

Individuos

Aquellos individuos cuyos puntos parciales o análogos (puntos que representan a cada individuo desde los diferentes grupos) se sitúen próximos, ilustran una misma posición relativa en los dos grupos analizados. Es la representación gráfica en un mismo espacio, de los individuos caracterizados por todas las variables (nube global media) y por cada uno de los grupos (nubes parciales).

Se busca una representación gráfica de la nube de individuos caracterizada por el conjunto de variables, y una representación superpuesta de las dos nubes de sitios caracterizada por la distribución de taxones y por variables ambientales.

La nube de los sitios descritos por los taxones y por las variables ambientales, es la "reunión" de dos sub-nubes, centradas y corresponde a las filas de la matriz $F'UX$. El cuadrado de la distancia entre dos sitios se escribe como la suma de los cuadrados ponderados de las distancias entre los sitios en cada una de las sub-nubes.

Variables

Las variables (taxones y variables ambientales) asociadas con los dos grupos están ubicadas en R^n , y por lo tanto generan dos nubes de puntos, correspondientes a los taxones y las variables descriptivas, respectivamente y una tercera nube que es la unión de las dos primeras.

La nube de taxones está centrada y la interpretación para esta nube de variables es análoga a la de un AC, ya que la distancia entre dos taxones es la distancia Chi-cuadrado y el enlace se expresa en un eje por su proximidad en el sentido euclídeo.

La nube de variables descriptivas no está centrada y se interpreta como en el ACP mediante las conexiones en un plano, utilizando el círculo de correlaciones.

La nube de taxones y variables descriptivas, ubicada en R^n tampoco está centrada, ya que es la unión de dos sub-nubes, una de las cuales no está centrada, y corresponde a las columnas de la matriz $F'UX$.

El vínculo entre un taxón j y una variable k se mide por la covarianza entre el perfil del taxón y la variable. Se interpretan utilizando las fórmulas de transición.

La coordenada de una variable k sobre un factor del AFM representa la correlación entre esta variable y el factor, de la misma manera que en el ACP clásico. La contribución de cada variable a un eje s sirve para seleccionar las variables continuas que dan más significado al eje (Morison, 2010).

Grupos de variables

Representación gráfica de una nube de dos puntos que representa los dos grupos sobre los ejes factoriales de las nubes de individuos y de variables.

Las coordenadas de los grupos toman valores entre 0 y 1, la representación de los grupos muestra cuales son similares (o diferentes) según el punto de vista de los factores del Análisis global, la suma de las coordenadas de los grupos en cada eje es igual al autovalor en el AFM, la contribución de cada grupo al eje es igual a la coordenada del grupo dividida por la suma de las coordenadas. El estudio para los grupos se completa con la calidad de representación de cada grupo ubicada sobre el primer cuadrante del plano factorial 1-2 (Morison, 2010).

Fórmulas de transición del AFM F'UX (Abdessemed y Escofier, 1996)

La necesidad de una interpretación conjunta de las representaciones de individuos y de variables lleva a superponer ambas, con el problema de representar sobre un mismo gráfico puntos de diferente naturaleza, que tienen lugar en espacios diferentes. Las fórmulas de transición relacionan la coordenada de un individuo sobre un eje con el conjunto de las coordenadas de todas las variables en el eje de ese mismo rango. No se puede interpretar la posición de un individuo en función de una sola variable (y recíprocamente) (Escofier y Pagès, 1992).

En el análisis global los sitios tienen los pesos f_{i+} , los taxones f_{+j}/λ_1^p y las variables explicativas $1/\lambda_1^m$.

Variable explicativa $H_s(k) = \frac{1}{\sqrt{Y_s}} \sum_{k=1}^m f_{i+} \frac{x_{ik}}{s_k} F_s(i)$

$$\text{Taxón} \quad G_s(j) = \frac{1}{\sqrt{\gamma_s}} \sum_{j=1}^p \frac{f_{ij}}{f_{+j}} F_s(i)$$

$$\text{Sitio} \quad F_s(i) = \frac{1}{\sqrt{\gamma_s}} \sum_{j=1}^p \frac{f_j}{\lambda_1^p} \left(\frac{f_{ij}}{f_{+j}} - 1 \right) G_s(j) + \sum_{k=1}^m \frac{1}{\lambda_1^m} \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right) H_s(k)$$

γ_s es el autovalor de rango s , $F_s(i)$ el factor de orden s para el sitio i , $G_s(j)$ el factor de orden s para la especie j , $H_s(k)$ el factor de orden s para la variable explicativa k .

Por lo tanto, en la representación superpuesta las coordenadas factoriales para los individuos i^1 (grupo de frecuencias) y i^2 (grupo de variables continuas) sobre el eje s obtenido en el ACP global son:

$$F_s(i^1) = \frac{1}{\lambda_1^p \sqrt{\gamma_s}} \sum_{j=1}^p \frac{f_{+j}}{f_{i+}} G_s(j) \quad \text{y} \quad F_s(i^2) = \frac{1}{\lambda_1^m \sqrt{\gamma_s}} \sum_{k=1}^m \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right) H_s(k)$$

λ_1^p , es el primer autovalor asociado al AC de la matriz de frecuencias y λ_1^m , es el primer autovalor asociado al ACP ponderado de la matriz de variables continuas.

2.3.1.3. Elementos de interpretación

El parecido entre las dos nubes parciales se puede evaluar globalmente mediante las siguientes ayudas adicionales:

Coefficiente Lg. El coeficiente Lg está basado en la inercia y puede utilizarse como: una medida de relación entre cada grupo y las dimensiones del AFM; una medida de relación entre los grupos; y un indicador de la dimensionalidad de los grupos. Lg es de alguna manera el "número de dimensiones comunes", cada una "ponderada" por su inercia (Pagès, 2014).

Por ejemplo, para la relación del primer componente principal del AFM y cada grupo, $Lg=0$ indica que las variables del grupo no están correlacionadas con el primer componente principal del AFM, y $Lg=1$ indica que el componente principal del AFM es el mismo que el primer componente principal del grupo.

El coeficiente Lg calculado entre un grupo y sí mismo, se puede calcular la dimensionalidad de dicho grupo. Si Lg es cercano a 1, indica que ese grupo es casi unidimensional, si Lg es cercano a 0, indica que varias dimensiones de variabilidad son importantes y por lo tanto el grupo es multidimensional.

Como desventaja, este coeficiente toma valores desde 0 y no tiene límite superior (Pagès, 2014).

Coefficiente RV (Escoufier, 1973). El coeficiente RV es una medida de relación entre dos conjuntos de variables y se basa en el principio de que dos conjuntos de variables están perfectamente correlacionados si existe una transformación ortogonal que haga coincidir los dos conjuntos (Josse *et al.*, 2008). RV es una generalización multivariada del coeficiente de correlación de Pearson (Borcard *et al.*, 2011).

Este coeficiente permite cuantificar la asociación entre dos matrices y toma valores entre cero (las configuraciones son ortogonales) y uno (la distancia entre las configuraciones es cero) (Medico *et al.*, 2019). El coeficiente RV refleja la cantidad de varianza compartida por las dos matrices (Abdi *et al.*, 2013).

Si \mathbf{X} e \mathbf{Y} están centradas por columnas, y con $\|A\| = \sqrt{\text{traza}(AA^T)}$ como la norma asociada, el coeficiente RV se define por:

$$RV(\mathbf{X}, \mathbf{Y}) = \frac{\text{traza}(\mathbf{XX}^T \mathbf{YY}^T)}{\sqrt{(\text{traza}(\mathbf{XX}^T))^2 \cdot \text{traza}(\mathbf{YY}^T)^2}} = \frac{\text{traza}(\mathbf{S}_{\mathbf{XY}} \mathbf{S}_{\mathbf{YX}})}{\sqrt{\text{traza}(\mathbf{S}_{\mathbf{XX}}^2) \cdot \text{traza}(\mathbf{S}_{\mathbf{YY}}^2)}}$$

Donde: \mathbf{XX}^T y \mathbf{YY}^T son las matrices de productos internos de \mathbf{X} e \mathbf{Y} , $\mathbf{S}_{\mathbf{XX}} = (1/(n - 1))\mathbf{XX}^T$ es la matriz de covarianzas de \mathbf{X} y $\mathbf{S}_{\mathbf{YY}} = (1/(n - 1))\mathbf{YY}^T$ es la matriz de covarianzas de \mathbf{Y} , y $\mathbf{S}_{\mathbf{XY}} = (1/(n - 1))\mathbf{X}^T \mathbf{Y}$ la covarianza entre \mathbf{X} e \mathbf{Y} (Josse *et al.*, 2008).

Al igual que la medición Lg, RV vale 0 si, y solo si, cada variable de un grupo no está correlacionado con todas las variables del otro grupo. Sin embargo, el coeficiente RV es siempre menor o igual a 1, valor que se obtiene cuando las

nubes de individuos asociados a cada uno de los grupos son homotéticas (en este caso no interviene la dimensión de las nubes de individuos) (Pagès, 2014).

Coefficiente de correlación entre grupos y factores del AFM. Mide la correlación entre las variables canónicas (proyección de los factores parciales obtenidos en el análisis individual de cada grupo sobre los ejes del análisis global) y los factores del análisis global del AFM.

Asimismo, se pueden evaluar la contribución y calidad de representación tanto de individuos como de grupos a través de:

Contribución (“Contr”): es una medida de cuánto contribuye el individuo o grupo a la construcción de cada eje del AFM.

Coseno cuadrado (“Cos²”): es la calidad de representación del individuo o grupo en el subespacio. Coseno cuadrado entre el punto y su proyección.

2.3.2. RDA – Análisis de Redundancia (Rao, 1964; Wollenberg, 1977)

El Análisis de Redundancia (RDA) es una extensión del ACP que modela explícitamente las variables respuesta como una función de las variables explicativas (Zuur *et al.*, 2007). Variables de respuesta y variables explicativas son denominaciones frecuentemente utilizadas para este método, y para este trabajo se encuentran definidas por las matrices \mathbf{Y} ($n \times p$) y \mathbf{X} ($n \times m$), respectivamente (Figura 2.4).

Wollenberg (1977) fue quien llamó al método “Análisis de redundancia” por referencia al índice de redundancia de Stewart y Love (1968), que es la proporción de la varianza de la matriz de datos de respuesta \mathbf{Y} que se explica por la matriz explicativa \mathbf{X} (Legendre y Legendre, 2012).

El RDA (Wollenberg, 1977) está diseñado para analizar una relación direccional entre dos conjuntos de datos multivariados. Es una forma de ordenación restringida que examina cuánto de la variación en un conjunto de variables explica la variación en otro conjunto de variables.

Conceptualmente, el RDA es una regresión lineal múltiple multivariada seguida de un ACP de la matriz de valores ajustados (Borcard *et al.*, 2011). Debido a su direccionalidad, el RDA siempre se considera la técnica analítica más apropiada cuando la relación esperada entre variables dependientes e independientes es lineal (o relación direccional). Si la relación esperada es gaussiana (no direccional), entonces el ACC es más apropiado (Lambert *et al.*, 1988, Guo *et al.*, 2015).

El RDA es recomendado cuando la matriz de respuesta, matriz de frecuencias \mathbf{Y} , que contiene la abundancia de taxones, puede ser analizada sola por un ACP. En otras palabras, cuando las variables \mathbf{y} están linealmente relacionadas entre ellas y la distancia Euclídea es apropiada para describir las relaciones entre objetos en un espacio factorial (Legendre y Legendre, 2012).

Como el RDA es básicamente una serie de pasos de la regresión múltiple se necesitan más observaciones que variables explicativas (Zuur *et al.*, 2007).

2.3.2.1. Pasos de un RDA

Legendre y Legendre (2012) resumen el análisis de redundancia (RDA) de una matriz \mathbf{Y} ($n \times p$) por una matriz \mathbf{X} ($n \times m$) en dos pasos, primero una regresión multivariada y luego un ACP de los valores ajustados, más un paso optativo de análisis de residuales.

Las columnas de las matrices \mathbf{X} e \mathbf{Y} se deben centrar para tener medias iguales a cero. La matriz de especies \mathbf{Y} solo se estandariza si están en distintas unidades.

a. Regresión multivariada

1) El primer paso del RDA es entonces una regresión multivariada de \mathbf{Y} en \mathbf{X} .

La matriz de coeficientes \mathbf{B} ($m \times p$) se calcula como $\mathbf{B} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Y}$, donde \mathbf{B} ($m \times p$) es la matriz de los coeficientes de regresión las p variables de \mathbf{Y} sobre las m variables de \mathbf{X} .

Dado que las matrices \mathbf{X} e \mathbf{Y} están centradas sobre sus respectivas medias, no hay parámetros de ordenada al origen en los vectores columna de los coeficientes

de regresión que forman la matriz \mathbf{B} , y los vectores columna $\hat{\mathbf{Y}}$ también están centrados.

Los valores ajustados $[\hat{y}]$ pueden ser calculados por la operación matricial $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{B}$

Reemplazando \mathbf{B} , por último, se tiene la matriz de valores ajustados $\hat{\mathbf{Y}}$ a través de la ecuación lineal:

$$\hat{\mathbf{Y}} = \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Y}$$

Esto es equivalente a realizar una serie de regresiones lineales múltiples de las variables individuales de \mathbf{y} sobre \mathbf{x} , donde se van calculando los vectores columna de valores ajustados o predichos que luego se unen para formar la matriz $\hat{\mathbf{Y}}$.

2) Se calcula la matriz de covarianzas de la matriz de valores ajustados $\hat{\mathbf{Y}}$ a través de la ecuación:

$$\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}} = \mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}'_{\mathbf{YX}}$$

Donde $\mathbf{S}_{\mathbf{YY}}$ es la matriz de covarianzas ($p \times p$) entre las variables de respuesta, $\mathbf{S}_{\mathbf{XX}}$ es la matriz de covarianzas ($m \times m$) entre las variables explicativas (que es la matriz $\mathbf{R}_{\mathbf{XX}}$ cuando las variables \mathbf{x} han sido estandarizadas).

3) Una vez que se tiene la matriz $\hat{\mathbf{Y}}$ se debe evaluar si la regresión es significativa para dar por válida la ordenación restringida y concluir sobre los resultados.

Legendre y Legendre (2012) presentan tres estadísticos para evaluar la significancia de la regresión a partir del estadístico R^2 :

- $R^2 = \frac{\text{Var}(\hat{\mathbf{Y}})}{\text{Var}(\mathbf{Y})} = \frac{\text{SC}(\hat{\mathbf{Y}})}{\text{SC}(\mathbf{Y})} = \frac{\text{Sum}(\hat{\mathbf{Y}})}{\text{Sum}(\mathbf{Y}_c)}$ y R^2 ajustado: $R_a^2 = 1 - (1 - R^2) \frac{(n-1)}{(n-m-1)}$

Las sumas de cuadrado (SC) de \mathbf{Y} centrada son la sumatoria de cada vector \mathbf{y} al cuadrado.

El R^2 mide la fuerza de la relación canónica entre \mathbf{Y} y \mathbf{X} calculando la proporción de la variación de \mathbf{Y} explicada por las variables en \mathbf{X} . El R^2 ajustado también mide la fuerza de la relación entre \mathbf{Y} y \mathbf{X} , pero aplica una

corrección del R^2 para tener en cuenta el número de variables explicativas. Cuanto más variables explicativas se usan, mayor es R^2 .

- $AIC = n \log_e \left(\frac{1-R^2}{n} \right) + 2k$. El Criterio de información de Akaike (AIC) se utiliza como medida de bondad del ajuste. La primera parte de la definición de AIC es una medida de bondad de ajuste, la segunda parte es una penalización por el número de parámetros en el modelo. El AIC se puede calcular para cada combinación posible de variables explicativas, y el modelo con el AIC más pequeño se elige como el mejor modelo (Zuur *et al.*, 2007).
- **Prueba de permutación.** El estadístico de prueba es el “pseudo-F”, siendo la proporción de inercia total restringida y no restringida, cada una dividida por sus respectivos grados de libertad. El algoritmo que calcula el pseudo-F, compara todos los términos con el mismo residuo del modelo completo (Legendre *et al.*, 2011).

$F = \frac{R_{Y|X}^2/m}{(1-R_{Y|X}^2)(n-m-1)}$ donde m es el número de variables explicativas o rango de la matriz de covarianzas de \mathbf{X} y n son los grados de libertad (ecuación utilizada para las pruebas de permutaciones en **vegan** de R y CANOCO (Legendre y Legendre, 2012)).

La prueba se realiza permutando las filas de \mathbf{Y} , y recalculando la descomposición canónica, esta operación se realiza un número alto de veces (generalmente superior a mil). Este valor indica si se ha explicado la mayoría de la variabilidad de \mathbf{Y} con la \mathbf{X} .

Al igual que los valores ajustados de una regresión lineal múltiple, los ejes canónicos (puntajes de ordenación de objetos) también son combinaciones lineales de las variables explicativas en \mathbf{X} (ter Braak y Prentice, 1988; ter Braak, 1995).

b. ACP de $\hat{\mathbf{Y}}$

El segundo paso es un ACP de la matriz $\hat{\mathbf{Y}}$. Este ACP produce los valores propios o autovalores canónicos y los vectores propios o autovectores también como los

ejes canónicos (coordenadas de ordenación de individuos). Este paso es realizado para obtener diagramas de ordenación del espacio reducido que muestran los objetos, las variables respuesta y las variables explicativas para los ejes más importantes de la relación canónica.

1) La matriz de valores ajustados $\hat{\mathbf{Y}}$ es sometida a un Análisis de Componentes Principales para reducir la dimensionalidad de la solución. Esto corresponde a resolver:

$$(\mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}'_{\mathbf{YX}} - \lambda_k\mathbf{I})\mathbf{u}_k = 0$$

que es la ecuación del RDA.

La matriz que contiene los vectores propios canónicos normalizados (norma 1) \mathbf{u}_k se llama \mathbf{U} . Los autovectores dan las contribuciones de las variables en la matriz $\hat{\mathbf{Y}}$ a los distintos ejes canónicos. La matriz \mathbf{U} es de tamaño $(p \times p)$.

Con el ACP de la matriz $\hat{\mathbf{Y}}$ se extraen los ejes de ordenación restringidos, mientras que con el ACP de la matriz de valores residuales $\hat{\mathbf{Y}}_{res}$ se extraen los ejes no restringidos.

Los coeficientes canónicos en la matriz normalizada \mathbf{U} dan las contribuciones de las variables de $\hat{\mathbf{Y}}$ a los ejes canónicos. Pueden ser interpretados como en un ACP. La matriz \mathbf{U} es utilizada para producir los biplots de escalado 1 o los diagramas triplot. Para los gráficos de escalado tipo 2, \mathbf{U} es reescalada de tal manera que la longitud de cada autovector es $\sqrt{\lambda_k}$.

Se deben realizar cálculos adicionales para producir el diagrama de triplot del RDA, que contiene tres tipos de elementos: variables de respuesta (por ejemplo, especies), objetos (por ejemplo, sitios) y variables explicativas.

2) La ordenación de los individuos en el espacio de las variables respuesta \mathbf{Y} es obtenido directamente de la matriz centrada \mathbf{Y}_c , utilizando la ecuación estándar para componentes principales \mathbf{F} y la matriz \mathbf{U} de autovectores \mathbf{u}_k :

$\mathbf{F} = \mathbf{Y}_c\mathbf{U}$ Ordenación de sitios en el espacio de \mathbf{Y} .

Los vectores de ordenación (columnas de \mathbf{F}) son llamados “vectores de coordenadas de sitios” (“*site scores*” en inglés). Tienen varianzas cercanas a los autovalores.

Las coordenadas de sitios se obtienen proyectando los datos originales de la matriz \mathbf{Y} en el eje k . Esto se aproxima a los datos observados que contienen los residuales ($\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{Y}_{\text{res}}$).

3) La ordenación de los individuos en el espacio de la \mathbf{X} se obtiene de:

$$\mathbf{Z} = \hat{\mathbf{Y}}\mathbf{U} = \mathbf{X}\mathbf{B}\mathbf{U} \text{ Ordenación de sitios en el espacio de } \mathbf{X}$$

Esta ordenación es combinación lineal de \mathbf{X} , por eso los vectores de ordenación son llamados “coordenadas ajustadas de sitios” (“*fitted site scores*” en inglés) o “coordenadas de muestras”, que son combinaciones lineales de las variables ambientales en CANOCO. Los vectores de ordenación tienen varianzas iguales a sus correspondientes autovalores.

Las coordenadas ajustadas de sitios son obtenidas proyectando los valores ajustados de la regresión múltiple (matriz $\hat{\mathbf{Y}}$) en el eje k y aproximan los datos ajustados.

4) La correlación r_k entre los vectores de ordenación en los espacios \mathbf{Y} y \mathbf{X} para la dimensión k se denomina "correlación especie-ambiente". Mide la fuerza de la relación entre los dos conjuntos de datos expresados por cada eje canónico k . Debe interpretarse con precaución porque un eje canónico con alta correlación especie-ambiente puede explicar una pequeña fracción de la variación en \mathbf{Y} , que viene dada por la cantidad (o proporción) de varianza de la matriz \mathbf{Y} explicada por cada eje canónico.

5) Para la contribución de las variables explicativas \mathbf{X} a los ejes de ordenación canónica, se pueden considerar los coeficientes de regresión o de correlación:

Matriz \mathbf{C} de coeficientes canónicos $\mathbf{C} = \mathbf{B}\mathbf{U}$

La matriz \mathbf{C} da directamente los pesos de las variables explicativas \mathbf{X} en la formación de la matriz de coordenadas ajustadas del sitio. La ordenación de sitios

en el espacio de las variables explicativas se puede encontrar directamente calculando \mathbf{XC} . Los coeficientes en las columnas de la matriz \mathbf{C} son idénticos a los coeficientes de regresión de las coordenadas de ordenación de $\mathbf{Z} = \hat{\mathbf{Y}}\mathbf{U} = \mathbf{XBU}$, en la matriz de variables explicativas estandarizadas \mathbf{X} .

Las correlaciones también se pueden calcular entre las variables en \mathbf{X} , por un lado, y los vectores de ordenación, en el espacio \mathbf{Y} o en el espacio \mathbf{X} por el otro. Las correlaciones entre \mathbf{X} y los vectores de ordenación en el espacio \mathbf{X} , matriz $\mathbf{R}_{\mathbf{XZ}} = \text{cor}(\mathbf{X}, \mathbf{Z})$ se utilizan para representar las variables explicativas en los biplots.

c. Ejes no canónicos

El RDA generalmente no explica completamente la variación en las variables respuesta (matriz \mathbf{Y}). Durante el paso de regresión, los residuos de regresión pueden calcularse para cada variable \mathbf{y} ; los residuales son las diferencias entre los valores observados y_{ij} en la matriz \mathbf{Y} y los valores ajustados correspondientes \hat{y}_{ij} en la matriz $\hat{\mathbf{Y}}$. La matriz de residuos (\mathbf{Y}_{res}) también es una matriz de tamaño $(n \times p)$.

Los residuos pueden analizarse mediante ACP, lo que conduce a $\min[p, n-1]$ autovalores y autovectores no canónicos. Cuando las variables en \mathbf{X} son buenas predictoras de las variables en \mathbf{Y} , los autovalores canónicos pueden ser mayores que los primeros autovalores no canónicos, pero este no es siempre el caso.

2.3.2.2. Gráficos del RDA

El resultado más importante de un método de ordenación aplicado a datos particulares es el diagrama o gráfico de ordenación. Usando este diagrama, es teóricamente posible reconstruir (con un cierto nivel de error) no solo la matriz de datos primaria (los datos de los taxones), sino también la matriz de (dis) similitudes entre muestras y la matriz de correlación de taxones (Leps y Smilauer, 2003).

En el RDA, se pueden dibujar diagramas llamados biplots, que contienen dos conjuntos de puntos como en el ACP o diagramas triplot que contienen tres conjuntos: las coordenadas del sitio (matrices \mathbf{F} o \mathbf{Z}), las variables de respuesta

de \mathbf{Y} y las variables explicativas de \mathbf{X} . Cada par de conjuntos de puntos se puede graficar en un biplot. Los biplots ayudan a interpretar la ordenación de los objetos en términos de \mathbf{Y} y \mathbf{X} . Cuando hay demasiados objetos o demasiadas variables en \mathbf{Y} o \mathbf{X} , se pueden graficar y presentar diagramas de ordenación separados para la respuesta y las variables explicativas. Al igual que en ACP, se pueden usar dos tipos principales de escalamiento (ter Braak, 1994; Legendre y Legendre, 2012).

Legendre y Legendre (2012) utilizaron los nombres de biplot de correlación (escalado 2) y biplot de distancia (escalado 1). Matemáticamente, la diferencia entre estas dos escalas son unas multiplicaciones simples que involucran a los autovalores. La interpretación ecológica de los dos tipos de triplots es diferente, si el interés está en las observaciones (sitios), entonces se debe usar el triplot de distancia (Zuur *et al.*, 2007).

En el triplot, las variables de respuesta (típicamente especies) están representadas por líneas, y las observaciones (en este caso sitios) por puntos. Lo nuevo que se introduce son las líneas para las variables explicativas cuantitativas (Zuur *et al.*, 2007). Los diagramas de esta ordenación muestran sitios, taxones y variables ambientales para que las relaciones entre ellos puedan derivarse lo más fácilmente a través de proyecciones, ángulos entre líneas o distancias entre puntos y líneas (Ramette, 2007).

RDA escalado 1 – Biplot de distancia (foco en los sitios)

Los autovectores en la matriz \mathbf{U} , que representan las coordenadas de las variables respuesta a lo largo de los ejes canónicos, se escalan a longitudes iguales a 1. Las coordenadas del sitio en el espacio \mathbf{X} se obtienen de la ecuación $\mathbf{Z} = \hat{\mathbf{Y}}\mathbf{U}$; estos vectores tienen varianzas iguales a λ_k .

Las coordenadas del sitio en el espacio \mathbf{Y} se obtienen de la ecuación $\mathbf{F} = \mathbf{Y}\mathbf{U}$. Las varianzas de estos vectores suelen ser ligeramente mayores que λ_k , porque \mathbf{Y} contiene los componentes ajustados y residuales y, por lo tanto, tiene más variación total que $\hat{\mathbf{Y}}$.

Las matrices \mathbf{Z} y \mathbf{U} , o \mathbf{F} y \mathbf{U} , se pueden usar juntas en biplots porque los productos de los vectores propios con las matrices de puntuación del sitio reconstruyen perfectamente las matrices originales: $\mathbf{ZU}' = \hat{\mathbf{Y}}$ y $\mathbf{FU}' = \mathbf{Y}$ como en el ACP.

En el escalado tipo 1, una variable explicativa cuantitativa \mathbf{x} se representa en el biplot o triplot usando el vector de correlaciones de \mathbf{x} con las coordenadas ajustadas del sitio, $\mathbf{r}_{\mathbf{xZ}} = \text{cor}(\mathbf{x}, \mathbf{Z})$, modificado multiplicando cada correlación por $\sqrt{\lambda_k / \text{Varianza total en } \mathbf{Y}}$, donde λ_k es el autovalor del eje correspondiente k . Esta corrección se explica por el hecho de que en esta escala las variaciones de las coordenadas del sitio difieren entre los ejes.

La matriz completa de coordenadas biplot en el escalado tipo 1 (\mathbf{BS}_1) para las variables explicativas se calcula de la siguiente manera (Legendre y Legendre, 2012):

$$\mathbf{BS}_1 = (\text{Varianza total en } \mathbf{Y})^{-1/2} \mathbf{R}_{\mathbf{xZ}} \mathbf{\Lambda}^{1/2}$$

Las gráficas resultantes de este escalado, llamadas biplots de distancia o triplots, enfocan la interpretación en la ordenación de individuos (sitios) porque las distancias entre individuos se aproximan a sus distancias Euclídeas en los espacios correspondientes a las matrices \mathbf{Y} o $\hat{\mathbf{Y}}$.

Las características principales de un **biplot o triplot de distancia** son las siguientes (ter Braak, 1994; Leps y Smilauer, 2003; Zuur *et al.*, 2007; Legendre y Legendre, 2012):

- Preserva la distancia Euclídea entre los sitios en la matriz $\hat{\mathbf{Y}}$.
- Los ángulos entre los vectores de las variables \mathbf{y} no son significativos, ya que no representan correlaciones.
- La proyección de un objeto individuo en ángulo recto sobre la variable respuesta o sobre una variable cuantitativa explicativa, aproxima la posición del individuo sobre dicha variable, e indica los valores (ajustados).
- En el biplot, los ángulos entre las variables respuesta (especies) y explicativas reflejan sus correlaciones.

- La relación entre el centroide de una variable explicativa y la variable respuesta, es encontrada proyectando el centroide en ángulo recto con la variable (como para los individuos o sitios).
- Las distancias dentro del centroide, y entre centroides y sitios individuales se aproximan a la distancia Euclídea.

RDA Escalado 2 – Biplot de correlación (foco en los taxones)

Los autovectores son escalados a la unidad de longitud.

Se obtienen las coordenadas de las variables respuesta por reescalado de los autovectores en la matriz \mathbf{U} a longitudes $\sqrt{\lambda_k}$, usando la transformación $\mathbf{U}\mathbf{\Lambda}^{1/2}$ como en el ACP.

Las coordenadas del sitio en el espacio \mathbf{X} obtenidas para escalado tipo 1 se reescalan a unidades de varianza utilizando la transformación $\mathbf{Z}\mathbf{\Lambda}^{-1/2}$ (es la misma transformación utilizada en ACP) para obtener la matriz \mathbf{G} de las coordenadas del sitio en la escala tipo 2.

Asimismo, las coordenadas del sitio en el espacio \mathbf{Y} obtenidas para el escalado tipo 1 se reescalan utilizando la transformación $\mathbf{F}\mathbf{\Lambda}^{1/2}$. Las varianzas de estos vectores suelen ser ligeramente mayores que 1 (al igual que el escalado 1).

Las matrices $\mathbf{Z}\mathbf{\Lambda}^{-1/2}$ y $\mathbf{U}\mathbf{\Lambda}^{1/2}$, o $\mathbf{F}\mathbf{\Lambda}^{-1/2}$ y $\mathbf{U}\mathbf{\Lambda}^{1/2}$, se pueden usar juntas en biplots porque los productos de los vectores propios con las matrices de coordenadas del sitio reconstruyen las matrices originales perfectamente: $\mathbf{Z}\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{1/2}\mathbf{U}' = \hat{\mathbf{Y}}$ y $\mathbf{F}\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{1/2}\mathbf{U}' = \mathbf{Y}$, como en el ACP.

En el escalado tipo 2, una variable explicativa cuantitativa x se representa en el biplot usando el vector de correlaciones de x con las coordenadas del sitio ajustado, $\mathbf{r}_{\mathbf{xZ}} = \text{cor}(\mathbf{x}, \mathbf{Z})$ sin más transformación. La matriz de coordenadas biplot (\mathbf{BS}_2) para las variables explicativas es entonces:

$$\mathbf{BS}_2 = \mathbf{R}_{\mathbf{xZ}} = \text{cor}(\mathbf{X}, \mathbf{Z})$$

Las reglas para la interpretación del **triplot de correlación RDA** (Triplot condicional de especies) son las siguientes (ter Braak, 1994; Leps y Smilauer, 2003; Zuur *et al.*, 2007; Legendre y Legendre, 2012):

Los gráficos resultantes de este escalado, llamados biplot de correlación o triplots, hacen foco en las variables de respuesta, matrices \mathbf{Y} o $\hat{\mathbf{Y}}$. Las características son:

- Preserva las covarianzas entre las especies.
- Las distancias entre los sitios no son aproximaciones a las distancias Euclídeas en el espacio multidimensional.
- Los ángulos entre las variables explicativas y respuesta y entre las variables explicativas entre sí, reflejan sus correlaciones.
- La proyección de un sitio en ángulo recto sobre una variable explicativa o respuesta aproxima el valor del sitio a lo largo de dicha variable.
- Los ángulos entre las variables reflejan sus correlaciones.
- La relación entre el centroide de una variable explicativa cualitativa y una variable respuesta (especie) se encuentra proyectando el centroide en ángulo recto sobre la variable.
- La distancia dentro de los centroides y entre centroides y objetos (sitios) individuales, no se aproxima a la distancia euclídea.

2.3.3. ACC – Análisis de correspondencias canónico (ter Braak, 1986)

El análisis de correspondencias canónico (ACC) es un método de ordenación canónico asimétrico desarrollado por ter Braak (1986), es un método de análisis multivariado directo del gradiente. Fue implementado por primera vez en el programa CANOCO en el año 1985 (ter Braak, 1985; ter Braak, 1988, 1990; ter Braak y Smilauer, 1998) y actualmente se encuentra disponible en varios paquetes computacionales y funciones de R para ecología comunitaria.

El ACC es la forma canónica (restringida) del Análisis de Correspondencias (AC), es decir es un ACP restringido sobre la matriz de especies, que ha sido transformada en una matriz de estadísticos de Chi-cuadrado de Pearson. Cualquier matriz que puede ser sujeta a un AC es adecuada para un ACC, en

particular para presencia-ausencia de especies y datos de abundancia (Legendre y Legendre, 2012). En ACC, el espacio restringido (o espacio canónico) es la parte del espacio total en el que limitamos la búsqueda de los ejes principales óptimos, el espacio no restringido (o espacio no canónico) es el resto del espacio total (Greenacre, 2008).

ACC combina aspectos de ordenación con aspectos del análisis directo del gradiente a través de la regresión. ter Braak (1986) lo presenta como un Análisis de Correspondencias donde los ejes son elegidos en base a las variables ambientales, que participan en forma activa en el análisis. El ACC restringe la ordenación de una matriz por medio de regresiones lineales múltiples con variables de una segunda matriz.

En el ACC, en vez de buscar la solución que mejor ajuste los ejes principales en el espacio completo de datos, lo que se hace, es ajustar los ejes principales en una parte limitada o restringida del espacio (por tanto, podría considerarse el ACC como un Análisis de Correspondencias restringido) (Greenacre, 2008).

El ACC es una forma ponderada de RDA, aplicado a la misma matriz \bar{Q} de contribuciones al estadístico Chi-cuadrado que el usado en el AC. La matriz \bar{Q} se calcula a partir de datos de frecuencia o presencia-ausencia, como en el Análisis de Correspondencias (AC).

El ACC, es similar al RDA, la diferencia es que preserva la distancia Chi-cuadrado (como en el AC), en lugar de la distancia Euclídea entre los datos de la matriz \hat{Y} . Los cálculos son un poco más complejos ya que la matriz \hat{Y} contiene valores ajustados obtenidos por regresión lineal ponderada con las variables explicativas X (Legendre y Legendre, 2012).

Es un método usual en el campo de la ecología, en este caso la matriz de frecuencias es una matriz de abundancia de diferentes especies (o una matriz presencia-ausencia) en diferentes sitios ecológicos mientras que la otra matriz describe dichos sitios por sus características ambientales. Como las especies son atraídas por condiciones favorables, las variables ambientales son consideradas como explicativas. Así, los dos conjuntos de variables no juegan un papel

simétrico. El ACC consigue introducir las variables explicativas dentro del análisis dándoles un papel activo y explicar los datos relacionándolos con ellas. En este caso selecciona la combinación lineal de las variables ambientales que maximiza la dispersión de las especies (ter Braak, 1987).

Como se mencionó anteriormente, el ACC funciona bien para sets de datos donde las variables de respuesta son unimodales.

Los métodos de ordenación unimodales, o métodos de promedio ponderado (AC, ACC y DCA), son basados en el modelo de respuesta de especies, donde se supone que una curva simétrica unimodal (en forma de campana) describe mejor el cambio de los valores de las especies a lo largo de los ejes de ordenación (Leps y Smilauer, 2003). El ACC se introduce como una extensión multivariada de la ordenación de promedio ponderada, que es un método simple para organizar las especies a lo largo de las variables ambientales.

El ACC se convierte en la técnica que selecciona la combinación lineal de las variables ambientales que maximiza la dispersión de las coordenadas de especies en el primer eje de ordenación (ter Braak, 1987; Jongman *et al.*, 1987). Los autovalores producidos por ACC miden esta separación (ter Braak, 1987).

La técnica propuesta por ter Braak (1986), estima simultáneamente el óptimo de especies, los coeficientes de regresión, y, por lo tanto, las coordenadas del sitio. Las coordenadas de especies representan el centroide de la especie, o el modo de la curva de respuesta unimodal de la especie.

ACC utiliza el modelo unimodal para modelar la respuesta de las especies a la variación ambiental como una simplificación matemática para permitir la estimación de una gran cantidad de parámetros y la identificación de una pequeña cantidad de ejes de ordenación (Ramette, 2007).

2.3.3.1. Pasos de un ACC

Existen diferentes algoritmos para realizar los cálculos del ACC. A continuación, se enuncian los pasos según Legendre y Legendre (1998, 2012):

a. Transformación de datos

La primera diferencia con el RDA es que no se utiliza la matriz de respuesta \mathbf{Y} centrada, sino que en ACC se usa la matriz $\bar{\mathbf{Q}}$, como en el AC, para trabajar con la distancia Chi-cuadrado.

Al igual que en AC, a partir de la matriz \mathbf{Y} de frecuencias se obtiene la matriz de frecuencias relativas \mathbf{Q} con término general $[p_{ij}]$. Luego, se estandariza la matriz \mathbf{Q} obteniéndose $\bar{\mathbf{Q}}_{(n \times p)}$:

$$\bar{\mathbf{Q}} = [\bar{q}_{ij}] = \left[\frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} \right]$$

La segunda diferencia con el RDA, es que una matriz diagonal de los pesos de las filas $\mathbf{D}(p_{i+})$ es utilizada en los cálculos de la regresión multivariada. Para cada fila de \mathbf{Y} , f_{i+} es la suma de los valores en la fila i , y p_{i+} es f_{i+} dividido el gran total f_{++} de las frecuencias en \mathbf{Y} .

La matriz \mathbf{X} es estandarizada a \mathbf{X}_{est} , usando los pesos de las frecuencias a través de la matriz $\mathbf{D}(f_{i+})$ (Legendre y Legendre, 2012). Se calculan las medias y varianzas de las columnas de \mathbf{X} ponderadas por los marginales filas de la matriz \mathbf{Y} (f_{i+}), a través de: $\sum_{i=1}^n f_{i+}x_{ik} = \bar{x}_k$ y $\sum_{i=1}^n f_{i+}(x_{ik} - \bar{x}_k)^2 = V(x_k)$, para estandarizar \mathbf{X} , obteniendo: $\mathbf{X}_{\text{est}} = \frac{x_{ik} - \bar{x}_k}{\sqrt{V(x_k)}}$

b. Regresión múltiple ponderada

Para obtener los coeficientes de regresión, se realiza una regresión múltiple ponderada. Los pesos o ponderaciones de filas de la matriz diagonal $\mathbf{D}(p_{i+})^{1/2}$ son aplicados a la matriz \mathbf{X} todas las veces que aparezca en la ecuación de la regresión multivariada:

$\mathbf{B} = [\mathbf{X}'_{\text{est}}\mathbf{D}(p_{i+})\mathbf{X}_{\text{est}}]^{-1}\mathbf{X}'_{\text{est}}\mathbf{D}(p_{i+})^{1/2}\bar{\mathbf{Q}}$ Estima la matriz de coeficientes de regresión.

$\hat{\mathbf{Y}} = \mathbf{D}(p_{i+})^{1/2}\mathbf{X}_{\text{est}}\mathbf{B}$ Estima los valores predichos.

Por lo tanto, la ecuación completa final para calcular la matriz de valores predichos es:

$$\hat{\mathbf{Y}} = \mathbf{D}(p_{i+})^{1/2} \mathbf{X}_{\text{est}} [\mathbf{X}'_{\text{est}} \mathbf{D}(p_{i+}) \mathbf{X}_{\text{est}}]^{-1} \mathbf{X}'_{\text{est}} \mathbf{D}(p_{i+})^{1/2} \bar{\mathbf{Q}}$$

La matriz de residuales se calcula como $\bar{\mathbf{Q}}_{\text{res}} = \bar{\mathbf{Q}} - \hat{\mathbf{Y}}$

c. Descomposición en valores singulares – Cálculo de coordenadas

Finalmente se realiza el ACP sobre la matriz de frecuencias relativas predichas $\hat{\mathbf{Y}}$.

La descomposición en valores singulares es llevada a cabo sobre la matriz:

$$\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} = \mathbf{S}_{\bar{\mathbf{Q}}\mathbf{X}} \mathbf{S}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{S}'_{\bar{\mathbf{Q}}\mathbf{X}}$$

si las covarianzas son: $\mathbf{S}_{\bar{\mathbf{Q}}\mathbf{X}} = \bar{\mathbf{Q}}'\mathbf{D}(p_{i+})^{1/2}\mathbf{X}$ y $\mathbf{S}_{\mathbf{X}\mathbf{X}} = \mathbf{X}'\mathbf{D}(p_{i+})\mathbf{X}$

La matriz normalizada $\hat{\mathbf{U}}$ en ACC no contiene los pesos de las filas de $\hat{\mathbf{Y}}$ en los ejes canónicos, sino que contiene los pesos (cargas o “loadings” en inglés) de las filas de $\bar{\mathbf{Q}}$ en los ejes de ordenación, como en el AC. Se obtiene:

$$\hat{\mathbf{U}} = \bar{\mathbf{Q}}\mathbf{U}\Lambda^{-1/2}$$

Las variables explicativas \mathbf{X} se proyectan a partir de sus correlaciones con los ejes de ordenación.

d. Ejes no canónicos

Los residuales pueden ser analizados aplicando la descomposición en valores singulares de la matriz $\bar{\mathbf{Q}}_{\text{res}}$. Esta descomposición es la que da, al igual que en el RDA, los ejes no canónicos.

e. Inercia

La inercia (o varianza total) en ACC se identifica de la misma manera que en AC. El autovalor de un eje viene dado por la desviación estándar ponderada del algoritmo ACC. Estos autovalores también se denominan autovalores canónicos. La cantidad de variación que puede explicarse por todas las variables ambientales

es igual a la suma de todos los autovalores canónicos. La cantidad de variación explicada por los dos primeros ejes puede expresarse como un porcentaje de la inercia total, y como un porcentaje de la varianza que puede explicarse por las variables ambientales (Zuur *et al.*, 2007).

La suma de cuadrados de todos los valores en la matriz $\bar{\mathbf{Q}}$, $\sum \bar{q}_{ij}^2$, mide la inercia total en $\bar{\mathbf{Q}}$. Es igual a la suma de todos los autovalores que son extraídos del Análisis de Correspondencias de $\bar{\mathbf{Q}}$.

2.3.3.2. Gráficos del ACC

Para las especies y sitios las interpretaciones son las mismas que en AC.

A continuación, se listan las diferencias del ACC con las salidas en el RDA (Legendre y Legendre, 2012):

- La variación ahora es expresada como “cuadrado medio del coeficiente de contingencia”. Es sesgada, pero no puede ajustarse fácilmente.
- El número máximo de ejes canónicos ACC es $\min[(p-1), m, n-1]$. El número mínimo de ejes residuales es $\min [(p-1), n-1]$.
- Las coordenadas de especies están representadas como puntos en el triplot.
- Las coordenadas de sitios son promedios ponderados (en lugar de sumas ponderadas) de las coordenadas de especies.

$\hat{\mathbf{U}}$ es utilizada para encontrar las coordenadas de sitios (matrices \mathbf{F} y $\hat{\mathbf{V}}$) en el espacio de las variables originales \mathbf{y} . Las coordenadas de sitios en el espacio de los valores ajustados $\hat{\mathbf{y}}$ serán encontradas usando \mathbf{U} en lugar $\hat{\mathbf{U}}$.

La matriz \mathbf{V} de las coordenadas de las especies del escalado tipo 1 y la matriz $\hat{\mathbf{V}}$ de las coordenadas de especies del escalado tipo 2, son también obtenidas de \mathbf{U} y $\hat{\mathbf{U}}$ utilizando las transformaciones descritas en el AC.

Para el espacio de las variables explicativas, las coordenadas de sitios que son combinaciones lineales de las variables ambientales corresponden a la ecuación del RDA: $(\mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}'_{\mathbf{YX}} - \lambda_k\mathbf{I})\mathbf{u}_k = \mathbf{0}$

Antes de calcular las coordenadas del biplot, la matriz \mathbf{Z} debe ser estandarizada a \mathbf{Z}_{est} utilizando el mismo procedimiento de estandarización de la matriz \mathbf{X} (Legendre y Legendre, 2012). Las matrices de coordenadas para el biplot (BS) de las variables explicativas puede ser calculado utilizando \mathbf{X}_{est} , \mathbf{Z}_{est} , y la matriz diagonal de los pesos de las filas $\mathbf{D}(p_{i+})$.

ACC Escalado 1 (foco en los sitios)

La posición de los individuos u objetos (sitios) en una variable explicativa cuantitativa, puede ser obtenida por la proyección de los sitios en ángulo recto sobre la variable.

En el espacio de \mathbf{Y} y de $\hat{\mathbf{Y}}$:

$\mathbf{V} = \mathbf{D}(p_{+j})^{-1/2}\mathbf{U}$ Coordenadas de especies escalado tipo 1.

$\mathbf{F} = \hat{\mathbf{V}}\mathbf{\Lambda}^{1/2}$ Coordenadas de sitios escalado tipo 1.

En el espacio de las variables ambientales \mathbf{X} :

$\mathbf{Z}_1 = \mathbf{D}(p_{i+})^{-1/2}\hat{\mathbf{Y}}\mathbf{U}$ Coordenadas de sitios para escalado tipo 1 que son combinaciones lineales de las variables ambientales.

Las matrices de las coordenadas de los biplots (BS “*Biplot scores*”) para las variables explicativas pueden ser calculadas utilizando la \mathbf{X} estandarizada, la \mathbf{Z} estandarizada y la matriz diagonal de pesos $\mathbf{D}(p_{i+})$ de la siguiente manera:

$\mathbf{BS}_1 = \mathbf{X}'_{\text{est}}\mathbf{D}(p_{i+})\mathbf{Z}_{\text{est}}\mathbf{\Lambda}^{1/2}$ para escalado tipo 1.

ACC Escalado 2 (foco en los taxones)

El óptimo de una especie a lo largo de una variable ambiental cuantitativa puede ser obtenido por la proyección de la especie en ángulo recto sobre la variable.

Las coordenadas de las filas (sitio) (vector “*rowscore*” en inglés) dan la ordenación de los sitios en el espacio de los datos originales de la especie (matriz $\hat{\mathbf{V}}$). Las coordenadas de las columnas (cargas de especies) están en el vector “*colscore*”,

que corresponden a la matriz $\hat{\mathbf{F}}$. Las coordenadas ajustadas de sitios (vector “*fitted-row score*”) dan la ordenación de los sitios en el espacio de las variables explicativas \mathbf{X} .

Coordenadas en el espacio de \mathbf{Y} y de $\hat{\mathbf{Y}}$:

$\hat{\mathbf{V}} = \mathbf{D}(p_{i+})^{-1/2} \hat{\mathbf{U}} = \mathbf{D}(p_{i+})^{-1/2} \bar{\mathbf{Q}} \mathbf{U} \mathbf{\Lambda}^{-1/2}$ Coordenadas de sitios escalado tipo 2.

$\hat{\mathbf{F}} = \mathbf{V} \mathbf{\Lambda}^{1/2}$ Coordenadas de especies escalado tipo 2.

Coordenadas en el espacio de las variables ambientales \mathbf{X} :

$\mathbf{Z}_2 = \mathbf{D}(p_{i+})^{-1/2} \hat{\mathbf{Y}} \mathbf{U} \mathbf{\Lambda}^{-1/2}$ Coordenadas de sitios para escalado tipo 2 que son combinaciones lineales de las variables ambientales.

La matriz de las coordenadas de los biplots para el escalado tipo 2 es:

$\mathbf{BS}_2 = \mathbf{X}'_{\text{est}} \mathbf{D}(p_{i+}) \mathbf{Z}_{\text{est}}$ para escalado tipo 2.

Triplot

El ACC proporciona las mismas posibilidades de elección de coordenadas para filas y columnas que el AC habitual. El triplot es un diagrama en el cual, además, se añaden vectores correspondientes a las variables explicativas. Para la visualización de las variables explicativas se utilizan los coeficientes de regresión estandarizados derivados de su relación con los ejes. La utilización de estos coeficientes refleja la idea de que en el ACC los ejes están relacionados linealmente con las variables explicativas (Greenacre, 2008).

Para el escalado tipo 1, el triplot es graficado utilizando la matriz \mathbf{V} para las especies, \mathbf{Z}_1 o \mathbf{F} , para los sitios y \mathbf{BS}_1 para las variables explicativas.

En el triplot, para el escalado tipo 1 las posiciones de las variables explicativas se modifican, y se calculan multiplicando cada correlación en las columnas de los k

ejes por: $\sqrt{\frac{\lambda_k}{\text{Varianza total en } \mathbf{Y}}}$

Para el escalado tipo 2, el triplot es graficado utilizando la matriz \hat{F} para las especies, Z_2 o \hat{V} , para los sitios y BS_2 para las variables explicativas.

Para el escalado tipo 2 se utilizan las correlaciones entre X y Z para representar las variables explicativas.

Legendre y Legendre (2012) mencionan un escalado tipo 3 (escalado simétrico), que puede ser realizado por los programas CANOCO y R, y es un compromiso entre los escalados 1 y 2. Este escalado no preserva la distancia Chi-cuadrado entre coordenadas de especies o de sitios.

2.3.3. Prueba de significancia del RDA y ACC

Cuando se utiliza una prueba t o F en una regresión lineal simple, los valores de estos estadísticos son comparados con los esperados según su distribución bajo H_0 . La distribución esperada de estas pruebas es la distribución original de los datos, asumiendo normalidad en la distribución de los residuales en regresión.

En un método restringido esta distribución no es conocida bajo H_0 y depende del número de variables respuesta, su estructura de correlación y la distribución de los valores de la variable respuesta. Esta distribución puede ser simulada por la **Prueba de permutación de Monte Carlo** (Leps y Smilauer, 2003).

Esta prueba es completamente libre de distribución. Esto significa que no depende de ningún supuesto sobre la distribución de los valores de abundancia de la especie (Leps y Smilauer, 2003; Legendre y Legendre, 2012).

En este caso, la H_0 postula la independencia entre las mismas filas (mismos sitios) de la matriz Y y la matriz X .

El procedimiento comienza reorganizando aleatoriamente (permutando) las muestras (filas) en la matriz de datos ambientales, manteniendo intactos los datos de las especies. Cualquier combinación de las especies y los datos ambientales obtenidos de esa manera es tan probable como el conjunto de datos "verdadero", si la H_0 es cierta.

Para cada uno de los conjuntos de datos permutados de esta manera, se calcula el modelo de ordenación restringida y se expresa su calidad de manera similar a la utilizada al juzgar la calidad de un modelo de regresión. En los métodos de ordenación restringida, igual que en un modelo de regresión, se utiliza un estadístico F, que es la relación entre la variabilidad de la variable de respuesta explicada por el modelo de regresión (dividido por el número de parámetros del modelo) y la variabilidad residual (dividida por el número de grados residuales de libertad).

Se registra el valor de dicho estadístico para cada versión permutada (aleatoria) del conjunto de datos. La distribución de estos valores define la distribución del estadístico de prueba bajo el modelo nulo. Si es altamente improbable que el valor del estadístico de prueba obtenido de los datos reales (sin permutación de las filas de la matriz de datos ambientales) provenga de esa distribución (es mucho mayor, es decir, corresponde a un modelo de ordenación de mayor calidad), se rechaza la H_0 .

Es difícil la elección de la definición de este estadístico para la ordenación canónica, debido a la multi-dimensionalidad de la solución obtenida. En general, la variabilidad de los datos de taxones, descrita por las variables ambientales es expresada por más de un eje canónico. Por lo tanto, se puede expresar la varianza utilizando todos los ejes canónicos o enfocarse en uno solo, típicamente el primero (Leps y Smilauer, 2003).

Para probar un solo eje a la vez se verifica si se puede obtener un valor propio igual o mayor bajo la hipótesis nula de que no hay relación entre la matriz respuesta y la matriz explicativa.

Para probar la importancia del análisis a nivel global, la base es la suma de todos los autovalores canónicos. Las hipótesis se plantean como:

- H_0 : no existe una relación lineal entre la matriz respuesta y la matriz explicativa;
- H_1 : existe una relación lineal entre la matriz respuesta y la matriz explicativa.

Según Legendre y Legendre (2012), originalmente, el estadístico de prueba era el autovalor o la suma de autovalores canónicos en sí mismo. Ahora, el estadístico F mencionado, es un estadístico "pseudo-F". Para probar todos los ejes canónicos que se define como:

$$\text{pseudoF} = \frac{\sum_{i=1}^m \lambda_i / m}{\text{RSS} / (n - m - 1)}$$

donde λ_i son los autovalores de los ejes canónicos, RSS es la suma de cuadrados residuales (diferencia entre la varianza total en los datos de taxones y la variabilidad explicada por todos los ejes canónicos ó suma de los autovalores no canónicos después de ajustar las variables explicativas), n el número de objetos o ejes canónicos, m el número de variables explicativas.

2.4. Esquemas de los métodos AFM, RDA y ACC

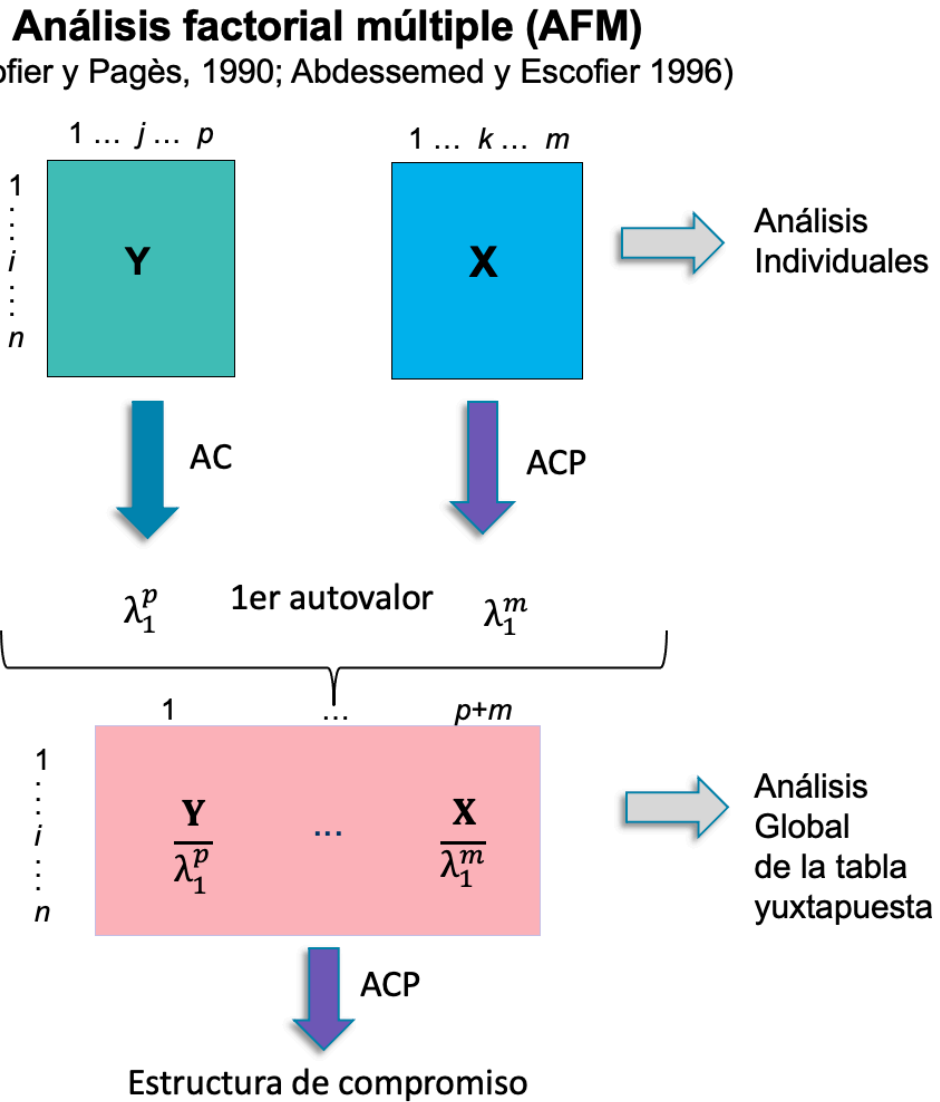


Figura 2.8. Esquema general del AFM.

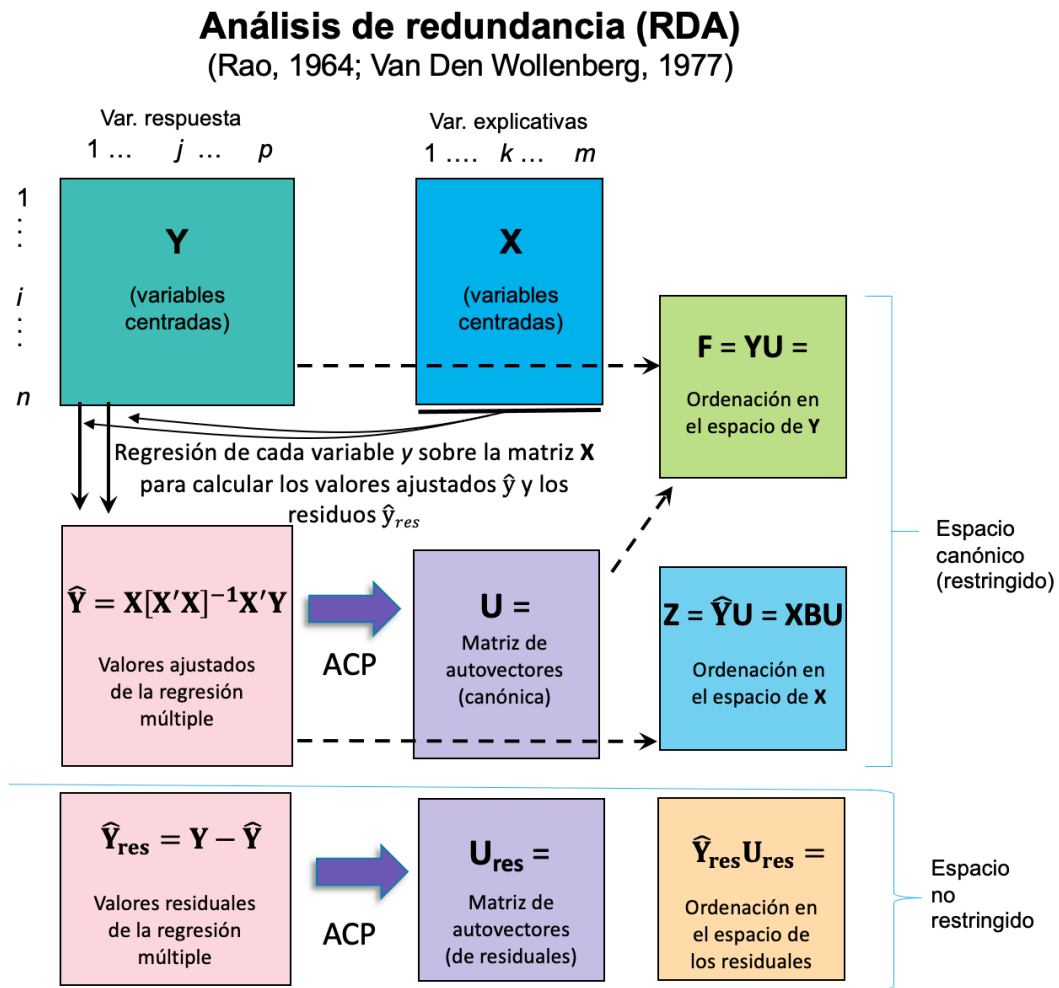


Figura 2.9. Esquema general de RDA (adaptado de Legendre y Legendre, 2012). Si la matriz Y es de presencia/ausencia o frecuencias, usar transformación de Hellinger.

Análisis de correspondencias canónico (ACC)

(ter Braak, 1986)

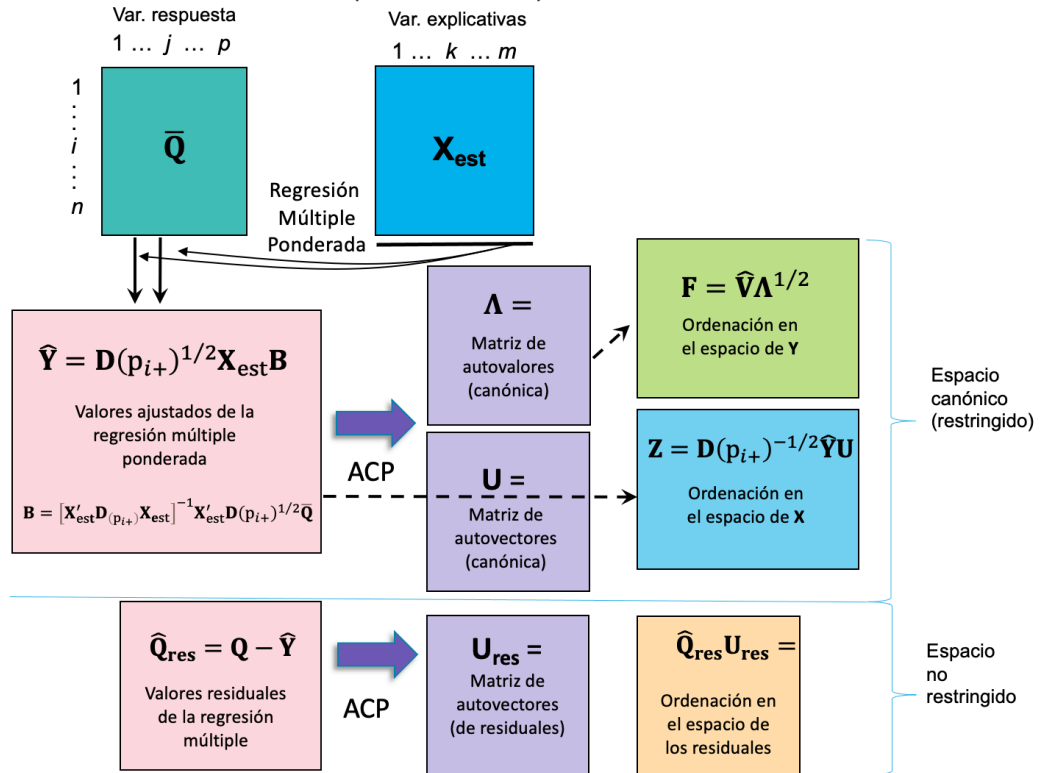


Figura 2.10. Esquema general del ACC. La matriz de entrada Y corresponde a \bar{Q} y las columnas de Y no están centradas, y X está estandarizada.

CAPÍTULO 3. Materiales

El objetivo del capítulo es presentar dos bases de datos diferentes para ejemplificar la aplicación y pertinencia de uso de cada uno de los métodos anteriormente descriptos: AFM, RDA y ACC.

3.1. Biomonitoreo cuenca Lacar - Hua Hum (LHH)

En el año 2015 en el marco de la capacitación en biomonitoreo de la US EPA (Agencia de Protección Ambiental de Estados Unidos) en la Provincia del Neuquén, Argentina, se muestrearon 10 sitios de la cuenca binacional Lacar - Hua Hum, donde se encuentra emplazada la localidad de San Martín de los Andes (71°21'12.13"O, 40°9'28.4"S). Luego, con el objetivo de avanzar en la determinación del estado ambiental de la mencionada cuenca, en el año 2016 se llevó a cabo el primer monitoreo interinstitucional, muestreándose 19 sitios (de los cuales 7 habían sido muestreados en 2015).

Para este trabajo se utilizaron los datos de muestreo de un total de 21 sitios de diferentes sistemas hidrológicos (arroyo, río, canal), realizados entre los meses de marzo y abril para ambos años.

Los sitios de muestreo fueron elegidos de acuerdo con sus características particulares en cuanto a la actividad antrópica y tipos de uso del suelo circundante. En la cabecera de cuenca se ubicó el sitio de control (identificado con el código PP0), sitio a mayor altura sobre el nivel del mar, sumándose sitios con diferentes grados de impacto antrópico hasta el cierre de cuenca, representando cada uno diferentes situaciones ambientales.

Se registraron parámetros fisicoquímicos *in situ* y de hábitat, y se tomaron muestras para análisis químicos y biológicos (macroinvertebrados). Los resultados fueron utilizados para analizar la calidad del agua a través de distintas variables físicas, químicas y biológicas. Los protocolos de muestreo tanto como la caracterización del hábitat fueron adaptados de Barbour *et al.* (1999) y Flotemersch *et al.* (2006).

En cada sitio fueron completadas planillas de caracterización del hábitat (Barbour *et al.*, 1999; AIC, 2012) para definir su calidad a partir del índice de “puntaje de evaluación de hábitat” (PEH) (Barbour *et al.*, 1996). Tabla XI Anexo.

Los individuos en este análisis estuvieron definidos por una combinación del sitio muestreado y el año. En la Figura 3.1 y Tabla I del Anexo se presentan los 21 sitios de muestreo que correspondieron a los biomonitoreos de la cuenca Lacar - Hua Hum, realizados en la época de estiaje de los años 2015 y 2016. Los sitios se asocian a tres subcuencas, Trabunco – Quitrahue (TQ), Pocahullo (P), Lacar – Hua Hum (L).

Para este caso de estudio, la matriz **Y** se conformó con datos de abundancia a nivel de familia de macroinvertebrados, en lugar de una clasificación taxonómica más fina como es la de géneros o especies. Se identificaron 46 familias, registrándose su frecuencia absoluta (Tabla VII del Anexo).

Se conformó la matriz **X** de variables ambientales, incluyéndose 13 variables en el estudio:

- conductividad eléctrica ($\mu\text{S}/\text{cm}$) “CE”
- oxígeno disuelto (mg/L) “OD”
- nivel de acidez o alcalinidad “PH”
- temperatura del agua ($^{\circ}\text{C}$) “TAG”
- altitud (m.s.n.m.) “ALT”
- nitrógeno total ($\mu\text{g}/\text{L}$) “NT”
- fósforo reactivo soluble ($\mu\text{g}/\text{L}$) “PRS”
- sólidos en suspensión totales ($\mu\text{g}/\text{L}$) “SST”
- hierro (mg/l) “FE”
- sílice (mg/l) “SI”

- dureza (mg/l) “DU”
- alcalinidad (mg/l) “ALC”
- puntaje de evaluación del hábitat “PEH”

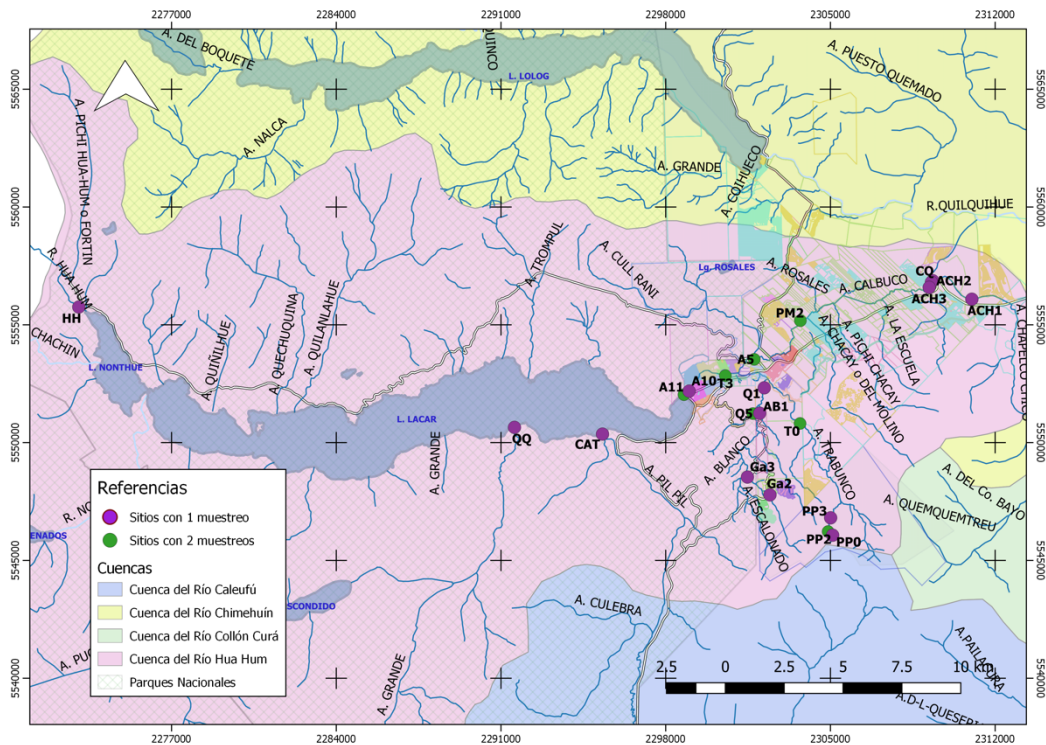


Figura 3.1. Mapa de distribución de sitios de muestreados. Biomonitoreo Lacar Hua Hum. Años 2015 y 2016.

Por lo tanto, la base de datos para el estudio conjunto de las matrices **X** e **Y** estuvo compuesta por 29 individuos ($n=29$), 13 variables ambientales ($m=13$) y 46 familias de macroinvertebrados ($p=46$), definiendo las matrices **X** de dimensión 29×13 e **Y** de dimensión 29×46 .



(a) Pradera del Puma, Cerro Chapelco.



(b) Plecópteros de la familia Austroperlidae, Pradera del Puma.



(c) Arroyo Trabunco.



(d) Red de tipo "D-Net".



(e) Arroyo Pocahullo.



(f) Muestra de macroinvertebrados.



(g) Arroyo Grande. Quila Quina, Lago Lacar.



(h) Lago Nonthue en nacimiento rio Hua Hum.

Figura 3.2. Imágenes de sitios del biomonitorio Lacar – Hua Hum. Años 2015 y 2016.

3.2. Monitoreo del alga exótica invasora *Didymosphenia geminata* (DIDYMO)

En el año 2010 se produjo el primer registro de presencia en Argentina del alga exótica invasora *Didymosphenia geminata*, en la cuenca del río Futaleufú, provincia de Chubut. Es por ello, que a principios del año 2011 se comenzó con el monitoreo para detección del alga en sitios distribuidos por toda la provincia del Neuquén, ocurriendo la primera detección en la provincia en el mes de noviembre del año 2011, sobre el río Collón Cura en Balsa Vieja.

En febrero de 2012 se implementó oficialmente el Programa de Vigilancia, Seguimiento y Control de *Didymosphenia geminata* (Lyngbye) Schmidt en la Provincia de Neuquén, a cargo de la Dirección General de Biología Acuática (SSA – MSTyA) y la Dirección Provincial de Fiscalización Hídrica (SRH – MESPyrN). En el marco de este programa se realizaron desde el año 2011 muestreos semestrales, y anuales desde el año 2015 para establecer la presencia-ausencia del alga exótica invasora de *D. geminata*,

Con el fin de realizar monitoreos integrales y ampliar el horizonte de conocimiento de la comunidad algal y su asociación con la especie exótica invasora, en cada campaña se obtuvieron muestras de material biológico de diferentes substratos (muestras multihábitat) y muestras de fitoplancton, y se determinó la composición y abundancia de toda la comunidad algal asociada. Se caracterizó el hábitat en cada uno de los sitios, se registraron parámetros fisicoquímicos y se extrajeron muestras de agua para análisis químicos. Detalles del muestreo se describen en el Anexo.

Los protocolos de muestreo fueron adaptados de Barbour *et al.* (1999) y Flotemersch *et al.* (2006) (AIC, 2012).

Para este trabajo de tesis se utilizaron datos de la VII campaña de muestreo del alga exótica invasora *D. geminata*, realizada entre los meses de abril y mayo de 2014. Se analizaron los datos disponibles de abundancia de 44 especies de algas (incluyendo *D. geminata*) de las muestras multihábitat y se incorporaron 9

variables ambientales para el estudio conjunto de las matrices **X** e **Y** (siglas y unidades iguales que en el monitoreo anterior):

CE, OD, PH, TAG, SST, ALC, PEH, PT (fósforo total - $\mu\text{g/L}$) y ORDEN (clasificación de ríos y arroyos de acuerdo con la cantidad de afluentes).

En la Figura 3.9 se presenta el mapa de distribución de los sitios del monitoreo del alga exótica invasora *D. geminata* en la Provincia del Neuquén, para el año 2014. Los sitios se asociaron a 4 cuencas: Río Colorado (RC), Río Hua Hum (LHH), Río Limay (RL), Río Neuquén (RN) (Tabla II del Anexo).

La base de datos analizada estuvo compuesta por 44 sitios ($n=44$), 9 variables ambientales ($m=9$) y 44 especies de algas ($p=44$). Definiendo las matrices **X** de dimensión 44×9 , e **Y** de dimensión 44×44 .

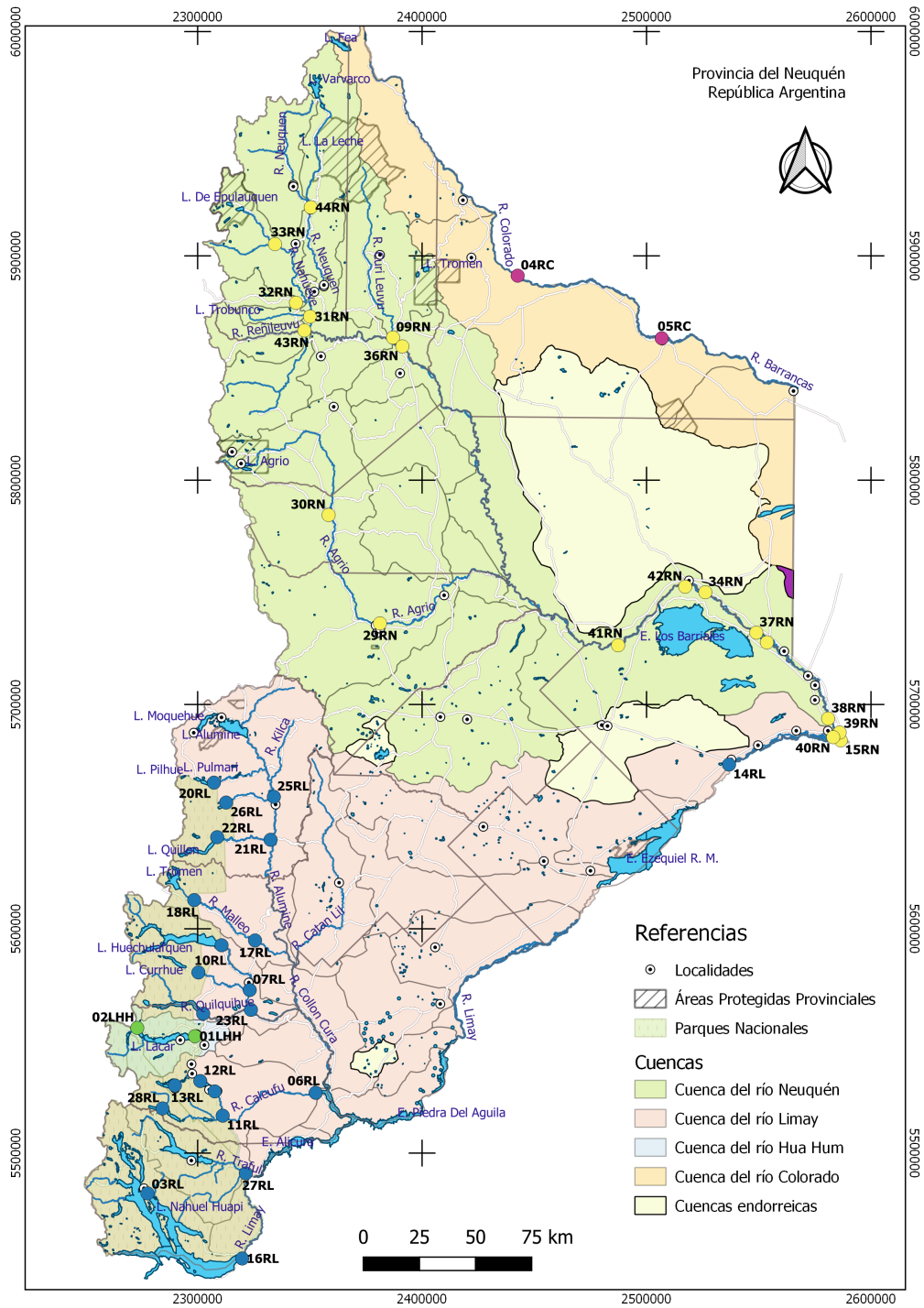


Figura 3.3. Mapa de distribución de sitios de muestreo. Los símbolos de sitios están representados en colores según la cuenca de pertenencia. Monitoreo del alga exótica invasora *Didymosphenia geminata*. Provincia del Neuquén. Año 2014.



(a) Río Varvarco.



(b) Río Chimehuín.



(c) Célula de *Didymosphenia geminata*.



(d) Arroyo El Once.



(e) Río Quilquihue.



(f) Río Limay.



(g) Río Meliquina.



(h) Río Neuquén.

Figura 3.4. Imágenes de sitios del monitoreo de *Didymosphenia geminata* en la Provincia del Neuquén.

3.3. Software

Para la realización de todos los análisis estadísticos se utilizaron los programas CANOCO 5 (Leps y Smilauer, 2003) y R (R Core Team, 2019) a través del entorno de desarrollo integrado RStudio.

Con el programa R, las funciones específicas utilizadas por librerías fueron:

“FactoMineR”: *MFA()*, *inertia.ratio()*, *Lg()*, *RV()*, *contrib()*, *cos2()*, *dimdesc()*, *plot()* (Le *et al.*, 2008). Para el AFM.

“vegan”: *decostand()*, *decorana()*, *rda()*, *anova.cca()*, *r.squared()*, *adj.r.squared()*, *coef()*, *goodness()*, *spenvcor()*, *inertcomp()*, *vif.cca()*, *intersetcor()*, *ordistep()*, *ordiR2step()*, *cca()*, *plot()* (Oksanen *et al.*, 2019). Para el RDA y ACC.

Tanto en CANOCO como en R, se ajustaron los parámetros gráficos para una mejor visualización de los resultados.

Para los gráficos de ordenación de métodos restringidos en R se detallan los siguientes elementos que se utilizan para trazar o extraer resultados mediante la función *plot()* (Borcard *et al.*, 2011; Oksanen *et al.*, 2019):

Tipos de Coordenadas (código de R en negrita)

- Coordenadas de especies – **“sp”** (“*stand for species*”). Coordenadas de las puntas de los vectores que representan las variables de respuesta en los biplots o triplots. Al igual que en ACP, dependen del escalado elegido.
- Coordenadas de sitios – **“wa”** (“*weighted averages*”). Coordenadas de los sitios expresadas en el espacio de las variables de respuesta **Y**.
- Coordenadas de sitios – **“lc”** (“*linear combinations of explanatory variables*”). Son combinaciones lineales de variables de restricción: coordenadas de los sitios en el espacio de las variables explicativas **X**. Son las coordenadas ajustadas del sitio (“*fitted site scores*”).

- Coordenadas de– “**bp**” (“biplot arrows”). Coordenadas de las puntas de los vectores que representan las variables explicativas. Las correlaciones se calculan entre las variables explicativas y las coordenadas ajustada de sitio, y luego estas correlaciones se transforman para producir las coordenadas del biplot.
- Centroides “**cn**” (“centroids of factor constraints”). Coordenadas de centroides para los niveles variables que sean factores.

Tipos de Escalamiento (código de R en negrita)

- Ninguno – “**0**”. No hay escalamiento.
- Sitios – “**1**”. Las coordenadas de sitios se escalan por autovalores, mientras que el otro conjunto de puntajes se deja sin escala.
- Especies – “**2**”. Las coordenadas de especies se escalan por autovalores, mientras que el otro conjunto de puntajes se deja sin escala.
- Simétrico – “**3**”. Sitios y Especies se escalan simétricamente por la raíz cuadrada de los autovalores.

El AFM fue realizado mediante R dado que CANOCO no realiza este análisis.

Con la función *cca()* se realizó el ACC y con la función *rda()* se realizó el RDA, estas funciones son similares en CANOCO, aunque la implementación es completamente diferente (Oksanen *et al.*, 2019). Estas funciones de R se basan en el algoritmo de Legendre y Legendre (2012): en *cca()* la matriz transformada de datos Chi-cuadrado se somete a una regresión lineal ponderada con variables restringidas, y sobre los valores ajustados se realiza un análisis de correspondencia mediante descomposición de valores singulares (DVS) (Punto 2.3.3.1. “b” y “c” de esta tesis). La función *rda()* es similar, pero utiliza una regresión lineal ordinaria no ponderada y DVS no ponderada (Punto 2.3.2.1. “a” y “b” de esta tesis). Legendre y Legendre (2012) en la Tabla 11.5 (página 650) dan un esqueleto del algoritmo del RDA de **vegan**, y el algoritmo de ACC es similar, pero implica la estandarización por pesos de fila y columna (Oksanen *et al.*, 2019).

Las sentencias utilizadas en R se especifican en el Anexo A2.

CAPÍTULO 4. Resultados y Discusión

4.1. Biomonitorio Cuenca Lacar - Hua Hum (LHH)

4.1.1. Resultados LHH

La proporción de ceros de la matriz **Y**, con datos de frecuencias absolutas de familias de invertebrados, fue de 0,65. La matriz **Y** fue transformada mediante la transformación “Hellinger” y se utilizó transformada para todos los análisis.

Existen situaciones donde se quiere reducir el número de variables explicativas, ya sea porque son demasiadas o porque lo requieren los análisis. A menudo es aconsejable eliminar una de las variables interrelacionadas en base a una decisión ecológica en lugar de estadística. En los casos en que están presentes varias variables explicativas correlacionadas, sin razones claras a priori para eliminar una u otra, se pueden examinar los factores de inflación de varianza (VIF). Los factores VIF miden cuánto se infla la varianza de la regresión o los coeficientes canónicos por la presencia de correlaciones entre las variables explicativas. Esto mide, de hecho, la inestabilidad del modelo de regresión. Como regla general, ter Braak y Smilauer (1998) recomiendan que las variables que tienen un VIF mayor que 20 se eliminen del análisis. Las variables se deben eliminar de una en una y se vuelve a calcular el análisis, ya que el VIF de cada variable depende de todas las demás. En este caso, básicamente fueron eliminados iones del análisis, dado que son variables altamente correlacionadas y cuya información química se encuentra incluida en el dato de pH, alcalinidad y conductividad.

4.1.1.1. Análisis indirecto del gradiente: AFM – Monitoreo Lacar - Hua Hum (AFM LHH)

El primer análisis realizado fue un AFM de **X** e **Y**, ejecutado con el programa R, dado que CANOCO no realiza este análisis. Cada una de las matrices se identificó como un grupo, denominado “Cuanti.lhh” para **X** y “Frec.hel.lhh” para **Y**. Entre los dos primeros ejes, el porcentaje de inercia explicada fue de 35,5 % (Tabla 4.1).

Tabla 4.1. Inercia de los cuatro primeros ejes del AFM LHH.

	Dim.1	Dim.2	Dim.3	Dim.4
Varianza	1,720	0,653	0,515	0,489
% de varianza	25,740	9,771	7,701	7,322
% acumulado de varianza	25,740	35,511	43,212	50,354

En la Tabla 4.2 se presentan los estadísticos que se utilizaron para evaluar la representación del AFM por grupos. La contribución del grupo “Cuanti.lhh” a la conformación del primer eje (Dim. 1) fue de 49,9 y del segundo eje (Dim. 2) 51,5; mientras que la contribución para el grupo “Frec.lhh” fue de 50,1 para el primer eje y 48,5 para el segundo. Por otro lado, la calidad de representación de cada grupo en el subespacio se definió a través del coseno cuadrado (\cos^2), resultando 0,56 y 0,43 para los grupos “Cuanti” y “Frec” con el primer eje, y significativamente menores con el segundo eje (0,09 y 0,06).

Tabla 4.2. Calidad de representación del AFM LHH por grupos. Coord.: coordenadas, Ctr.: contribución, Cos2.: coseno cuadrado.

Grupos	Coord. Dim.1	Ctr. Dim.1	Cos2 Dim.1	Coord. Dim.2	Ctr. Dim.2	Cos2 Dim.2
Cuanti.lhh	0,859	49,941	0,566	0,336	51,525	0,087
Frec.lhh	0,861	50,059	0,435	0,317	48,475	0,059

En la Tabla 4.3 se presentan los resultados de los coeficientes Lg y RV para los grupos “Cuanti” y “Frec” y para la estructura consenso del AFM. De acuerdo con el coeficiente RV, ambos grupos se relacionaron más con la configuración de consenso (0,86 y 0,89) que, entre ellos, con un valor de RV igual a 0,54 que indica que las variables de un grupo están medianamente correlacionadas con la del otro grupo.

Tabla 4.3. Coeficientes Lg y RV entre grupos. AFM LHH.

		Cuanti.lhh	Frec.lhh	AFM
Coeficiente Lg	Cuanti.lhh	1,305		
	Frec.hel.lhh	0,801	1,030	
	AFM	1,224	1,456	1,558
Coeficiente RV	Cuanti.lhh	1,000		
	Frec.hel.lhh	0,540	1,000	
	AFM	0,860	0,890	1,000

La Figura 4.1 permite comparar globalmente los grupos y se cuantifica la relación entre grupos y con los ejes principales, a través del coeficiente Lg. La relación con el primer eje del AFM fue similar para ambos grupos y cercana a 1, y similar para ambos grupos, pero baja, en relación con el segundo eje del AFM, lo que indica que estos grupos tienen el primer componente del AFM en común, mientras que la segunda dimensión del AFM no está tan relacionada con ninguna dimensión en particular de cada grupo. Ambos grupos se dispusieron próximos en el gráfico, lo que significa que inducen las mismas estructuras en los individuos.

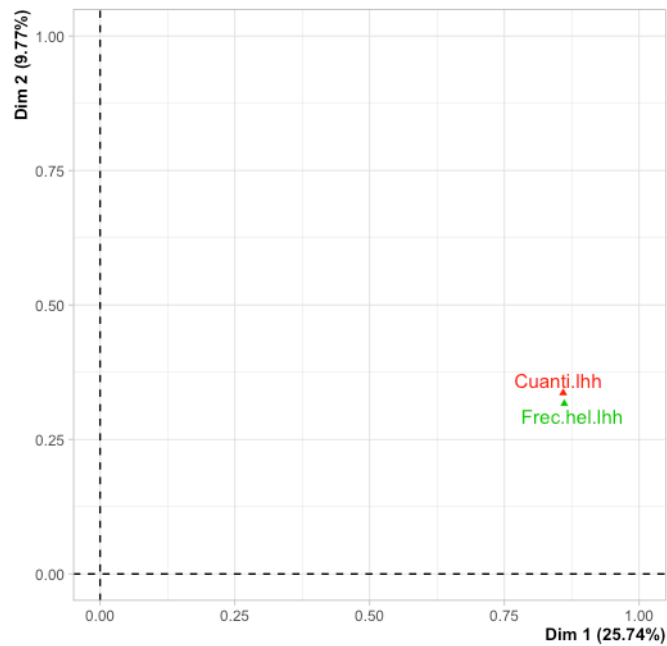


Figura 4.1. Representación de los grupos conformados por las matrices **X** (Cuanti.lhh) e **Y** (Frec.hel.lhh). Plano principal del AFM LHH.

En la Figura 4.2 se puede observar la representación de los sitios en el plano principal del AFM y sus puntos parciales. Se detallan las abreviaturas de sitios en la Tabla I del Anexo. Varios sitios tuvieron representaciones sumamente diferentes de acuerdo con cada uno de los dos grupos representados por las matrices **X** e **Y**, entre ellos: 01PP0, 02PP2, 03PP 04PP3, 20A5, 21A5, 22A5, 23T3,24T3, 25A10, 26A11, 27A11. Los sitios con vectores más largos son los que presentaron las mayores inercias: 25A10, 26A11, 27A11, 01PP0, 02PP2, 03PP 04PP3. Se calculó la inercia para los puntos parciales obteniéndose el valor de 0,88 para el primer eje y 0,53 para el segundo eje, indicando que para el primer eje los puntos parciales son similares unos con otros, mientras que no lo son en referencia al segundo eje. Esto se visualiza observando la dirección de mayor longitud en la que se encuentran los vectores de puntos parciales.

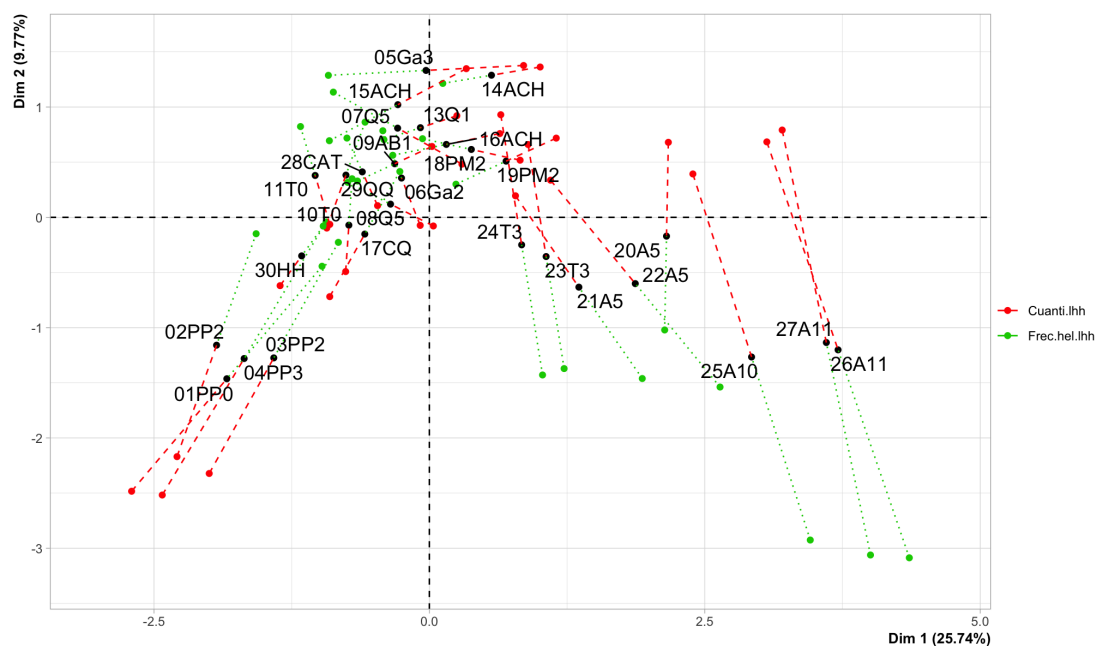


Figura 4.2. Representación de sitios con baricentros y puntos parciales. Plano principal del AFM LHH.

En la Figura 4.3 se presenta el círculo de correlación de las variables ambientales en función de las dos primeras dimensiones del AFM. Las variables están representadas por flechas y su importancia por la longitud de estas. La relación entre las variables está definida por el ángulo que forman las respectivas proyecciones sobre los ejes del AFM.

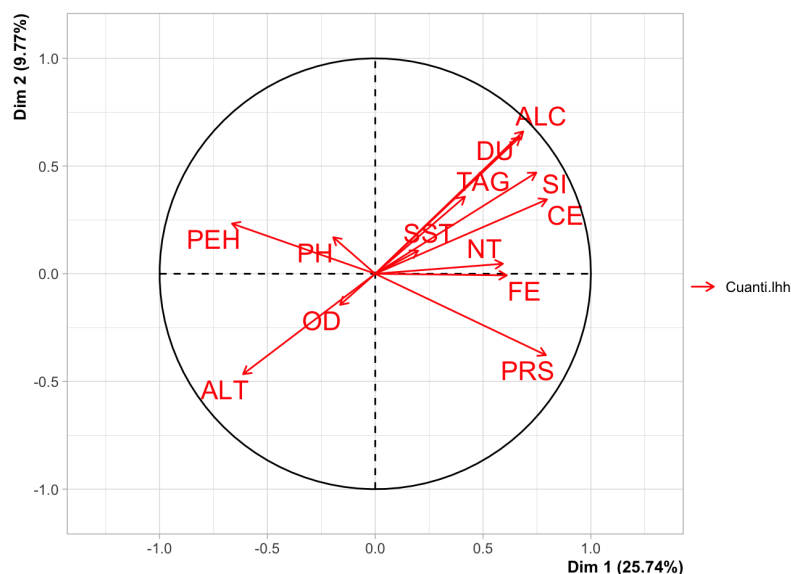


Figura 4.3. Círculo de correlación de variables ambientales de la matriz **X**. Plano principal del AFM LHH.

Se calcularon las correlaciones de las variables con los dos primeros ejes del AFM. En la Tabla 4.4 se presentan las variables que presentaron valores de correlación significativos (p -valor $< 0,05$). Para el primer eje las variables ambientales más correlacionadas positivamente fueron CE, PRS, SI, ALC y negativamente PEH y ALT. Mientras que las familias más correlacionadas positivamente fueron Chironomidae, Glossiphoniidae, Hyalellidae, Tubificidae y negativamente la familia Gripopterygidae. Para el segundo eje las variables ambientales con mayores correlaciones fueron ALC y DU, y algunas familias con correlaciones positivas y negativas iguales o menores a 0,5.

Tabla 4.4. Correlaciones significativas (p -valor $< 0,05$) de variables ambientales y taxones con los dos primeros ejes del AFM LHH.

Dim.1	Correlación	Dim.2	Correlación
CE	0,796	ALC	0,661
PRS	0,791	DU	0,634
Chironomidae	0,776	Gyrinidae	0,552
SI	0,747	Sericostomidae	0,473

Glossiphonidae	0,709	Hydropsychidae	0,472
Hyaellidae	0,702	SI	0,470
Tubificidae	0,688	Tipulidae	0,439
ALC	0,686	Leptophlebiidae	0,436
DU	0,671	PRS	-0,377
ALT	-0,613	Limnaeidae	-0,396
PEH	-0,664	Chironomidae	-0,459
Gripopterygidae	-0,710	ALT	-0,466

Para conseguir una mejor interpretación de la ordenación de los sitios, en el plano principal, en la Figura 4.4 se realizó una categorización de los sitios en 3 subcuencas que se indican con distintos colores, y se agregó un círculo de color a cada etiqueta de familia de macroinvertebrados que indica su grado de sensibilidad a la contaminación (azul: alto, verde: moderado, amarillo: bajo; elaboración propia). Además, con flechas azules se incorporaron los gradientes de las variables continuas de acuerdo con la importancia de sus cargas sobre los dos primeros ejes del plano principal, tomando como referencia los valores que superen los $2/3$ (0,7) de la carga máxima (Beavers *et al.*, 2013, Garson, 2013). Detalles de abreviaturas de sitios y familias se encuentran en las Tablas I y III del Anexo, y en las Tablas IV y V se encuentra el resumen de los valores observados de las variables ambientales en general y por sitio.

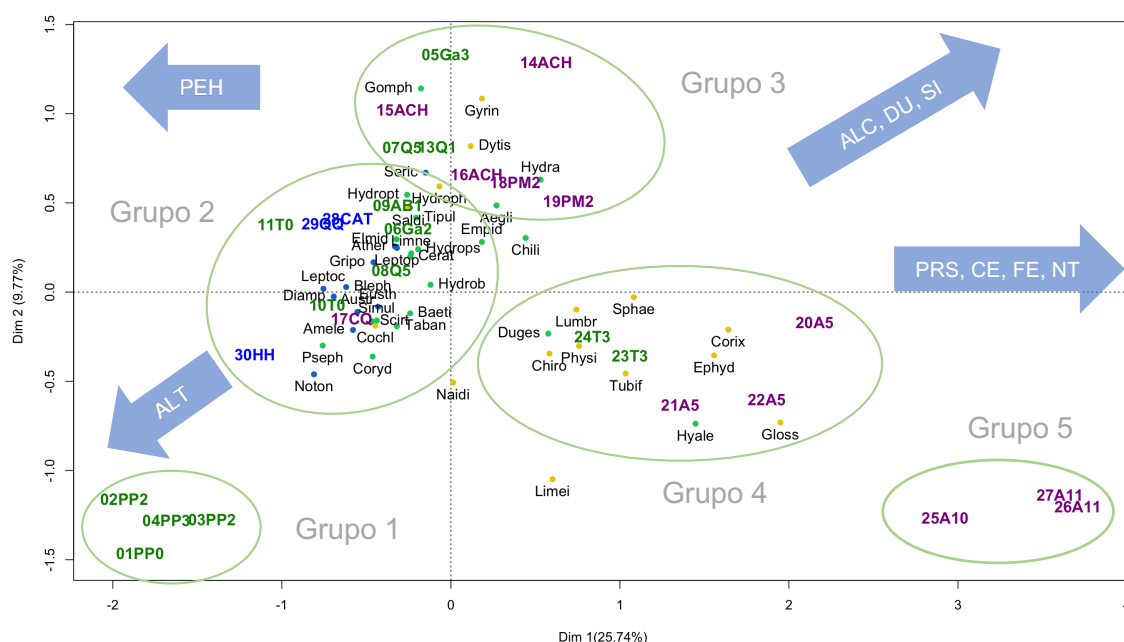


Figura 4.4. Caracterización conjunta del AFM LHH en el plano principal.

Representación de individuos (sitios) y frecuencias (familias de macroinvertebrados). Colores de etiquetas de sitios: subcuencas (violeta: Pocahullo, verde: Trabunco-Quitrahue, azul: Lacar). Colores de círculos: grado de sensibilidad de familias a la contaminación (azul: alto, verde: moderado, amarillo: bajo).

Se identificaron 5 grupos principales, al grupo 1 se asocian los sitios coincidentes con las cabeceras de cuenca (01PP0, 02PP2, 03PP 04PP3), con altos valores de PEH, y bajos valores de PRS, CE, FE, NT, ALC, DU y SI, mientras que, en el otro extremo, en el grupo 5 se encuentran los sitios más asociados a la desembocadura del Arroyo Pocahullo sobre el lago Lacar (25A10, 26A11, 27A11) con altos valores de PRS, CE, FE, NT, valores medios de ALC, DU y SI, y mínimos valores de PEH.

Los grupos centrales en la configuración (2, 3 y 4), se caracterizaron por asociarse con valores medios de las variables ambientales, dentro de eso con el grupo 4 asociado a mayores valores de PRS, CE, FE y NT y menores valores de PEH.

Si bien los grupos extremos (1 y 5) no se asocian directamente con ninguna de las familias de macroinvertebrados, se observa una distribución de familias más tolerantes a la contaminación (por ejemplo: Glossiphoniidae, Corixidae, Ephydriidae, Tubificidae), del lado derecho en el plano de ordenación, y familias más sensibles hacia el centro de la ordenación (por ejemplo: Notoneumoridae, Blephariceridae, Leptoceridae, Ameletopsidae), que pueden ser asociadas a dichos grupos.

4.1.1.2. Análisis directo del gradiente: RDA - Monitoreo Lacar Hua Hum (RDA LHH)

El DCA de la matriz **Y** arrojó un valor de longitud del primer eje igual a 2,61 DE con lo que se utilizó el RDA como análisis directo del gradiente.

El RDA de las matrices **X** (29x13) e **Y** (29x46) fue realizado con CANOCO y R. Cuando el análisis fue ejecutado con CANOCO, se importó la matriz transformada, dado que en el menú de opciones no aparece como opción la transformación de Hellinger para la matriz **Y**. Mientras que ambos programas utilizaron la matriz de datos **X** estandarizada.

Como primer paso se evaluó la significancia del RDA mediante una prueba de permutación. La prueba de significancia del RDA, arrojó un resultado significativo con pseudo-F = 2,1516; pero ligeramente distinto p-valor informado por R (p-valor = 0,002) y CANOCO (p-valor = 0,001).

La varianza total reportada en R fue de 0,3788 (representa una proporción de 1) y contiene la proporción de varianza de **Y** explicada por las variables **X** (proporción restringida, 0,65), la varianza no explicada de **Y** (proporción no restringida, 0,35). En CANOCO la variación total fue de 10,607; el resultado se muestra como un 65,09% de variación asociada a las variables explicativas (R^2) y 34,84% a la variación explicada ajustada (R^2_a).

Mediante el uso del R también se calcularon los estadísticos R^2 y R^2_a , con los mismos resultados, $R^2 = 0,65$ (equivalente a la proporción de varianza restringida) y $R^2_a = 0,34$.

En la Tabla 4.5 se presenta el resumen de los cuatro primeros ejes del RDA, autovalores, proporciones explicadas y la proporción acumulada de cada eje canónico (cada eje canónico = cada variable de restricción, en este caso, las variables ambientales).

Tabla 4.5. Resumen de los 4 primeros ejes canónicos del RDA LHH. Salidas de R y CANOCO.

Resultados		RDA1	RDA2	RDA3	RDA4
R (vegan)	Autovalor	0,133	0,032	0,018	0,016
	Proporción explicada	0,351	0,084	0,049	0,043
	Proporción acumulada	0,351	0,435	0,484	0,527
<hr/>					
CANOCO	Autovalor	0,351	0,084	0,049	0,043
	Variación explicada (acumulada en %)	35,100	43,490	48,370	52,720
	Correlación pseudo-canónica	0,912	0,850	0,913	0,832
	Variación explicada ajustada (acumulada)	53,920	66,810	74,320	81,000

Mientras que CANOCO informa los autovalores escalados (su suma es igual a 1), R presenta autovalores que son una variación total sin escala. No obstante, los valores son exactamente lo mismo, ya que los autovalores de CANOCO son los de R por una constante (2,639).

Junto con la salida del análisis del RDA, CANOCO reportó diferentes estadísticos diagnósticos relacionados tanto con las variables ambientales como con las frecuencias de familias. En la Tabla 4.6 se presentan los VIF (factores de inflación de varianza), las correlaciones de cada variable ambiental con cada uno de los dos ejes principales y sus coordenadas.

Tabla 4.6. Valores de factor de inflación de la varianza (VIF), correlación y coordenadas de las variables ambientales con los dos primeros ejes canónicos.

RDA LHH.

	VIF	Correlación RDA1	Correlación RDA2	Coordenadas RDA1	Coordenadas RDA2
CE	11,672	0,608	-0,196	0,666	-0,231
OD	3,272	-0,026	-0,111	-0,029	-0,130
PH	1,640	-0,279	-0,304	-0,306	-0,357
TAG	5,556	0,293	-0,190	0,321	-0,221
ALT	11,837	-0,437	0,314	-0,480	0,370
NT	2,644	0,498	-0,131	0,546	-0,154
PRS	3,117	0,788	0,075	0,863	0,088
SST	1,576	0,169	0,022	0,186	0,026
FE	2,767	0,525	-0,188	0,575	-0,221
SI	6,610	0,476	-0,212	0,522	-0,250
DU	27,798	0,365	-0,122	0,400	-0,143
ALC	23,682	0,343	-0,183	0,376	-0,216
PEH	2,480	-0,724	-0,194	-0,793	-0,229

Tanto CANOCO como R ofrecen una serie de estadísticos para evaluar la bondad y calidad de representación de todos los elementos en la ordenación. En R se obtuvieron con el paquete **vegan** mediante la función *goodness()*.

Para el caso de las familias de macroinvertebrados las 20 familias con mejor ajuste fueron (nombre completo en Tabla III del Anexo): Tubificidae, Chironomidae, Glossiphonidae, Limnaeidae, Saldidae, Tabanidae, Elmidae, Hydropsychidae, Scirtidae, Simuliidae, Tipulidae, Baetidae, Hyalellidae, Hydrobiosidae, Limnephilidae, Ameletopsidae, Athericidae, Austroperlidae, Gripopterygidae, Leptophlebiidae.

La representación de los sitios de acuerdo con la bondad de su ajuste en el RDA fue, en orden decreciente de importancia: 25A10, 26A11, 27A11, 30HH, 15ACH, 01PP0, 20A5, 06Ga2, 04PP3, 05Ga2, 29QQ, 22A5, 02PP2, 21A5, 11T0, 13Q1,

09AB1, 14ACH, 23T3, 03PP2, 07Q5, 10T0, 28CAT, 24T3, 17CQ, 08Q5, 16ACH, 19PM2, 18PM2.

Para visualizar los resultados del RDA, se extrajeron los dos primeros ejes canónicos y se graficaron los biplots y el triplot con escalado 2 (Figuras 4.5 y 4.6).

En la Figura 4.5 se presenta el biplot con escalado 1 (a), que relaciona sitios y variables ambientales y el biplot escalado 2 (b), que relaciona las familias de macroinvertebrados con las variables ambientales. Estos diagramas son salidas de CANOCO, mientras que en el Anexo A2 se presentan las sentencias para realizar los mismos con R. Detalle de abreviaturas de familias en Tabla III del Anexo.

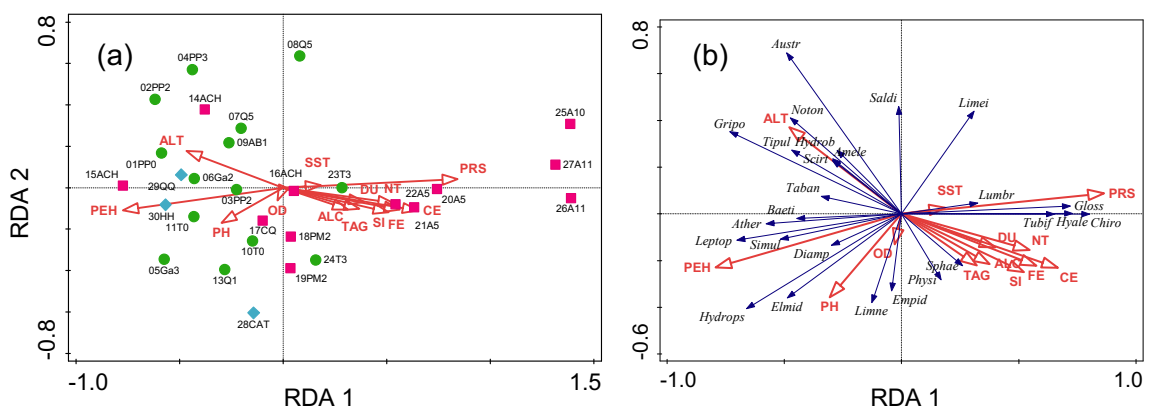


Figura 4.5. RDA Biplots, primeros dos ejes canónicos. Salida CANOCO. (a) Biplot de sitios y variables ambientales, escalado 1 (subcuencas: círculo verde, Trabunco-Quitrahue; rombo celeste, Lacar-Hua Hum; cuadrado rojo, Pocahullo). (b) Biplot de familias y variables ambientales, escalado 2.

Las variables mayormente asociadas al primer eje canónico del RDA fueron, con correlación positiva el PRS, CE y FE, y con correlación negativa el PEH, coincidiendo con el resultado de AFM, y con mayor importancia en la correlación con el segundo eje canónico del RDA se asoció la ALT positivamente, y el PH y SI negativamente, mientras que para el AFM el segundo eje estuvo relacionado principalmente a la ALT.

Ambos biplots de la Figura 4.5 representan dos escalas diferentes, por lo que permite visualizar diferencialmente la relación entre las variables ambientales y los sitios en (a) y las variables ambientales y las familias en (b) y ayudan a la interpretación cuando los elementos del gráfico son numerosos.

La Figura 4.6 representa el triplot con escalado 2 (CANOCO), este escalado pone foco en la correlación entre especies. Para una mejor visualización se incluyeron las familias mejor representadas, de acuerdo con su distancia al centro del plano de ordenación.

Se definieron grupos con características similares. El primer eje representó un gradiente asociado fundamentalmente al grado de contaminación o degradación del ambiente, observándose en concordancia con eso, familias de macroinvertebrados con diferentes grados de sensibilidad a la contaminación. En el grupo 1 las familias mas representativas fueron Austroperlidae y Gripopterygidae, coincidiendo con los sitios de altas cumbres, y siendo característicos bajos valores de PRS, CE, FE, SI, PH y altos valores de PEH. El grupo 2, con menores valores de ALT, y mayores valores de PH y SI, compartió similares características de las variables importantes sobre el primer eje canónico (bajos valores de PRS, CE, FE, SI, PH y altos valores de PEH) y se asociaron familias como Elmidae, Hydropsychidae y Leptophlebiidae.

Se asociaron a un tercer grupo los sitios 25A10, 26A11, y 27A11, 20A5, 21A5 y 22A5, con los mayores valores de PRS, CE y FE y menores valores de PEH, coincidiendo los tres primeros sitios con el grupo 5 del AFM. A este grupo se asociaron directamente las familias Chironomidae, Tubificidae, Glossiphoniidae y Hyalellidae, relacionados con sitios altamente urbanizados de la cuenca Pocahullo.

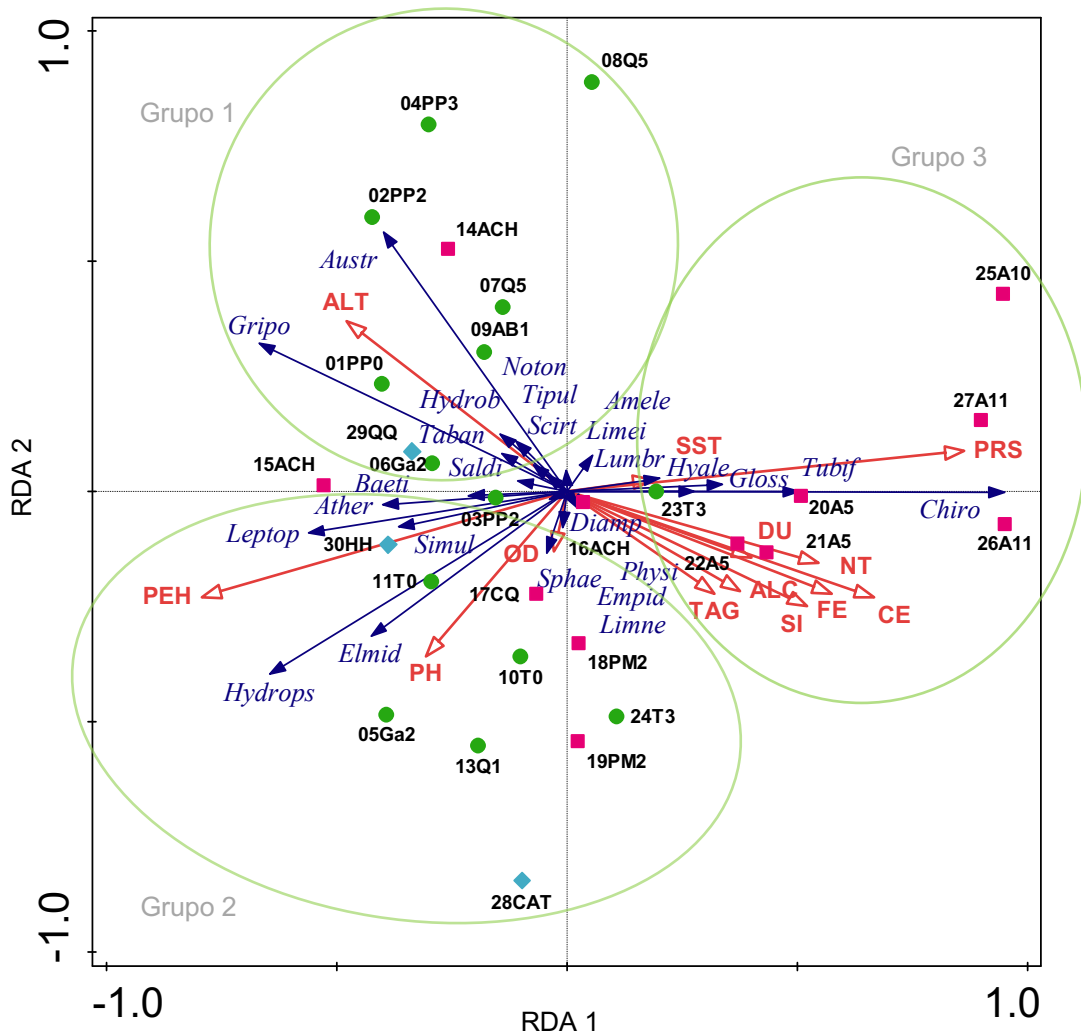


Figura 4.6. Triplot del RDA LHH, primeros dos ejes canónicos. Salida del CANOCO. Escalado 2. Relación sitio-familia-ambiente. Los vectores rojos representan las variables predictoras ambientales, los vectores azules las frecuencias de familias de macroinvertebrados y los puntos los sitios, identificando con color las subcuencas (fucsia: Pocahullo, verde: Trabunco-Quitrahue, azul: Lacar).

El paquete **vegan** del R ofrece varias funciones para seleccionar el mejor modelo reducido, las funciones *ordiR2step()* y *ordistep()* y la función *forward.sel()* del paquete **adespatial** (solo para RDA). El proceso de selección de variables a través del *ordiR2step* incorporó tres variables significativas: PRS, PEH, ALT. Por su lado, CANOCO también ofrece la posibilidad de realizar una selección de variables de

acuerdo con un proceso de selección forward en el que se va indicando la significancia de cada variable que ingresa al modelo y en el que el usuario puede decidir que variables incorporar de acuerdo con su criterio.

4.1.2. Discusión de aspectos biológicos

En este trabajo quedó manifestado que existe un gradiente ambiental en la cuenca Lacar – Hua Hum evidenciado por las variables que cobraron mayor relevancia en la ordenación, el fósforo (PRS), la conductividad eléctrica (CE) y el puntaje de hábitat (PEH), y asimismo definido por la variación de diferentes familias de macroinvertebrados bentónicos, que no necesariamente responden a la categorización en subcuencas, sino al grado de antropización generada por el crecimiento de la población y los asentamientos en la localidad de San Martín de los Andes y sus alrededores. Las técnicas de ordenación se han utilizado con éxito para interpretar la estructura de la comunidad de macroinvertebrados.

El porcentaje de inercia explicado por el AFM fue de un 35,5% y 43,49% con el RDA entre los dos primeros ejes, permitiendo encontrar un gradiente ambiental en el primer eje asociado al grado de degradación o contaminación.

En cuanto a la ordenación de sitios, el AFM y el RDA arrojaron resultados similares. El AFM permitió identificar por lo menos 5 grupos de sitios con diferentes características, dado que define un gradiente de contaminación entre los sitios prístinos en las cabeceras de cuencas y los más impactados. El primer grupo contuvo el sitio de referencia o blanco, Vertiente del Mocho, el cual se encuentra en las altas cumbres y es la vertiente que define la naciente de la subcuenca Trabunco - Quitrahue, por lo que este grupo es el que define los sitios de mejor calidad ambiental. Sucesivamente, cada grupo se asoció a diferentes grados de contaminación o degradación llegando al quinto grupo en el que se incorporaron los sitios de la subcuenca Pocahullo antes de desembocar al lago Lacar, los cuales han recibido los aportes de todo el pueblo que atraviesan y en consecuencia resultaron los más impactados y de peor calidad ambiental. Este último se asocia a aguas duras, con alta conductividad, y altos valores de fósforo y nitrógeno revelando una evaluación desfavorable del hábitat de este arroyo (Muñiz Saavedra *et al.*, 2015; Pereyra *et al.*, 2016)

En el RDA no toma tanta relevancia el sitio de referencia PP0, Vertiente del Mocho, pero igualmente la ordenación de grupos de sitios se da de acuerdo con un gradiente de degradación del ambiente en el primer eje, similar al del AFM.

Miserendino (2001) también describe relaciones claras de factores ambientales con macroinvertebrados, viéndose afectados principalmente por la velocidad del agua, la conductividad, el tamaño del sustrato y la abundancia de plantas acuáticas, y secundariamente por la latitud y la temperatura del agua. Sus resultados indican diferencias en los conjuntos de macroinvertebrados dependiendo de la vegetación de la cuenca.

Asimismo, Miserendino *et al.* (2008) mencionan que los arroyos urbanos patagónicos con frecuencia experimentan alteraciones en sus regímenes hidrológicos, canales extensos y modificaciones ribereñas (incluyendo dragado y limpieza regular) y aporte de contaminantes orgánicos y sedimentos, proponiendo un enfoque que combina la evaluación biológica directa con la física, química y análisis ambiental para diagnosticar la degradación.

Se encontraron similares resultados que Mauad *et al.* (2015) quienes estudiaron el sistema lótico Challhuaco-Ñireco encontrando a través de un análisis de ordenamiento (RDA) que las distribuciones de los ensambles de especies de macroinvertebrados responden al gradiente climático y topográfico (temperatura y elevación), pero también se asocian con variables relacionadas al impacto antrópico (conductividad, contenido de nitratos y fosfatos). Encontraron que los ensambles en cabecera están representados por insectos sensibles, mientras que los taxones tolerantes como Tubificidae, Lumbriculidae, Chironomidae y el crustáceo *Aegla* sp. son dominantes en los sitios urbanizados.

En la Tabla I del Anexo se especifican los valores del índice biótico BMPS para cada sitio muestreado. El índice BMPS (“*Biotic Monitoring Patagonian Streams*”) adaptado para Patagonia por Miserendino y Pizzolón (1999), es un índice biótico que se utiliza para la evaluación ecológica de la calidad de agua, combinando la presencia de los diferentes macroinvertebrados con un valor de sensibilidad o tolerancia/intolerancia a la contaminación. El índice se calcula sumando las puntuaciones asignadas a cada una de las familias de macroinvertebrados

identificadas en el sitio de muestreo. El puntaje se asigna una sola vez por familia, independientemente de la cantidad de individuos recolectados, valores más altos de puntaje indican baja tolerancia de la familia de macroinvertebrados a la contaminación. La suma total de los puntajes de las familias presentes en el sitio muestreado le otorgan un valor de índice que permite determinar la calidad del agua de acuerdo con las categorías especificadas (Tablas I y II del Anexo).

Si bien los grupos “Cuanti.lhh” y “Frec.lhh” se dispusieron próximos (ver Figura 4.1) indicando que inducen las mismas estructuras en los sitios, en la configuración de consenso del AFM, no lograron identificarse claramente las asociaciones de familias con los grupos de sitios más extremos, grupos 1 y 5, pero se evidenció una distribución de familias con las más sensibles a la contaminación sobre los cuadrantes izquierdos, y las más tolerantes en los cuadrantes derechos del plano principal. El RDA si permitió identificar las familias más representativas asociadas a los grupos de sitios, por ejemplo, familias asociadas a la menor sensibilidad (más tolerantes) a la contaminación, cuyo puntaje oscila entre 1 y 3 para el índice BMPS (Miserendino y Pizzolón, 1999) (Tubificidae, Chironomida, Glossosomatidae) y estuvieron asociadas al grupo 5 del AFM y grupo 3 del RDA, mientras que familias con mayor sensibilidad a la contaminación como Austroperlidae, Gripopterygidae y Leptophlebiae estuvieron asociados al grupo 1 del AFM y grupo 1 del RDA, coincidiendo con mejores condiciones de hábitat de acuerdo al índice de evaluación PEH.

La evaluación del hábitat se define como la evaluación de la estructura del hábitat físico circundante que influye en la calidad del recurso hídrico y la condición de la comunidad acuática residente para los arroyos. Un enfoque integral para evaluar la estructura del hábitat incluye una evaluación de la variedad y calidad del sustrato, la morfología del canal, la estructura del banco y la vegetación ribereña (Barbour *et al.*, 1996; Barbour *et al.*, 1999).

Un estudio realizado en el río Medina, Tucumán, utiliza el análisis de correspondencia canónica (ACC) y encuentran un gradiente que se correlaciona significativamente ($r = 0,663$, $p < 0,05$) con la conductividad a la que parecen responder los hirudíneos, y los insectos como Elmidae (Coleoptera) y Ephydriidae (Diptera), e inversamente correlacionado ($r = -0,854$, $p\text{-valor} < 0,05$) con el

oxígeno disuelto, causa a la que sí responde Leptohyphidae (Ephemeroptera). El análisis señala que los nutrientes son difíciles de seguir en sus ciclos y fuentes, con fuertes variaciones temporales y destacan que los bioindicadores no son capaces de detectar un aumento de los mismos (fosfatos y nitratos especialmente) por actividades antrópicas. Menciona que su efecto es indirecto, pero por ello no menos importante (Fernández *et al.*, 2006).

En coincidencia con Miserendino *et al.* (2008) se menciona que los análisis de la comunidad de macroinvertebrados son valiosos para la detección del deterioro de los arroyos en la Patagonia y se recomiendan para monitorear los esfuerzos de recuperación biológica, conservación y restauración en el futuro. Similares resultados son encontrados por Macchi *et al.* (2020), quienes evidencian la potencial aplicación de los macroinvertebrados para el diseño de herramientas de monitoreo y vigilancia ambiental del río Negro.

Los sistemas de cabecera, como los estudiados en este trabajo (sitios PP0, PP2, PP3) proporcionan hábitats únicos, tienen diferentes características en sus atributos físicos y químicos, que configuran una serie correspondiente de entornos únicos para los organismos. No debe haber ninguna duda de que los sistemas de cabecera soportan sistemas biológicos únicos sobre la base de la composición de especies, la estructura del ensamble y las tasas de diversos procesos del ecosistema (Richardson y Danehy, 2007).

Los cambios en el uso del suelo de bosques nativos a tierras usadas en agricultura, ganadería y plantaciones forestales, afectan las comunidades acuáticas, ya que la pérdida de vegetación ribereña altera directa e indirectamente la calidad del agua, el tipo de sustrato y afecta las condiciones físicas del hábitat acuático y las comunidades bióticas asociadas, disminuyendo el material alóctono que ingresa al cuerpo hídrico y la disponibilidad de sustratos aportados por los bosques adyacentes y alterando la estructura y la función trófica de las comunidades de insectos acuáticos (Urdanigo *et al.*, 2019). De acuerdo con los resultados de este trabajo, esto se observa principalmente en los sitios en los que han avanzado las urbanizaciones, sobre sectores medios y bajos de la subcuenca Tabunco-Quitrahue y la subcuenca Pocahullo.

La comunidad bentónica permite hacer, en un aspecto aplicado, diagnósticos a dos niveles: a) calidad del agua; b) situación de la cuenca. En el primer caso que es casi una necesidad en nuestros días, permite tomar decisiones de manejo y administración. En el segundo caso, al ampliar la escala a un nivel de paisaje, se reconoce que el resultado final sobre el curso de agua está en la cuenca, con una acumulación de efectos que a su vez interactúan de modos no siempre claros (Fernández *et al.*, 2006).

Domínguez *et al.* (2020) expresan que los resultados del estado de los ecosistemas acuáticos obtenidos con algunos métodos de diagnóstico (por ejemplo, los químicos), por lo general no son fácilmente compatibles con otros (como los biológicos). Algunas veces, mientras que los resultados de los parámetros fisicoquímicos denotan una calidad del agua “aceptable”, los biológicos la reconocen como de baja calidad. Y no son contradicciones: es que se están midiendo distintos aspectos de los sistemas acuáticos tal vez con diferentes objetivos. En este sentido, en este trabajo de tesis se propusieron algunos de los métodos que conjugan los aspectos fisicoquímicos y biológicos, evaluando el ambiente con una visión integral.

4.2. Monitoreo del alga exótica invasora *Didymosphenia geminata* (DIDYMO)

4.2.1. Resultados DIDYMO

La proporción de ceros de la matriz **Y**, con datos de frecuencias relativas porcentuales de especies de algas, fue de 0,75.

4.2.1.1. Análisis indirecto del gradiente: AFM - Monitoreo *D. geminata* (AFM DIDYMO)

Se realizó primeramente un AFM de las matrices **X** e **Y**. Cada una de las matrices se identificó como un grupo, denominado “Cuanti.didy” para **X** y “Frec.didy” para **Y**. Entre los dos primeros ejes, el porcentaje de inercia explicada fue de 27,8 % (Tabla 4.7).

Tabla 4.7. Inercia de los cuatro primeros ejes del AFM DIDYMO.

Autovalores	Dim.1	Dim.2	Dim.3	Dim.4
Varianza	1,666	0,819	0,661	0,631
% de varianza	18,675	9,175	7,405	7,063
% acumulado de varianza	18,675	27,832	35,237	42,300

En la Tabla 4.8 se presentan los estadísticos de evaluación de la representación del AFM DIDYMO. La contribución del grupo “Cuanti.didy” a la conformación del primer eje (Dim. 1) fue de 51,5 y del segundo eje (Dim. 2) 9,99; mientras que la contribución para el grupo “Frec.didy” fue de 48,47 para el primer eje y 90,0 para el segundo. Por otro lado, la calidad de representación definida a través del coseno cuadrado (\cos^2), resultó 0,52 y 0,21 para los grupos “Cuanti.didy” y “Frec.didy” con el primer eje y 0,005 y 0,180 con el segundo eje.

Tabla 4.8. Calidad de representación AFM DIDYMO por grupos. Coord.: coordenadas, Ctr.: contribución, Cos2.: coseno cuadrado.

Grupos	Coord. Dim.1	Ctr. Dim.1	Cos2 Dim.1	Coord. Dim.2	Ctr. Dim.2	Cos2 Dim.2
Cuanti.didy	0,859	51,527	0,526	0,082	9,990	0,005
Frec.didy	0,808	48,473	0,216	0,738	90,010	0,180

En la Tabla 4.9 se presentan los resultados de los coeficientes Lg y RV para los grupos “Cuanti.didy” y “Frec.didy” y para la estructura consenso del AFM. El valor del coeficiente RV entre ambos fue bajo (0,34), indicando muy baja asociación entre las variables que conforman cada grupo, mientras que cada grupo por separado presentó alta correlación con la configuración de consenso (0,74 y 0,89).

Tabla 4.9. Coeficientes Lg y RV entre grupos. AFM DIDYMO.

		Cuanti.lhh	Frec.lhh	AFM
Coeficiente Lg	Cuanti.lhh	1,400		
	Frec.lhh	0,706	3,017	

	MFA	1,264	2,234	2,099
Coeficiente RV	Cuanti.lhh	1,000		
	Frec.hel.lhh	0,344	1,000	
	MFA	0,737	0,888	1,000

En la Figura 4.7 se observa la representación de los grupos con respecto a los dos primeros ejes del AFM. Para el primer eje ambos grupos estuvieron representados de manera similar, sin embargo, solo el grupo “Frec.lhh” presentó alta relación con el segundo eje. Ambos grupos se dispusieron distantes en el gráfico, lo que significa que inducen estructuras diferentes en la ordenación de los individuos.

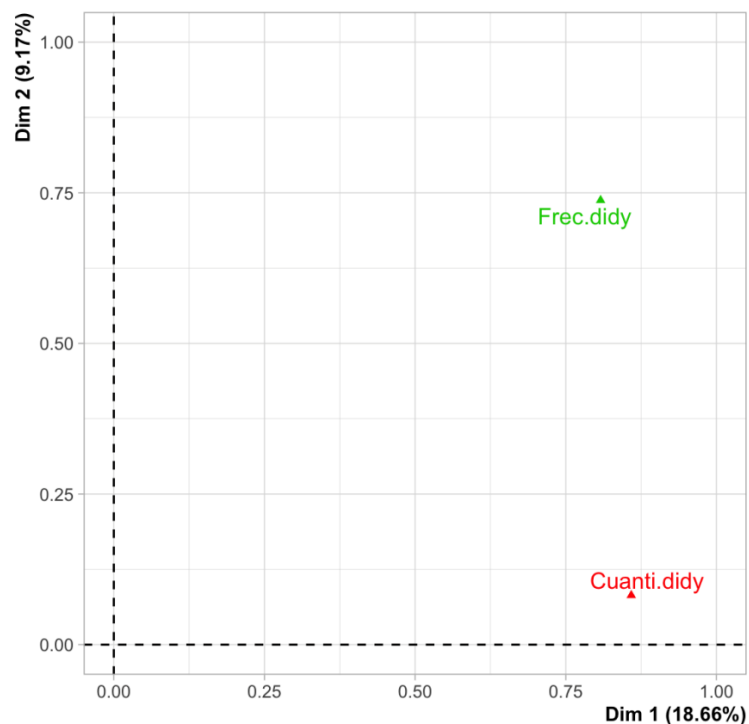


Figura 4.7. Representación de los grupos conformados por las matrices X (Cuanti.didy) e Y (Frec.didy) en el plano principal del AFM DIDYMO.

En la Figura 4.8 se observa la representación de los sitios y sus puntos parciales. En la ordenación de los sitios en el gráfico no se identificaron a priori agrupaciones, viéndose que varios sitios tuvieron representaciones diferentes de acuerdo con cada uno de los dos grupos “Cuanti.didy” y “Frec.didy”. Se calculó la inercia para los puntos parciales obteniéndose el valor de 0,86 para el primer eje y 0,53 para el segundo eje. Las abreviaturas utilizadas para los sitios se detallan en la Tabla VI del Anexo.



Figura 4.8. Representación de sitios con baricentros y puntos parciales en el plano principal. AFM DIDYMO.

En la Tabla 4.10 se presentan los valores de mayores correlaciones significativas de las variables con los dos primeros ejes. Para el primer eje las variables ambientales más correlacionadas positivamente fueron el PT, la ALC, la CE, la TAG y los SST. Mientras que la especie *Calothrix sp.* fue la más correlacionada negativamente. Para el segundo eje, la única variable ambiental significativa, pero baja correlación fue SST, mientras que las especies más correlacionadas fueron *Diatoma sp.* y *Navicula sp.* con correlación positiva y *Lyngbya sp.*, *Calothrix sp.* y *Nostoc sp.* con correlación negativa.

Tabla 4.10. Correlaciones significativas (p-valor < 0,05) de variables ambientales y taxones con los dos primeros ejes del AFM DIDYMO.

Dim.1	Correlación	Dim .2	Correlación
PT	0,734	<i>Diatoma sp.</i>	0,620
ALC	0,724	<i>Navicula sp.</i>	0,369
CE	0,653	SST	0,303
TAG	0,639	<i>Nostoc sp.</i>	-0,329
SST	0,581	<i>Calothrix sp.</i>	-0,515
<i>Calothrix sp.</i>	0,569	<i>Lyngbya sp.</i>	-0,642
<i>Epithemia sp.</i>	0,524		
<i>Diatoma sp.</i>	0,471		
ORDEN	0,371		
OD	-0,500		
PEH	-0,548		
<i>Gomphonema sp.</i>	-0,647		

En la Figura 4.9 se presenta el círculo de correlación de las variables ambientales en función de las dos primeras dimensiones del AFM. En el primer eje se identificó

un gradiente de ALC, CE y TAG con mayores valores sobre la margen derecha del primer eje del plano principal, y menores valores de OD. De acuerdo con las cargas de las variables con los ejes principales, no se distinguieron variables con carga importante sólo sobre el eje 2, y si se encontró un gradiente oblicuo con mayores valores de PEH sobre el cuadrante inferior izquierdo y mayores valores de PT y SST sobre el cuadrante superior derecho.

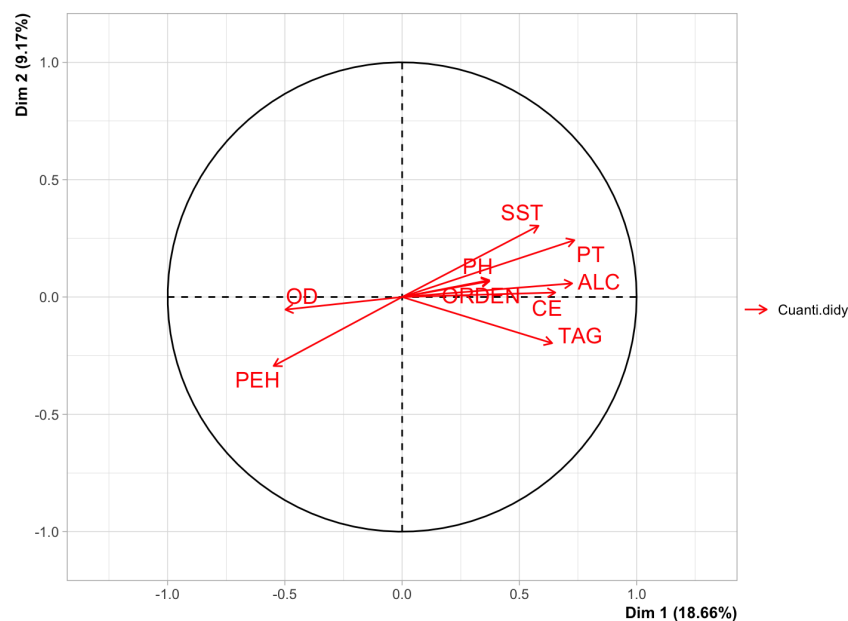


Figura 4.9. Círculo de correlación de variables ambientales de la matriz **X**. Plano principal del AFM DIDYMO.

En la Figura 4.10 se presenta la representación conjunta de sitios y especies del AFM DIDYMO, incorporándose con colores la categorización de los sitios de acuerdo con las cuencas a las que pertenecen (amarillo: Río Neuquén, púrpura: Río Colorado, azul: Río Limay, verde: Lago Lacar). Con flechas azules se agregaron los gradientes de las variables continuas de acuerdo con sus cargas sobre los dos primeros ejes del plano principal (Beavers *et al.*, 2013, Garson, 2013).

Si bien el porcentaje de inercia explicado por el segundo eje no fue grande, pudieron distinguirse tres grupos, teniendo en cuenta las variables más correlacionadas con el primer eje y los gradientes oblicuos definidos por PT, SST

y PEH. El grupo 1 con menores valores de ALC, CE y TAG y mayores valores de OD, y con valores medianos a grandes de PEH, y bajos valores de PT y SST, asociándose con la mayoría de los sitios de las cuencas del río Limay, Hua Hum. A este grupo se asociaron *Asterionella sp.*, *Rhopalodia sp.*, *Gomphoneis sp.*, *Gomphonema sp.*, *Cyclotella sp.*, *Mougeotia sp.*, *Didymosphenia sp.*, *Cosmarium sp.*, *Hannaea sp.*, *Zygnema sp.*

Por su parte, a los grupos 2 y 3 se asociaron la mayoría de los sitios de las cuencas de los ríos Neuquén y los dos sitios de la cuenca del río Colorado, caracterizados por altos valores de ALC, CE, TAG, bajos valores de OD. El grupo 2 se asoció a un amplio rango de valores de PEH, PT y SST, y se relacionaron especies tales como *Diploneis sp.*, *Frustulia sp.*, *Calothrix sp.*, *Lyngbya sp.*, *Nostoc sp.*, *Achnanthes sp.* Mientras que el grupo 3 se asoció a los mayores valores de PT y SST, y a especies tales como *Diatoma sp.* y *Gyrosigma sp.*

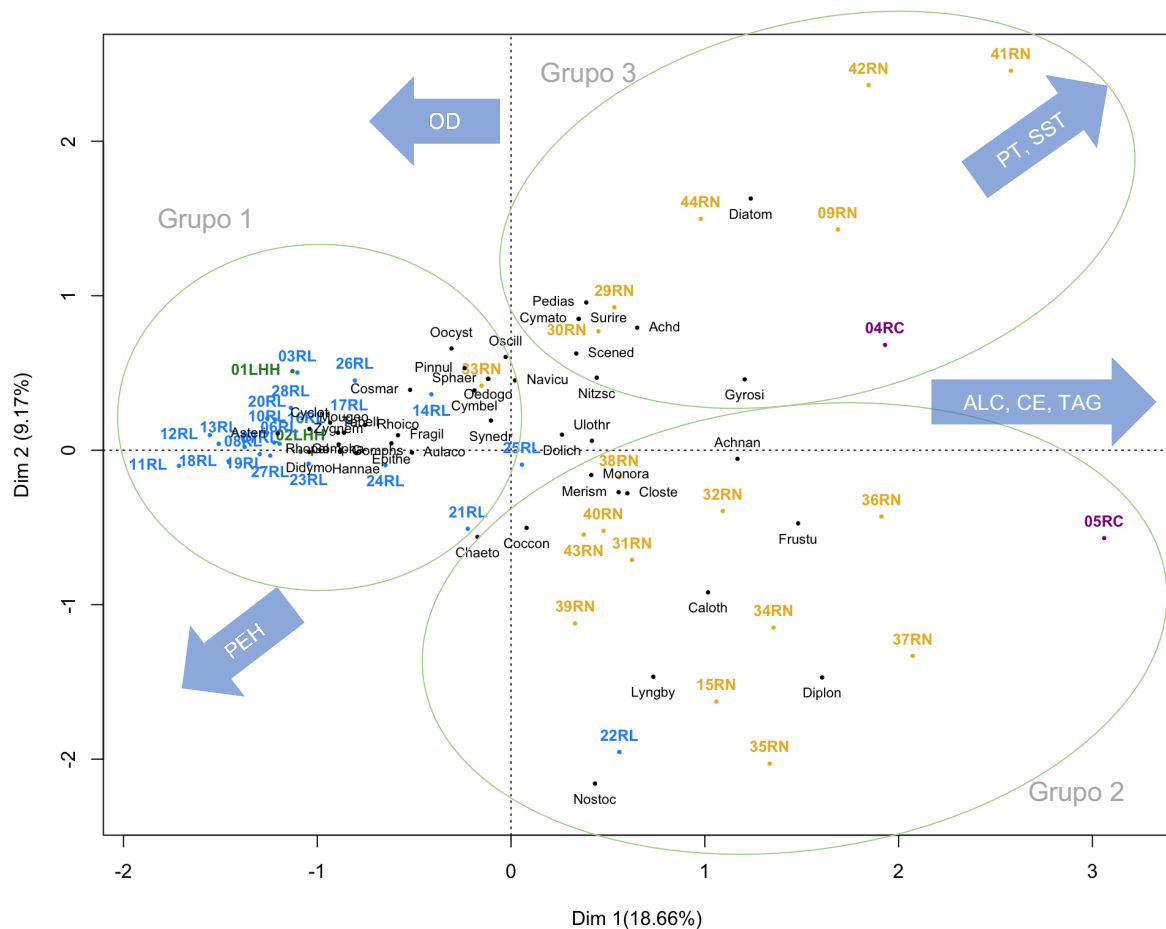


Figura 4.10. Caracterización conjunta AFM DIDYMO. Representación conjunta de individuos (sitios) y frecuencias (especies de algas). Los colores de etiquetas de sitios identifican las cuencas (púrpura: río Colorado, verde: río Hua Hum, amarillo: río Neuquén, azul: río Limay).

4.2.1.2. Análisis directo del gradiente: ACC – Monitoreo Didymo (ACC DIDYMO)

El DCA de la matriz **Y** arrojó un valor de longitud del primer eje igual a 3,22 con lo que se utilizó el ACC como análisis directo del gradiente.

La prueba de significancia de los ejes canónicos del ACC, arrojó un resultado significativo con pseudo-F = 1,4046 y p-valor (CANOCO) = 0,003, p-valor (R) = 0,006.

Los resultados de la partición del Chi-cuadrado escalado arrojaron una varianza o inercia total de 4,219 (restringida igual a 1,144 y no restringida 3,076), representando una proporción restringida de 0,271 y proporción no restringida de 0,729. En CANOCO la variación total también fue expresada como 4,219, lo que representó un 27,10% de variación asociada a las variables explicativas y 7,81% asociada a la variación explicada ajustada.

En la Tabla 4.11 se presentan los autovalores y proporción de varianza explicada de los cuatro primeros ejes del ACC. El porcentaje explicado por el primer eje canónico es de 9,51%, y del segundo eje canónico 5,10%, acumulando entre los dos el 14,61%.

Tabla 4.11. Resumen de los 4 primeros ejes canónicos del ACC DIDYMO.
Salidas de R y CANOCO.

Resultados		ACC1	ACC2	ACC3	ACC4
R (cca vegan)	Autovalor	0,401	0,215	0,171	0,112
	Proporción explicada	0,095	0,051	0,040	0,026
	Proporción acumulada	0,095	0,146	0,186	0,213
CANOCO	Autovalor	0,401	0,2152	0,1712	0,1125
	Variación explicada (acumulada en %)	9,510	14,610	18,670	21,340
	Correlación pseudo- canónica	0,823	0,779	0,730	0,807
	Variación explicada ajustada (acumulada)	35,100	53,920	68,890	78,720

En la Tabla 4.12 se presentan los valores obtenidos de VIF y las correlaciones con los dos primeros ejes de las variables ambientales.

Tabla 4.12. Valores de factores de inflación de la varianza (VIF), de correlación y coordenadas de las variables ambientales con los dos primeros ejes canónicos.

ACC DIDYMO.

	VIF	Correlación ACC1	Correlación ACC2	Coordenadas ACC1	Coordenadas ACC2
CE	1,833	0,482	0,038	-0,586	0,049
OD	1,826	-0,499	-0,028	0,607	-0,036
PH	1,770	0,293	-0,076	-0,356	-0,098
TAG	1,876	0,589	0,240	-0,716	0,308
ORDEN	1,227	0,223	-0,024	-0,271	-0,030
PEH	2,487	-0,415	0,059	0,504	0,076
PT	2,263	0,618	-0,327	-0,752	-0,420
ALC	2,700	0,516	0,369	-0,627	0,473
SST	2,001	0,380	-0,128	-0,463	-0,165

Las especies mejor representadas de acuerdo con el estadístico de bondad de ajuste (función *goodness()*, paquete **vegan** de R) fueron: *Achnantheidium sp.*, *Aulacoseira sp.*, *Calothrix sp.*, *Closterium sp.*, *Cocconeis sp.*, *Cyclotella sp.*, *Cymato sp.*, *Diatoma sp.*, *Didymosphenia sp.*, *Diploneis sp.*, *Epithemia sp.*, *Fragilaria sp.*, *Frustulia sp.*, *Gomphoneis sp.*, *Gomphonema sp.*, *Gyrosigma sp.*, *Lygnbya sp.*, *Mougeotia sp.*, *Pediastrum sp.*, *Rhopalodia sp.*, *Scenedesmus sp.*, *Suirella sp.*, *Tabellaria sp.*, *Ulothrix sp.*

La representación de los sitios de acuerdo con la bondad de su ajuste en el ACC fue, en orden decreciente: 05RC, 11RL, 41RN, 36RN, 27RL, 13RL, 01LHH, 10RL, 26RL, 37RN, 12RL, 16RL, 07RL, 06RL, 04RC, 08RL, 42RN, 25RL, 40RN, 19RL, 14RL, 18RL, 29RN, 34RN, 30RN, 09RN, 31RN, 28RL, 22RL, 35RN, 39RN, 02LHH, 33RN, 20RL, 03LNH, 44RN, 32RN, 21RL, 43RN, 24RL, 17RL, 15RN, 38RN, 23RL.

Para visualizar los resultados del ACC, se extrajeron los dos primeros ejes canónicos y se graficaron dos biplots (escalado 1 y 2) y el triplot con escalado 2. Detalle de abreviaturas de especies en Tabla VII del Anexo.

En la Figura 4.11 se presentan los biplots con escalado 1 que relaciona sitios y variables ambientales (a) y con escalado 2 que relaciona las especies de algas con las variables ambientales (salida CANOCO). El primer eje canónico del ACC DIDYMO estuvo relacionado principalmente a la magnitud de CE, OD y PEH, y por la importancia de las cargas sobre ambos ejes, se generaron gradientes oblicuos de PT, con los mayores valores en el cuadrante inferior derecho y de ALC y TAG con mayores valores en el cuadrante superior derecho del plano principal de ordenación del ACC.

Con respecto a las especies, el método para evaluación gráfica consiste en proyectar (en ángulo recto) la especie sobre la flecha que representa cada variable. Esto da una aproximación de los promedios ponderados de las especies con respecto a las variables ambientales.

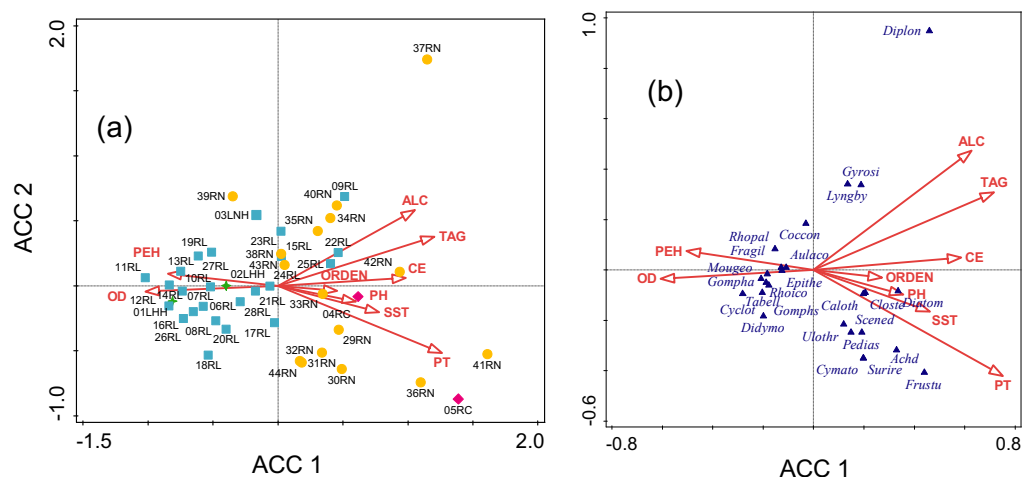


Figura 4.11. ACC Biplots, primeros dos ejes canónicos. Salida CANOCO. (a) Biplot de sitios y variables ambientales, escalado 1. Los colores de símbolos identifican las cuencas (púrpura: río Colorado, verde: río Hua Hum, amarillo: río Neuquén, azul: río Limay). (b) Biplot de especies y variables ambientales, escalado 2.

La Figura 4.12 representa el triplot con escalado 2, salida CANOCO. Este escalado se enfoca en la correlación entre especies y para una mejor visualización se incluyeron solo las familias más representativas en la ordenación, de acuerdo con su distancia al centro del gráfico. Tal como fuera determinado en el AFM, se definieron también tres grupos con características similares, distinguiéndose en el grupo 1 los sitios de la cuenca del río Limay, y Lacar – Hua Hum, con los mayores valores de OD y PEH, y bajos valores de CE, siendo algunas de las especies asociadas la carofita *Mougeotia sp.* y las diatomeas *Cyclotella sp.*, *Didymosphenia sp.*, *Gomphonema sp.*, *Gomphoneis sp.*, *Rhopalodia sp.* y *Tabellaria sp.* A los grupos 2 y 3 se asociaron los sitios de la cuenca del río Neuquén y del río Colorado, con los menores valores de OD y PEH, y altos valores de CE. El grupo 2 se caracterizó por presentar los mayores valores de ALC y TAG, y valores medios de PT, y a este grupo se asociaron especies tales como las diatomeas *Diploneis sp.* y *Gyrosigma sp.* y la cianobacteria *Lygnbya sp.* Mientras que el grupo 3 se caracterizó por altos valores de PT, SST y valores medios de ALC y TAG, con especies asociadas tales como las clorofitas *Closterium sp.*, *Pediastrum sp.*, *Scenedesmus sp.*, *Ulothrix sp.*, las diatomeas *Achnantheidium sp.*, *Cymatopleura sp.*, *Diatoma sp.*, *Frustulia sp.*, *Surirella sp.*, y la cianobacteria *Calothrix sp.*

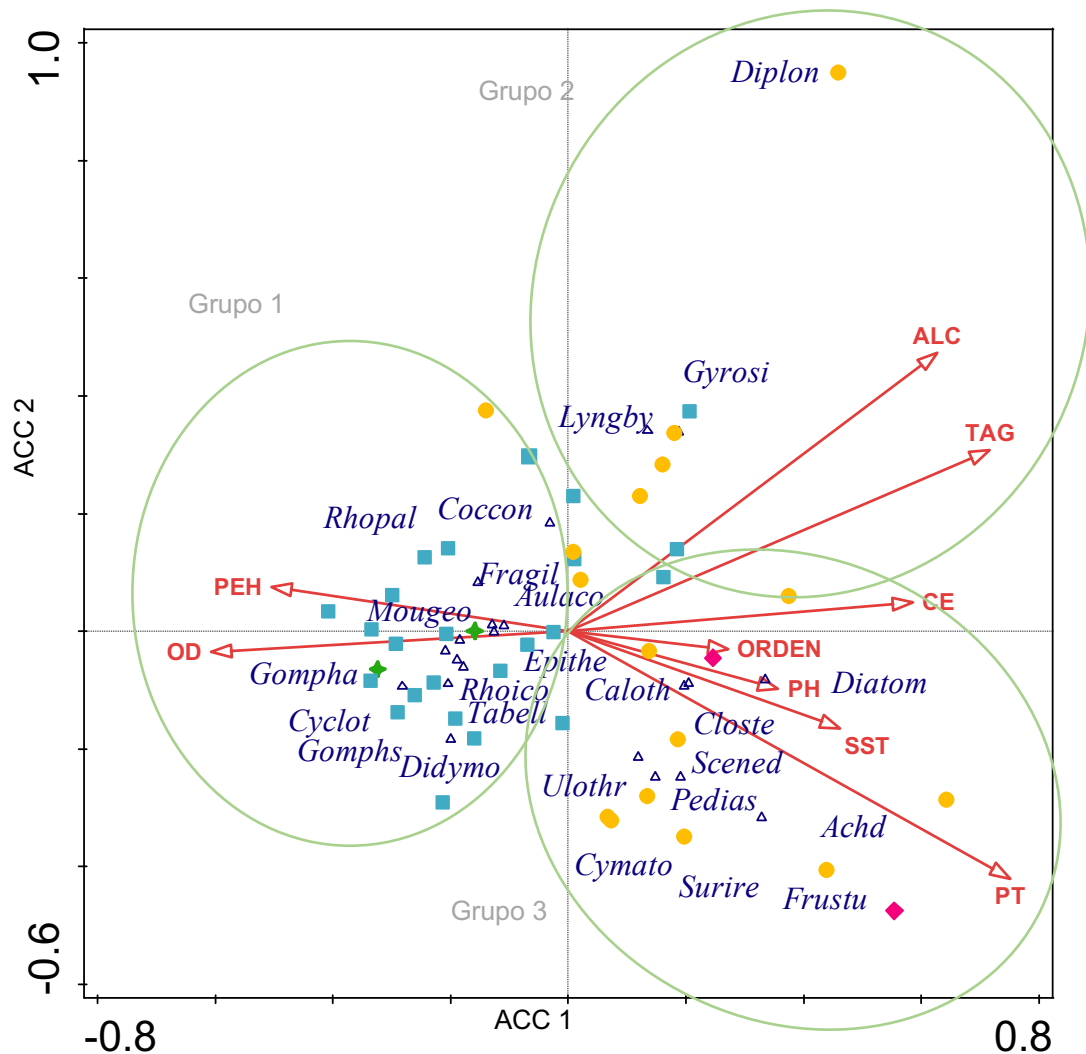


Figura 4.12. Triplot del ACC DIDYMO, primeros dos ejes canónicos. Salida de CANOCO. Escalado 2. Los vectores rojos representan las variables predictoras ambientales, los nombres en azul representan las especies de algas. Los colores de los símbolos de sitios identifican las cuencas (púrpura: río Colorado, verde: río Hua Hum, amarillo: río Neuquén, azul: río Limay).

Por último, se realizó un proceso de selección del mejor modelo reducido, el cual incorporó las variables PT y ALC.

4.2.2. Discusión de aspectos biológicos

Las configuraciones halladas tanto con el AFM, como con el ACC, fueron análogas en el sentido de poder establecer grupos similares de sitios asociadas con diferentes especies algales y determinadas características físico-químicas del agua.

Las proyecciones ortogonales de los puntos sobre los vectores definidos por las variables indican el valor óptimo del individuo en la variable. Por ejemplo, si se coloca el foco en la especie *D. geminata*, cuya presencia originó el monitoreo integral, se aprecia que en el estudio de la campaña del año 2014 se asoció principalmente a sitios de bajo ORDEN, con los valores más bajos de CE, PT, ALC y TAG y valores altos de OD y PEH. Este resultado y una evaluación ampliada en el tiempo puede encontrarse en el trabajo de Beamud *et al.* (2019), un estudio realizado sobre los datos de todas las campañas del mismo programa asociado a los datos de esta tesis.

Para el estado de invasión en el año 2014, *D. geminata* puede ser considerada una especie rara por su mínima abundancia en relación con las demás especies, siendo la cuarta especie de menor abundancia entre las 44 registradas en todos los sitios evaluados en este trabajo (Tabla VII del Anexo). Tanto en el AFM como en el ACC esta especie se encontró asociada al grupo 1, sitios de las cuencas del río Limay, y Lacar – Hua Hum. Las actividades recreativas se consideran la principal fuente de dispersión de *D. geminata* en las cuencas hidrográficas (Kilroy y Unwin, 2011). Sin embargo, si bien la llegada de células es un requisito previo, un establecimiento exitoso requiere que el río tenga condiciones adecuadas para la supervivencia y floración celular (Montecino *et al.*, 2014).

Si bien el porcentaje de inercia explicado no fue muy grande, alcanzando un 27,8% con el AFM y 14,61% con el RDA entre los dos primeros ejes, en primera instancia permitió observar las características diferenciales de los sitios de las diferentes cuencas, siendo más relevante entre sitios de la cuenca del Río Limay y Neuquén, esta última coincidente con altos niveles de fósforo, alcalinidad y conductividad. Si bien ambas son cuencas reguladas y presentan diferencias menores en los sectores por encima y por debajo de las represas, existen

diferencias entre las mismas en cuanto a los parámetros mencionados según el reporte de la AIC (2014), coincidiendo entonces con las agrupaciones encontrada en este trabajo. En dicho informe para el río Limay se reportan valores promedios de 54,4 $\mu\text{S}/\text{cm}$ de CE, 1,7 mg/L de SST y 24,2 mg $\text{CO}_3\text{Ca}/\text{L}$ de ALC, mientras que para el río Neuquén reportan promedios mayores para los mismos parámetros (253 $\mu\text{S}/\text{cm}$ de CE, 34,1 mg/L de SST y 56,1 mg $\text{CO}_3\text{Ca}/\text{L}$ de ALC), un contraste similar entre los ríos a los registrados en este estudio, con promedios de 63,6 $\mu\text{S}/\text{cm}$ de CE, 13,5 mg/L de SST y 25,5 mg $\text{CO}_3\text{Ca}/\text{L}$ de ALC para el río Limay, y promedios mayores de para el río Neuquén (191,5 $\mu\text{S}/\text{cm}$ de CE, 52,5 mg/L de SST y 56,1 mg $\text{CO}_3\text{Ca}/\text{L}$ de ALC).

Es valioso definir la relevancia de las variables ambientales que se incorporan en los estudios, dado que en el contexto de una investigación el costo del análisis de las muestras es un punto clave a la hora de planificar. Si bien prácticamente todas las variables resultaron importantes, resulta conveniente la selección de dos variables significativas que definen un modelo reducido con solo dos variables en el ACC, que son PT y ALC, ambas variables determinadas en laboratorio.

Es importante enfocarse en la parte presupuestaria de los estudios biológicos, dado que los monitoreos deben ser rentables, pero científicamente válidos. Es trascendental recopilar, compilar, analizar e interpretar rápidamente datos ambientales para facilitar las decisiones de gestión y las acciones resultantes para el control o la mitigación del deterioro. Para un análisis exploratorio, ciertamente se deben incluir variables que se consideran relacionadas con los determinantes más importantes de la composición de especies. Sin embargo, a menudo también es deseable incluir otras variables que sean fáciles y económicas de medir. Para esto se necesita:

- Procesamientos de datos con rápidos resultados que sean aplicados en las decisiones de gestión.
- Informes científicos fácilmente traducidos a los gestores y al público en general.
- Procedimientos ambientalmente benignos.

4.3. Discusión general de aspectos estadísticos

¿Cuál es la finalidad de los métodos descritos en esta tesis?

- Reducir la complejidad de los datos, mediante su representación en un diagrama de ordenación de pocas dimensiones.
- Conservar la máxima cantidad de información posible, mediante la generación de factores latentes que se ordenan según la varianza o inercia explicada.
- Describir gráficamente los patrones de composición de taxones y las relaciones entre sitios y variables.
- Facilitar la interpretación de la distribución de los sitios y taxones en relación con gradientes determinados por variables ambientales conocidas (directos) o desconocidas (indirectos).

Los datos analizados permitieron realizar un estudio exploratorio integrando información sobre los factores ambientales que definen los sitios muestreados y la ocurrencia y abundancia de las comunidades biológicas. En los estudios que involucran ensambles de taxones, generalmente es más interesante describir la variabilidad de la estructura del ensamble como un todo, es decir, la variación cuantitativa observada a través del espacio o el tiempo (Hurlbert, 1984) que mirar cada especie independientemente.

4.3.1 AFM vs RDA

En el AFM la similitud entre las representaciones geométricas derivadas de cada grupo de variables se mide por el coeficiente RV. Este coeficiente está normalizado y tiene rango entre 0 y 1, tiende a ser mayor cuando el número de individuos es chico o el número de variables es grande. Además, se puede evaluar su significancia mediante permutaciones (Josse *et al.*, 2008).

El coeficiente Lg utilizado en el análisis AFM no está normalizado, por lo que es más difícil de interpretar.

El AFM propone un punto de vista simétrico, exploratorio, donde las estructuras correlativas están expuestas sin ninguna referencia a una direccionalidad de

posibles relaciones causales. Tampoco se prueba una hipótesis direccional formal. Por lo tanto, este enfoque no está adaptado al modelado de relaciones asimétricas, una tarea dedicada a RDA o CCA. Sin embargo, el AFM podría usarse en las primeras etapas de un proyecto de investigación como una técnica neutral de exploración de datos para ayudar a generar hipótesis causales, que luego podrían probarse en un conjunto de datos independiente (Borcard *et al.*, 2011).

El AFM se ha utilizado principalmente en evaluación sensorial y química hasta ahora, pero el potencial para aplicaciones ecológicas es importante como en Beamud *et al.* (2010). Este método es muy útil para explorar las complejas relaciones entre varios grupos de descriptores ecológicamente significativos, independientemente de su número y tipo (Borcard *et al.*, 2011), como es el caso de este trabajo.

En cuanto a la utilización de los métodos de ordenación canónica, son ampliamente utilizados en evaluaciones similares a la que se planteó en esta tesis. Por ejemplo, Miserendino *et al.* (2008), utilizaron el análisis de correspondencia canónica (ACC) para evaluar las relaciones entre ensamblajes de macroinvertebrados y variables ambientales. Incluyeron todas las variables ambientales inicialmente para evaluar la respuesta de especies y sitios a gradientes ambientales. Transformaron algunas variables a $(\log x + 1)$, las variables que estaban fuertemente correlacionadas con otras (aquellas con un factor de inflación > 10) en el análisis inicial fueron eliminadas y un análisis adicional fue llevado a cabo con las 12 variables ambientales restantes. Utilizaron la opción de selección forward proporcionada por CANOCO y mantuvieron para el análisis las variables con $p < 0,01$ (prueba de permutación de Monte Carlo).

De la misma forma, Macchi (2017) también utiliza el Análisis de Correspondencias Canónico (ACC) y el Análisis de Redundancia (RDA), para examinar la relación entre las asociaciones de invertebrados acuáticos o métricas seleccionadas y las variables ambientales, para evaluar la degradación ambiental en mallines patagónicos. Importante trabajo que destaca la vulnerabilidad de los mallines frente a los disturbios antrópicos debidos a los diferentes usos del suelo y el rol de las comunidades de macroinvertebrados como potenciales bioindicadores.

Según Palmer (1993), dado que la significación estadística de un RDA o el ACC se determina mediante una prueba de aleatorización, no es necesario transformar los datos para cumplir con los supuestos estadísticos. Sin embargo, las transformaciones se pueden usar para atenuar la influencia de los valores atípicos. La elección de la transformación afecta las coordenadas de los sitios, los taxones y las variables ambientales. Una transformación de amortiguación (por ejemplo, raíz cuadrada) tiende a hacer que las muestras y las especies se distribuyan de manera más uniforme. Solo en raras ocasiones la transformación de las variables ambientales cambiará la interpretación general de una ordenación.

En esta tesis, cuando el RDA fue realizado con CANOCO, por defecto el programa no sugirió transformación de la matriz Y . Sin embargo, se probaron los análisis (AFM y RDA) tanto con la matriz original, como con la transformada mediante la transformación "Hellinger". Los resultados numéricos fueron diferentes, pero no contradictorios, optándose por el uso de la matriz transformada para todos los análisis. Además, el porcentaje de ceros de la matriz Y fue un valor alto (65%), resultado que no es raro según la bibliografía para los datos de abundancia de especies, pero para evitar el uso de doble ceros como indicadores de semejanza entre sitios, se aplicó la transformación a los datos de frecuencias (Legendre y Legendre, 2012).

Vegan informa autovalores sin escala, CANOCO informa directamente valores propios escalados en la forma en que su suma es igual a uno, no una variación total. Los autovalores canónicos miden la cantidad de varianza explicada por el modelo RDA.

Una diferencia importante en el uso del software es que CANOCO está programado para evaluar los datos y ofrecer la mejor alternativa a su criterio, muchas veces incluyendo transformaciones de la matriz Y y de la matriz X (Leps y Smilauer, 2003), mientras que con R cada paso del análisis debe ser elegido por el investigador, a su criterio.

El análisis de gráficos es fundamental para la interpretación de resultados, es por esto que tanto en R como en CANOCO, pueden ser acomodados para una mejor visualización mediante sentencias y a través del menú del gráfico. En R se pueden

crear gráficos con la función *plot()* de escalado 1 y 2 (son los que salen por defecto). En el escalado 1, la distancia entre los objetos se aproxima a sus distancias Euclídeas, mientras que en el escalado tipo 2, los ángulos entre las variables *x* e *y* reflejan su correlación. La definición de que gráfico incluir y que ajustes hacerle es fundamental, y así como se menciona en Legendre y Legendre (2012) en este trabajo se optó por biplots y triplot, dado que la cantidad de objetos a incluir en los diagramas fue grande.

Los resultados obtenidos con el RDA pueden considerarse un primer paso, dado que existe la posibilidad de estimar un modelo reducido, que también resulte significativo. Las funciones *ordiR2step()* y *ordistep()* del paquete **vegan** y la función *forward.sel()* del paquete **adespatial**, se pueden utilizar para realizar un proceso de selección de variables. A través del *ordiR2step* se estimó un modelo reducido que incorporó tres variables significativas: PRS, PEH, ALT. Con estas 3 variables podría generarse una nueva matriz, realizar un nuevo RDA y evaluar la nueva ordenación. Por su lado, CANOCO ofreció estos mismos resultados junto con todas las salidas mencionadas anteriormente.

En resumen, los mismos resultados fueron obtenidos con R y con CANOCO, la diferencia es que en R la ejecución de cada parte del análisis debe ser programada cuando no salió por defecto, mientras que CANOCO directamente proporcionó todos los resultados en una misma salida.

4.3.2 AFM vs ACC

El AFM da la posibilidad de introducir matrices de variables de distinto tipo (continuas, cualitativas y de tipo frecuencia) como activas en un mismo análisis (Escofier y Pagès, 1984; Moncada Herrera, 2007; Abdi *et al.*, 2013). Así, por ejemplo, se podrían involucrar características continuas o cualitativas (por ejemplo, índices o indicadores ambientales). Se podría obtener con el AFM, un gráfico en el cual cada grupo esté representado por un único punto. Este gráfico permite una comparación sintética de los grupos y una evaluación rápida de su mutua semejanza, lo que es particularmente útil cuando el número de grupos es elevado. Permitiría, por ejemplo, extender el estudio presentado aquí y analizar no sólo un año sino también la evolución a lo largo de varios años.

Según Abdessemed y Escofier (1996), el AFM debe ser un requisito previo para un ACC (o un RDA) que pueda realizarse para un análisis más detallado, especialmente si se han detectado estructuras comunes. AFM y ACC son dos métodos complementarios, si con el AFM, tenemos la posibilidad de detectar estructuras comunes o específicas, el ACC (o RDA) refinará útilmente los resultados encontrados, esencialmente en el caso donde hay estructuras comunes en al menos dos grupos.

Abdessemed y Escofier (1996) mencionan que la aplicación de AFM y ACC, nos permite encontrar un primer factor que es el mismo en los dos análisis. Esto se explica ya que el primer factor del AFM es un factor absolutamente común para ambos grupos. Sin embargo, el segundo factor tanto del AFM como del ACC son muy diferentes. Esto se debe al hecho de que el segundo factor del AFM es un factor de distribución de la especie que está muy poco relacionado con las variables descriptivas y que no puede aparecer en el ACC.

Los triplots, que recogen la información de los tres elementos que componen las matrices X e Y (sitios, variables ambientales y frecuencias de algas) son sumamente útiles a la hora de evaluar sus relaciones. Pero muchas veces se necesita observar los elementos de a dos para poder describir mejor dichas relaciones, sobre todo como en este caso en el cual el número de elementos es grande (Legendre y Legendre, 2012). En este sentido se utilizaron los biplots con escalado 1 y 2 para representar: sitios y variables ambientales, y especies y variables ambientales, respectivamente (Borcard *et al.*, 2011). Esto mismo aplica para las dos bases de datos analizadas en este trabajo.

Aquí podríamos intentar predecir la abundancia de una especie en particular en una muestra en función de los valores de uno o más predictores (variables ambientales y/o ejes de ordenación en el contexto de los métodos de ordenación) Leps y Smilauer (1999).

La manera exacta en la que trabaja el ACC es compleja y está explicada con todo detalle en ter Braak (1986; 1987). El autor afirma que esta técnica es una extensión del Análisis de Correspondencias (alias "medias recíprocas" (RA) HILL (1973)) en la que los ejes de ordenación se eligen en función de las variables

ambientales conocidas, imponiendo la restricción adicional de que los ejes deben ser combinaciones lineales de las variables ambientales (Fernández Gómez *et al.*, 1996).

El ACC proporciona las mismas posibilidades de elección de coordenadas para filas y columnas que el AC habitual. El triplot es un diagrama en el cual, además, añadimos vectores correspondientes a las variables explicativas. Para la visualización de las variables explicativas tenemos dos posibilidades. Una es utilizar sus coeficientes de correlación con los ejes para definir sus posiciones; los representaríamos como si fueran puntos adicionales. La otra posibilidad es usar los coeficientes de regresión estandarizados derivados de su relación con los ejes. Nosotros consideramos que esta última posibilidad es mejor ya que refleja la idea de que en el ACC los ejes están relacionados linealmente con las variables explicativas (Greenacre, 2008).

En relación con el tipo de coordenadas, Palmer (1993) sugirió el uso de restricciones lineales ("coordenadas lc") en los diagramas de ordenación, porque estos dan mejores resultados en las simulaciones, mientras que las coordenadas del sitio ("coordenadas wa") son un paso del análisis restringido al no restringido. Sin embargo, McCune (1997) demostró que las variables ambientales ruidosas (y todas las mediciones ambientales son ruidosas) destruyen las "coordenadas lc", mientras que las "coordenadas wa" son poco afectadas (Oksanen *et al.*, 2019).

El Análisis de Correspondencia Canónica (ACC) es un método cada vez más popular para el análisis multivariado de datos de la comunidad ecológica. Sin embargo, según McCune (1997) es uno de los métodos multivariados más potencialmente engañosos para el análisis comunitario. Afirma que la inclusión de variables ambientales ruidosas o irrelevantes puede distorsionar la representación de gradientes en la estructura de la comunidad. No obstante, menciona que estos problemas se pueden evitar mediante el uso de métodos tradicionales de ordenación indirecta, donde se expresa la estructura comunitaria pura, sin ninguna restricción impuesta por las variables ambientales.

Debe ser considerado que el ACC es sensible a especies raras que ocurren en muestras pobres de especies, y la reducción de peso de tales especies ayudan a reducir este problema (Ramette, 2007; Legendre y Legendre, 2012).

Una de las cuestiones que debe resolverse es, ¿dentro de los métodos canónicos, cual utilizar? Para el análisis de sitios que representan gradientes cortos, ACP y RDA pueden ser adecuados. Para gradientes más largos, muchas especies se reemplazan por otras a lo largo del gradiente y esto genera muchos ceros en la matriz de datos de especies. Los ecólogos comunitarios han argumentado repetidamente que la distancia Euclídea (y, por lo tanto, ACP y RDA) es inapropiada para los datos de abundancia de especies crudas que involucran abundancias nulas (Legendre y Legendre, 2012). En términos ecológicos esto implica que la distancia Euclídea sobre enfatiza la dominancia de los valores de las especies cuya abundancia es alta y puede dar lugar a una alta afinidad artificial entre entidades que no tienen en común muchos atributos (Boesch, 1977; Herrera Moreno, 2000). Por esa razón, ACC es a menudo el método preferido por los investigadores que analizan datos de composición, a pesar del problema que plantean las especies raras. (Legendre y Borcard, 2001)

Al decidir si utilizar el tipo de método de ordenación lineal o unimodal, debe tenerse en cuenta otra diferencia importante entre ellos. Los métodos unimodales siempre funcionan implícitamente con datos estandarizados. Los métodos ACC, AC o DCA resumen la variación en las frecuencias relativas de las variables respuesta (especies). Una implicación importante de este hecho es que estos métodos no pueden funcionar con muestras 'vacías', es decir, registros en los que no hay especies presentes. Además, los métodos unimodales no se pueden usar cuando las variables respuesta no tienen las mismas unidades (Leps y Smilauer, 1999).

Ter Braak (1986) mostró que el AC es una aproximación de la ordenación gaussiana y el ACC es una aproximación de la regresión gaussiana restringida. Esta es la razón ecológica del AC y ACC.

Zuur *et al.* (2007) proponen algunos lineamientos para cuando usar cada método:

1. RDA debe usarse para analizar relaciones lineales entre especies y variables ambientales.
2. ACC puede usarse para analizar relaciones unimodales entre especies y variables ambientales.
3. ACP o RDA deben usarse si la diversidad beta es pequeña, o si el rango de las muestras cubre solo una pequeña parte del gradiente.
4. Un gradiente largo tiene una alta diversidad beta, y esto indica que se debe usar AC o ACC.

CAPÍTULO 5. Conclusiones

Se corrobora la hipótesis de que las técnicas disponibles para relacionar una matriz de frecuencias y un grupo de variables continuas que describen un mismo conjunto de individuos (AFM, RDA y ACC) no dan exactamente los mismos resultados.

El estudio integral de ambas matrices permitió asegurar que el AFM es adecuado para realizar un primer análisis exploratorio del conjunto de datos, mientras que los métodos asimétricos de análisis directo del gradiente (RDA y ACC), resultan útiles para explicar estructuras encontradas por el AFM y definir modelos que puedan explicar las relaciones entre el ambiente y la biota. Se encuentra entonces que estos métodos pueden ser complementarios.

Se debe tener en cuenta que para comparar el método simétrico (AFM) con los asimétricos (RDA y ACC), se utilizaron exactamente las mismas matrices, con lo cual todo el potencial de los datos no está volcado, ya que para que esto sea posible se quitaron algunas variables del análisis. Para una evaluación más eficaz se sugiere utilizar el AFM con todos los datos, ya que no presenta restricciones y luego avanzar en un siguiente análisis a través del RDA o ACC según corresponda.

Como ventaja, el AFM resulta sumamente útil y versátil dado que para la evaluación de un mismo conjunto de individuos el método no se restringe a solo dos matrices o tablas de datos, y puede incorporar variables de diferente naturaleza. El método tampoco tiene restricciones respecto al número de individuos y variables, siendo que el RDA o el ACC necesariamente deben contar con más individuos que variables explicativas. El AFM se recomienda para un primer análisis exploratorio.

Por las características de la ordenación, el AFM estudiará la similaridad entre individuos respecto a uno o más grupos de variables y las relaciones entre las variables, mientras que los métodos como el RDA y el ACC pondrán de relevancia las relaciones funcionales y los taxones que realmente aparecen con condiciones diferenciales del ambiente, aquellas apariciones que son explicadas por la

heterogeneidad ambiental. Así, para que los métodos RDA o ACC puedan ser utilizados, debe haber una relación funcional entre las matrices, y el primer paso de cada método, que es una regresión lineal múltiple, debe ser significativa, de lo contrario no es correcto continuar con el análisis.

Es importante entender el fundamento teórico de los métodos de análisis directo del gradiente para utilizarlos de manera correcta. Si bien el AFM no puede realizarse en CANOCO, comparando la implementación de los métodos RDA y ACC en CANOCO y R, se obtuvieron similares resultados, con algunas mínimas diferencias en cómo se presentan las salidas.

El programa CANOCO es una excelente herramienta para los biólogos, ecólogos y otros, que necesitan analizar datos complejos, dado que intrínsecamente el software incorpora conocimiento ecológico y sugiere los análisis apropiados según el conjunto de datos. Posee una batería extensa de métodos de análisis, y muy buenas opciones gráficas.

Utilizar el lenguaje R es el mayor desafío que afrontan quienes no tienen grandes conocimientos en estadística ni en programación, pero sin dudas es una de las herramientas más potentes en el mundo de la estadística hoy en día, por lo que su utilización debe incentivarse en todos los ámbitos de la ciencia y la investigación.

Desde lo biológico y la gestión del patrimonio acuático, la Provincia del Neuquén ha avanzado de manera significativa en los monitoreos relacionados al ambiente acuático y la evaluación integral del ecosistema a nivel de cuencas, incorporando las condiciones hidrológicas, la biota, calidad del agua y el hábitat. Es importante entonces, traducir la información obtenida en políticas o recomendaciones para la gestión, planificación y toma de decisiones, a la par de procurar el uso eficiente de recursos económicos y humanos. Los métodos estadísticos utilizados en este trabajo resultaron sumamente útiles para analizar datos ecológicos procedentes de biomonitoreos acuáticos y las posibles relaciones entre las frecuencias de taxones y las variables ambientales.

Esta Tesis aporta sólo una mínima parte del universo de interacción entre la estadística y la realidad del ambiente, sobre todo en un medio tan difícil de estudiar como el acuático, pero se espera que sea de utilidad para encarar las evaluaciones de datos similares de la mejor manera.

BIBLIOGRAFÍA

Abdessemed, L. y Escofier, B. (1996) Analyse factorielle multiple de tableaux de fréquences; Comparaison avec l'analyse canonique des correspondances. Journal de la société statistique de Paris, tome 137, no 2 (1996): 3-18.

Abelli B., M., Baffico, G., Beamud, G., Bonetto, M., Castiñeira, L., Diaz, M., Laffitte, L., Laurenzano, B., Morzenti, P., Muñiz Saavedra, J., Pedrozo, F., Rivera, D., Temporetti, P. y Ugolini, F. (2012) Programa de vigilancia, seguimiento y control de *Didymosphenia geminata* (Lyngbye) Schmidt 1899 en la Provincia del Neuquén". CAL 5, Santa Fe, Argentina.

Abdi, H. y Williams, L.J. (2010) Principal Component Analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2, 433-459. <http://dx.doi.org/10.1002/wics.101>

Abdi, H., Williams, L.J., y Valentin, D. (2013) Multiple factor analysis: principal component analysis for multitable and multiblock data sets. WIREs Comput Stat. doi: 10.1002/wics.1246.

AIC. (2007) Calidad del agua del río Negro – Años 2001 a 2006. SGA, DPA, SSMA, DPRH.

AIC. (2012) Protocolo para la detección y seguimiento de *Didymosphenia geminata* (Lyngbye) Schmidt en el área andina patagónica, Argentina. AIC, SSAyDS, DGBA, DPRH, DPA.

AIC. (2014) Red básica de monitoreo de la calidad del agua en ríos. Síntesis de Resultados Serie 2001-2013. AIC, SSAyDS, DGBA, DPRH, DPA.

Anderson, M.J., T.O. Crist, J.M. Chase, M. VellendInouye, B.D., Freestone, A.L., Sanders, N.J., Cornell, H.V., Comita, L.S., Davies, K.F., Harrison, S.P., Kraft, N.J.B., Stegen, J.C. y Swenson, N.G. (2011) Navigating the multiple meanings of Beta diversity: a roadmap for the practicing ecologist. Ecol. Lett. 14: 19-28. [258, 260]

Aumassanne, C. (2019) Impacto de los cambios en el uso y la cobertura de la tierra y de la variación del clima, sobre los principales componentes del balance de agua en la cuenca del río Colorado, Argentina. Tesis doctoral. UNLP, Argentina. <https://doi.org/10.35537/10915/82130>

Baccalá, N. (2004) Contribuciones al análisis de matrices de datos multivía: tipología de las variables. *Doctoral Thesis*, Universidad de Salamanca, Spain.

Barbour, M.T., Stribling, J.B., Gerritsen, J. y Karr, J.R. (1996) Biological Criteria: Technical Guidance for Streams and Small Rivers. EPA 822-B-96-001. U.S. Environmental Protection Agency, Washington, DC.

Barbour, M.T., Gerritsen, J., Snyder, B.D., y Stribling, J.B. (1999) Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, DC.

Bartlett, R. (1939) Measurement in Psychology. *Nature* 144: 498–502.
<https://doi.org/10.1038/144498a0>

Beavers, A.S, Lounsbury, J.W., Richards, J.K., Huck, S.W., Skolits G.J., y Esquivel, S.L. (2013) Practical Considerations for Using Exploratory Factor Analysis in Educational Research. *Practical Assessment, Research & Evaluation*, 18(6). <http://pareonline.net/getvn.asp?v=18&n=6>

Beamud, S.G., Laffitte, L., Muñiz Saavedra, J., Baffico, G., Laurenzano, B., Díaz, M., y Pedrozo, F. (2019) Controlling factors affecting the presence and distribution of *Didymosphenia geminata* in lakes and rivers of the Neuquén province, Argentina. *Ecohydrology*, e2148. <https://doi.org/10.1002/eco.2148>

Beamud, S.G., Diaz, M., Baccalá, N.B., y Pedrozo, F.L. (2010) Analysis of patterns of vertical and temporal distribution of phytoplankton using multifactorial analysis: Acidic Lake Caviahue, Patagonia, Argentina. *Limnologica* 40: 140–147.

Benzécri, J.P. (1969) Statistical analysis as a tool to make patterns emerge from data. 35-60 in: S. Watanabe [ed.] *Methodologies of pattern recognition*. Academic Press, New York.

Boesch, D.F. (1977). Application of numerical classification in ecological investigations of water pollution. *Ecological Res. Ser.*, EPA-600/3-77-033,115pp.

Borcard, D., Gillet, F., y Legendre, P. (2011) *Numerical Ecology with R*. Springer Science & Business Media, Springer-Verlag New York, USA.

Bran, D., Ayeza, J., Lopez, C., (2002) Áreas ecológicas de Neuquén. Informe Técnico Instituto Nacional de Tecnología Agropecuaria (INTA). Laboratorio de Teledetección -SIG. Bariloche.

Carmona Jiménez, J., Carmona Jiménez, R., y Ramírez, R. (2016) Estudio del valor indicador de las comunidades de algas bentónicas: Una propuesta de evaluación y aplicación en el Río Magdalena, Ciudad de México. *Rev. Int. Contam. Ambie.* 32 (2): 139-152.

Diaz, M.M., Pedrozo, F.L., Reynolds, C., y Temporetti, P. (2007) Chemical composition and the nitrogen-regulated trophic state of Patagonian lakes. *Limnologica*, 37, 17–27. <https://doi.org/10.1016/j.limno.2006.08.006>

Domínguez, E., Giorgi, A. y Gómez, N. (2020) La bioindicación en el monitoreo y evaluación de los sistemas fluviales de la Argentina: Bases para el análisis de la integridad ecológica. Editorial Eudeba.

Chessel, D., Lebreton, J. y Yoccoz, N. (1987) Propriétés de l'analyse canonique des correspondances; une illustration en hydrobiologie, *Revue Statistique Appliquée* 35(4): 55–72.

Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* 29, 751–760. Escoufier, Y. doi:10.2307/2529140

Escoufier, B., y Pagès, J. (1984) L'analyse factorielle multiple: une méthode de comparaison de groupes de variables, in E. Diday, ed., 'Data Analysis and Informatics, III', Elsevier Science, Amsterdam, pp. 41–56. Proceedings of the Third International Symposium on Data Analysis and Informatics.

Escoufier, B., y Pagès, J. (1990) *Analyses factorielles simples et multiples*, Paris, Dunod.

Escoufier, B., y Pagès, J. (1992) *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*. Ed. Universidad del País Vasco, Bilbao.

Fernández Gómez, M.J., Galindo, M.P., Barrera, I., Vicente-Villardón, J.L., y Martín, A. (1996) Alternativa al análisis canónico de correspondencias basada en los métodos Biplot. *Mediterránea. Serie de estudios biológicos*, 63-71.

Fernández, H., Domínguez, E., Romero, F. y Cuezco, G. (2006) La calidad del agua y la bioindicación en los ríos de montaña del Noroeste Argentino. *Serie Conservación de la Naturaleza N° 16*. Fundación Miguel Lillo, Tucumán, Argentina.

Fisher, R.A. (1935) *The Design of Experiments*. Edimburgo. Oliver and Boyd.

Fisher, R.A. (1936) the use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7: 179-188. doi:10.1111/j.1469-1809.1936.tb02137.x

Flotemersch, J.E., Muñiz Saavedra, J.M., Laffitte, L., Laurenzano, B., Abelli Bonardi, M., Blocksom, K.A. (2017) Benthic macroinvertebrate field sampling effort required to produce a sample adequate for the assessment of rivers and streams of Neuquén Province, Argentina. *Limnologica* 65 (2017) 55–60.

Flotemersch, J.E., Stribling, J.B. y Paul, M.J. (2006) Concepts and Approaches for the Bioassessment of Non-wadeable Streams and Rivers. EPA 600-R-06-127. US Environmental Protection Agency, Cincinnati, Ohio.

Friberg, N., Bonada, N., Bradley, D.C., Dunbar, M.J., Edwards, F.K., Grey, J., Hayes, R.B., Hildrew, A.G., Lamouroux, N., Trimmer, M., y Woodward, G. (2011) Biomonitoring of human impacts in natural ecosystems: The good, the bad and the ugly. *Adv. Ecol. Res.* 44, 1-68. doi: 10.1016/B978-0-12-374794-5.00001-8

Garson, G.D. (2013) Factor Analysis. Statistical Associates Blue Book Series, BOOK 15. Statistical Associates Publishing. Asheboro, NC.

Gauch, H.G. Jr. (1982) Multivariate analysis in community ecology. Cambridge University Press, Cambridge. 298 pp.

Gerhardt, A. (1999) Biomonitoring of Polluted Water. Reviews on Actual Topics. Trans Tech Publ, Zürich, Switzerland. 301pp.

Gibson, G.R., Barbour, M.T., Stribling, J.B, y Gerritsen, J. (1996) Biological Criteria: Technical Guidance for Streams and Small Rivers. U.S. Environmental Protection Agency; Office of Water; Washington, DC.

Goodall, D.W. (1954) Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Aust. J. Bot.* 2: 304-324.

Greenacre, M.J. (1983) Theory and applications of correspondence analysis. Academic Press, London. 364pp.

Greenacre, M. (2008) La práctica del análisis de correspondencias. Fundación BBVA. 375pp.

Guo, C., Park, Y.S., Liu, Y., y Lek, S. (2015) Toward a new generation of ecological modelling techniques: Review and bibliometrics. *Developments in Environmental Modelling* 27:11-44. DOI: 10.1016/B978-0-444-63536-5.00002-8. In book: *Advanced Modelling Techniques Studying Global Changes in Environmental Sciences* Chapter: 2. Elsevier.

Gower, J.C. (1984) Ordination, multidimensional scaling and allied topics. 727-781 in: W. Lederman [ed.] *Handbook of Applicable Mathematics*. Vol. VI: E. Lloyd [ed.] *Statistics*. Wiley, Chichester.

Hankel, G.E., Molineri, C y Domínguez, E. (2019) Correspondence between stream benthic macroinvertebrate assemblages and ecoregions in northwestern Argentina. *Freshwater Science* 38, no. 1: 64-76. <https://doi.org/10.1086/701467>

He, F., y Legendre, P. (1996) On species-area relations. *Am. Nat.* 148: 719-737.

- He, F., y Legendre, P. (2002) Species diversity patterns derived from species-area models. *Ecology* 83:1185-1198.
- Hellinger, E. (1909) Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik* (en alemán). 136 : 210–271, doi : 10.1515 / crll.1909.136.210 , JFM 40.0393 .01
- Herrera Moreno, A. (2000) La clasificación numérica y su aplicación en la ecología. Instituto Tecnológico Santo Domingo, Santo Domingo. 88pp.
- Hill, M.O. (1974) Correspondence analysis: a neglected multivariate method. *Appl Stat.* 23:340-354.
- Hill, M.O. y Gauch, H.G. Jr (1980) Detrended correspondence analysis, an improved ordination technique. *Vegetatio* 42: 47-58.
- Hotelling, H. (1931) The Economics of Exhaustible Resources. *Journal of Political Economy*, 39 (2): 137-175.
- Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6): 417–441.
- Josse, J., Pagès, J. y Husson, F. (2008) Testing the significance of the RV coefficient. *Computational Statistics and Data Analysis* 53, 82–91.
- Jongman, R.H.G., ter Braak, C.J.F. y van Tongeren, O.F.R. (1987) Data analysis in community and landscape ecology. Pudoc, Wageningen, The Netherlands. Cambridge Univ. Press, Cambridge, England.
- Kilroy, C. y Unwin, M. (2011) The arrival and spread of the bloom-forming, freshwater diatom, *Didymosphenia geminata*, in New Zealand. *Aquatic Invasions* (2011) Volume 6, Issue 3: 249–262. doi: 10.3391/ai.2011.6.3.02
- Kroonenberg, P.M. (2008) Applied multi-way data analysis. Wiley & Sons, Inc., New Jersey. 579 pp. ISBN: 978-0-470-16479-6
- Le, S., Josse, J., y Husson, F. (2008) FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, 25(1), 1-18. 10.18637/jss.v025.i01
- Lebart, L., Morineau, A. y Piron, M. (1995) Statistique exploratoire multidimensionnelle, Dunod, Paris.
- Legendre, P., y Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271–280.
- Legendre, P., y Legendre, L. (1998) Numerical Ecology, 2nd English edition. Elsevier Science BV, Amsterdam, The Netherlands. 870pp

Legendre, P., y Legendre, L. (2012) Numerical Ecology, 3th English edition. Elsevier Science BV, Amsterdam, The Netherlands.990pp

Legendre, P., Oksanen, J. y ter Braak, C.J. (2011) Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*: 2, 269–277.

Lepš, J., y Šmilauer, P. (1999) Multivariate analysis of ecological data 110pp. <http://regent.bf.jcu.cz/textbook.pdf>

Lepš, J., y Šmilauer, P. (2003) Multivariate Analysis of Ecological Data Using CANOCO. Cambridge University Press, Cambridge.

Mauad, M., Miserendino, M.L., Risso, M.A., y Massaferro, J. (2015) Assessing the performance of macroinvertebrate metrics in the Challhuaco-Ñireco System (Northern Patagonia, Argentina). *Iheringia, Série Zoologia*, Porto Alegre, 105(3):348-358. doi: 10.1590/1678-476620151053348358

Macchi, P.A. (2017) Macroinvertebrados acuáticos como indicadores ecológicos de cambios en el uso del suelo en mallines del sudoeste de la Provincia de Río Negro. Tesis de Doctorado en Ciencias Naturales. FCNyM, UNLP. Argentina.

Macchi, P.A., Bernardis, A.M., Lavalle, A.L., Baeza, L., Encina, M., Mora, G.A., Navarro, M., y Saade, I. (2020) Invertebrados acuáticos como indicadores del estado ambiental del río Negro. *Revista de la Asociación Argentina de ecología de Paisajes* 9(1): 28-31.

Mahalanobis, P.C. (1936) On the Generalized Distance in Statistics. *Proceedings of the National Institute of Science of India*, 2: 49-55.

McCune, B. (1977) Influence of noisy environmental data on canonical correspondence analysis. *Ecology*, 78(8): 2617–2623.

Miserendino, M.L. y Pizzolón, L.A. (1999) Rapid assessment of river water quality using macroinvertebrates: a family level biotic index for patagonic Andean zone. *Acta Limnologica Brasiliensia* vol 11(2) 137-148.

Miserendino, M.L. (2001) Macroinvertebrate assemblages in Andean Patagonian rivers and streams. *Hydrobiologia* 444: 147–158.

Miserendino M.L., Brand, C. y Di Prinzio, C.Y. (2008) Assessing Urban Impacts on Water Quality, Benthic Communities and Fish in Streams of the Andes Mountains, Patagonia (Argentina). *Water Air Soil Pollution*: 194:91–110. DOI 10.1007/s11270-008-9701-4

Morison, M.L.V. (2010) Comparación entre el análisis canónico de correspondencias y el análisis factorial múltiple en tablas de frecuencias-variables

continuas. Tesis de Maestría en Ciencias Estadística. Universidad Nacional de Colombia.

Moncada Herrera, J. (2007) El Análisis Factorial Múltiple: Un Paso más en la Superación de la Dicotomía Cualitativo-Cuantitativo. VI Congreso Chileno de Antropología. Colegio de Antropólogos de Chile A. G, Valdivia.

Montecino, V., Molina, X., Kumar, S., Castillo, M.L.C., y Bustamante, R.O. (2014) Niche dynamics and potential geographic distribution of *Didymosphenia geminata* (Lyngbye) M. Schmidt, an invasive freshwater diatom in Southern Chile. Aquatic Invasions. Volume 9, Issue 4: 507–519. <http://dx.doi.org/10.3391/ai.2014.9.4.09>

Montecino V., Molina, X., Bothwell M.L., Muñoz, P., Carrevedo, M.L., Salinas F., y Bustamante, R.O. (2016) Spatio temporal population dynamics of the invasive diatom *Didymosphenia geminata* in central southern Chilean rivers. Science of the Total Environment, 568, 1135–1145. <https://doi.org/10.1016/j.scitotenv.2016.03.080>

Muñiz Saavedra J., Laffitte L.N., Morzenti, P., Rivera, D., Godoy L., Muñoz, J.L., Torres, M., y Lof Mapuche Vera. (2015) Monitoreo de Aguas Superficiales de la Subcuenca Trabunco-Quitrahue Comisión de Seguimiento Ambiental Decreto Pcial. N° 2092/08 y Res. N° 484/08. Informe Técnico DGBA. SEAyDS, Neuquén.

Oksanen, J., Guillaume Blanchet, F., Friendly, M. Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Szoecs, E. y Wagner, H. (2019) vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>

Pagès, J. (2014) Multiple Factor Analysis by Example Using R. Chapman & Hall/CRC. DOI: 10.1201/b17700

Palmer, M.W. (1993) Putting things in even better order: The advantages of canonical correspondence analysis. Ecology, 74(8): 2215-2230.

Park, Y.S., Lek, S., Baehr, C., y Jørgensen, S.E. (2015) Advanced Modelling Techniques Studying Global Changes in Environmental Sciences, Volume 27. 1st Edition: Elsevier. 380 pp. Hardcover ISBN: 9780444635365. eBook ISBN: 9780444635433.

Parmar, T.K., Rawtani, D., y Agrawal, Y.K. (2016) Bioindicators: the natural indicator of environmental pollution, Frontiers in Life Science, 9:2, 110-118, doi: 10.1080/21553769.2016.1162753

Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2:559-572.

Peña, D. (2002) *Análisis de datos Multivariantes*. Mc Graw Hill. Madrid. 539pp. ISBN 84-481-3610-1.

Pereyra, F.X., Lara, J., Carut, A., Muñiz Saavedra, J., Torre, F., Kalmbach, R., Oliva, R., Salaberry, G., Laffitte, L.N., Morzenti, P., Rivera, D., Castiñeira, L., y Tobío, M.I. (2016) *Estudio geocientífico aplicado al ordenamiento territorial*, San Martín de los Andes. Provincia del Neuquén. República Argentina. Instituto de Geología y Recursos Minerales, Servicio Geológico Minero Argentino. Serie Contribuciones Técnicas – Ordenamiento Territorial N° 9. 152 pp. Buenos Aires.

Rodríguez-Olarte, D., Barrios, M., Caputo, L., Fierro, P., Jiménez-Prado, P., Navarro, E., Macchi, P., Mojica, J. I., Molinero, J., Montoya, J. V., Pantoja, A., Pompêo, M., Ríos- Touma, B., Teixeira de Mello, F., Tobón, F., Torremorell, A., Villalba, A., Villamarín, C. (2020) *Criterios para la evaluación de estresores y parámetros en la estimación del estado ecológico de ríos en Suramérica*. Serie Publicaciones Especiales. Museo de Ciencias Naturales. Universidad Centroccidental Lisandro Alvarado (UCLA). Barquisimeto, Lara. Venezuela. 68 pp.

R Core Team. (2019) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Ramette, A. (2007) *Multivariate analyses in microbial ecology*. *FEMS Microbiol Ecol* 62 142–160. <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1574-6941.2007.00375.x>

Rao, C.R. (1952) *Advanced Statistical Methods in Biometric Research*. J. Willey, New York.

Rao, C.R. (1964) *The use and interpretation of principal component analysis in applied research*. *Sankhya* 26(1), 329–359.

Rao, C.R. (1995). *Use of Hellinger distance in graphical displays*. In E.-M. Tiit, T. Kollo, & H. Niemi (Ed.): *Multivariate statistics and matrices in statistics*. Leiden (Netherlands): Brill Academic Publisher. pp. 143–161

Reid, B., y Torres, R. (2014) *Didymosphenia geminata* invasion in South America: Ecosystem impacts and potential biogeochemical state change in Patagonian rivers. *Acta Oecologica*, 54, 101–109. <https://doi.org/10.1016/j.actao.2013.05.003>

Richardson, J.S. y Danehy, R.J.A. (2007) *Synthesis of the Ecology of Headwater Streams and their Riparian Zones in Temperate Forests*. *Forest Science* 53(2).

Robert, P. y Escoufier, Y. (1976) *A unifying tool for linear multivariate statistical methods: the RV-coefficient*. *Appl. Stat.–J. Roy. St. C* 25: 257-265

Smol, J.P., Birks, H.J., y Last, W.M. (2001) *Tracking Environmental Change Using Lake Sediments Volume 3: Terrestrial, Algal, and Siliceous Indicators*. Kluwer Academic Publishers, Dordrecht, The Netherlands. ISBN 978-0-306-47668-6.

Sokal, R. R. y Rohlf, F.J. (1995) *Biometry – The principles and practice of statistics in biological research*. 3rd edition. W. H. Freeman, New York.

Spearman, C.H. (1904) General intelligence objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

Stewart, D. y Love, W. (1968) A general canonical correlation index. *Psychol. Bull.* 70: 160-163.

Team Cowi. (1996) *Plan Maestro de Desagües Pluviales y de Desagües Cloacales*. Municipalidad de San Martín de los Andes, San Martín de los Andes.

ter Braak, C.J.F. (1985) CANOCO: A FORTRAN program for canonical correspondence analysis and detrended correspondence analysis. IWIS-TNO, Wageningen, The Netherlands.

ter Braak, C.J.F. (1986) Canonical correspondence analysis: A new technique for multivariate direct gradient analysis. *Ecology* 67(5).

ter Braak, C.J.F. (1987) The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* 69: 69-77.

ter Braak, C.J.F. y Prentice, I.C. (1988) A theory of gradient analysis. *Adv. Ecol. Res.* 18: 271-317.

ter Braak, C.J.F. y P. Šmilauer. (1998) *CANOCO Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (version 4)*. Microcomputer Power, Ithaca, NY, USA.

Thioulouse, J., Dray, S., Dufour, A.B., Siberchicot, A., Jombart, T., y Pavoine, S. (2018) *Multivariate Analysis of Ecological Data with ade4*. doi: 10.1007/978-1-4939-8850-1 ISBN: 978-1-4939-8848-8.

Trishala, K.P., Rawtani, D., y Agrawal, Y.K. (2016) Bioindicators: the natural indicator of environmental pollution, *Frontiers in Life Science*, 9:2, 110-118. doi: [10.1080/21553769.2016.1162753](https://doi.org/10.1080/21553769.2016.1162753)

Urdanigo, J.P., Díaz Ponce, M., Tay-Hing Cajas, C., Sánchez Fonseca, C., Yong Benitez, R., Armijo Albán, K., Guerrero Chúez, N. y Mancera-Rodríguez, N.J. (2019) Diversidad de macroinvertebrados acuáticos en quebradas con diferente cobertura ribereña en el bosque Protector Murocomba, Ecuador. *Revista de Biología Tropical*, 67(4), 861-878.

Whittaker, R.H. (1967) Gradient analysis of vegetation. *Biol. Rev. (Camb.)* 42: 207-264.

Wilks, S.S. (1932) Moments and Distributions of Estimates of Population Parameters from Fragmentary Samples. *Ann. Math. Statist.*, 3 (3): 163-195.

Wilks, S.S. (1935) The Likelihood Test of Independence in Contingency Tables. *Ann. Math. Statist.* 6 (4): 190-196. doi:10.1214/aoms/1177732564

Wollenberg, A.L. van den. (1977) Redundancy analysis. An alternative for canonical correlation analysis. *Psychometrika* 42: 207-219

Woodward, G. Gray, C., y Baird, D.J. (2013) Biomonitoring for the 21st Century: new perspectives in an age of globalisation and emerging environmental threats. *Limnetica*, 32 (2): 159-174. DOI: 10.23818/limn.32.14.

Zuur, A.F., Ieno, E.N. y Smith, G.M. (2007) *Analysing ecological data*. Springer. New York. 686pp.

ANEXO

A1. Descripción ampliada de áreas de estudio y metodologías de muestreo

A1.1. Biomonitorio cuenca Lacar - Hua Hum

A1.1.1. ÁREA DE ESTUDIO

El área de estudio se localiza dentro de la ecoregión Cordillera Patagónica septentrional (Pereyra *et al.*, 2016). En la cuenca Lacar el clima es frío húmedo, con temperatura media anual menor a 10 °C. Los vientos predominantes son generados en el anticiclón del Pacífico meridional con dirección oeste a este, que descargan su humedad al ascender por la cordillera. El régimen pluviométrico es muy variable, con precipitaciones entre 2000 y 2300 mm/año en la región piedemontana, 5000 mm/año en la zona cordillerana y menos de 1200 mm/año en la parte oriental de la cuenca (Rodríguez-Olarte *et al.*, 2020). Pereyra *et al.* (2016) mencionan precipitaciones de hasta menos de 600 mm en la zona oriental del área urbana.

La cuenca se encuentra en el Departamento Lacar, suroeste de la Provincia del Neuquén, República Argentina, y vierte sus aguas al océano Pacífico. Consiste en una gran depresión con eje este/oeste, rodeada por cordones montañosos de más de 2.000 m.s.n.m. del que drenan cursos en forma radiada. El punto más elevado es el Cerro Chapelco (2.394 m.n.m.), donde se emplaza uno de los centros de ski más importantes de Argentina. En las altas cumbres se encuentran las nacientes o cabeceras de cuenca de los arroyos que fluyen entre faldeos montañosos y humedales (o mallines), que, por último, atraviesan el casco céntrico de San Martín de los Andes. La red hídrica se conforma por el eje definido por el arroyo Maipu, arroyos Quitrahue, Trabunco y Calbuco que confluyen en el arroyo Pocahullo, el cual desemboca en el Lago Lacar. En el extremo occidental de la cuenca se encuentra el lago Nonthue, a partir del cual nace el río Hua Hum, que da origen al lago Pirehueico en Chile. A su vez, diversos arroyos confluyen en el

lago Lacar, por ejemplo el arroyo Tren Tren en playa Catritre y el arroyo Grande en el paraje Quila Quina, sobre su margen sur.

La red hídrica, que constituye el sistema de drenaje natural del área, comprende un conjunto de arroyos de carácter permanente y semipermanente de ambientes típicamente ritrónicos de montaña. El régimen hidrológico de los ríos y arroyos es pluvio-nival, con dos picos de crecida, uno en invierno y otro en primavera (periodo de deshielo). Desde el punto de vista biogeográfico, la cuenca se localiza en el área de transición o ecotono entre dos grandes regiones. Las provincias subantárticas de la región Antártica y, las provincias Patagónica y Altoandina de la región Neotropical. La vegetación dominante está formada por bosques de ciprés de la cordillera (*Austrocedrus chilensis*) y fagáceas como coihue (*Nothofagus dombeyi*), raulí (*N. nervosa*), roble pellín (*N. oblicua*), ñire (*N. antartica*) y lenga (*N. pumilio*) (Rodríguez-Olarte *et al.*, 2020).

Dentro de la cuenca confluyen varias jurisdicciones, comunidades de pueblos originarios, sectores de territorio provincial en las altas cumbres con arroyos de cabecera, y lindante al ejido de San Martín de los Andes, se emplaza el Parque Nacional Lanín.

El Municipio de San Martín de los Andes es parte de un aglomerado urbanístico dentro del denominado Corredor de los Lagos Andino – Patagónicos (Team Cowi, 1996). La localidad de San Martín de los Andes (SMA), ha experimentado un sostenido crecimiento poblacional en las últimas décadas, llegando a casi 40.000 habitantes según el censo del año 2010. El gran crecimiento poblacional implicó un avance progresivo sobre todo el ámbito del ejido municipal. En líneas generales la urbanización ha tenido un carácter desordenado por lo que actualmente se han generado numerosos problemas ambientales al ocuparse terrenos poco aptos. (Pereyra *et al.*, 2016)

El desarrollo económico de la localidad es principalmente turístico, por lo tanto, el mayor impacto sobre la cuenca es su uso para actividades recreativas, a lo que se suman las urbanizaciones, que se concentran cercanas al casco céntrico y en la denominada Vega Maipú. Rodríguez-Olarte *et al.* (2020) mencionan que el mayor riesgo ecológico de los sistemas acuáticos a lo largo de la cuenca se

relaciona con el impacto combinado del cambio de uso de la tierra, expansión urbana (ej. vertido de efluentes, drenaje de zonas para construcción), actividades en las riberas y llanuras de inundación, canalizaciones, deforestación, agricultura, con efectos adversos en calidad de agua y sedimento de los ríos y lagos.

A1.1.2. METODOLOGÍA DE MUESTREO

En cada sitio se procedió al registro de parámetros fisicoquímicos mediante el empleo de medidores multiparamétricos marca HACH, modelo HQ 40d, con sus electrodos correspondientes: conductividad eléctrica ($\mu\text{S}/\text{cm}$) “CE”, oxígeno disuelto (mg/L) “OD”, nivel de acidez o alcalinidad “PH”, temperatura del agua ($^{\circ}\text{C}$) “TAG”.

Para la obtención de las muestras de agua para análisis químico de iones y nutrientes se utilizaron bidones de 1 y 2 L de capacidad de polietileno de alta densidad. Se mantuvieron congelados a -18°C hasta su envío al laboratorio GECARA del CRUB-UNCo.

Se utilizaron macroinvertebrados como bioindicadores de la calidad ecológica del agua. Para la obtención de las muestras de macroinvertebrados bentónicos en cada lugar se identificó un tramo de 100 m sobre el cual se establecieron 11 transectas cada 10 m de distancia. La muestra por sitio estuvo conformada por las 11 submuestras de cada transecta. La recolección fue efectuada en zonas de profundidad no mayor a 40 cm (tomado como referencia aproximada la altura del codo al sumergir el brazo) y se utilizó una red de tipo D-net de $500\ \mu\text{m}$ de abertura de malla (Figura 3.2_d) cuya área del cuadrante fue de $30,5\ \text{cm}^2$. La red fue apoyada en el fondo del lecho, y dentro del área definida por su marco, los operadores procedieron a limpiar las piedras con las manos sin sacarlas del agua o a remover el sustrato con los pies por 15 segundos aproximadamente, capturando así los organismos que viven en el lecho del río. El material recolectado se colocó en un envase rotulado y se le adicionó alcohol etílico al 70%, para su conservación, hasta su envío al laboratorio CIRA-UNCo. Mayor detalle sobre el muestreo se describe en Flotemersch *et al.* (2017).

A1.2. Monitoreo del alga exótica invasora *Didymosphenia geminata*

A1.2.1. ÁREA DE ESTUDIO

La provincia de Neuquén presenta una importante heterogeneidad natural, dada principalmente por dos gradientes climáticos: el de precipitaciones que varía en sentido longitudinal, con abundantes lluvias hacia el límite con Chile, que decrecen rápidamente hacia el Este; y el de temperaturas que está asociado con la altimetría, con temperaturas más bajas a medida que la altura es mayor. A estos gradientes se suma la complejidad geomorfológica de la provincia que determina diferentes tipos de suelos y sistemas de drenaje (Bran *et al.*, 2002). Bran *et al.* (2002), amplía las divisiones de las provincias fitogeográficas de Cabrera (1971) y define para la provincia del Neuquén 6 áreas ecológicas: Cordillera Norte, Cordillera Sur, Precordillera, Sierras y Mesetas Norte, Sierras y Mesetas Sur, Monte Austral.

Los ambientes acuáticos de la provincia y de la Patagonia, en general, se ubican dentro de la categoría oligotrófica a mesotrófica, con aguas circunneutras y bien oxigenadas (Díaz *et al.*, 2007).

Los ríos Neuquén y Limay son regulados, desde aguas arriba hacia aguas abajo, sobre el río Limay se encuentran los Embalses de Alicurá, Piedra del Águila, Pichi Picún Leufú, Exequiel Ramos Mexía y Arroyito y sobre el río Neuquén, Los Barreales, Mari Menuco y El Chañar. Las principales actividades productivas son la generación de energía hidroeléctrica, extracción de petróleo y gas, fruticultura, turismo, minería y ganadería, las que en distinto grado vuelcan sus efluentes tratados o sin tratar en estos cursos de agua. (AIC, 2007).

La cuenca del río Neuquén tiene una superficie de 39.731 km², con un régimen de escurrimiento pluvionival. El periodo de estiaje se presenta entre marzo-abril y las dos avenidas, una en invierno entre mayo-agosto (debido a las lluvias en la parte baja de la cuenca activa) y otra de primavera entre noviembre-diciembre

(producto de la fusión nival en la alta cordillera). La carencia de lagos en cabeceras, la escasa vegetación y las fuertes pendientes hacen que este río tenga un carácter torrencial (AIC, 2007). El río Neuquén se encuentra con el río Limay cerca de la ciudad de Neuquén para formar el río Negro, que continúa su camino hacia el este hasta el Océano Atlántico.

La cuenca del río Limay, con una superficie de 61.597 km², la mayor parte del caudal del río se origina en las precipitaciones invernales que se producen en la alta cuenca, principalmente en forma de lluvia y nieve. Esta cuenca posee un alto grado de regulación natural debido a la presencia de vegetación boscosa y numerosos lagos de cabecera en la zona cordillerana (AIC, 2007).

El río Colorado se origina en la confluencia de los principales afluentes, los ríos Grande y Barrancas, y recorre alrededor de 1200 km atravesando parte de las provincias de Mendoza, Neuquén, Río Negro, La Pampa y Buenos Aires. Tiene un derrame anual promedio de 4.380 hm³ y la superficie total de su cuenca es de 47.459 km². Su caudal medio anual es de 138,8 m³/s presentando una marcada estacionalidad primavero-estival, atribuida a su régimen nival. El río Colorado adquiere la tonalidad que le da nombre, a partir de los sedimentos que provienen de secuencias clásticas (areniscas, pelitas y conglomerados). El río Colorado es uno de los cauces argentinos con mayor cantidad de arrastre promedio, pues llega comúnmente hasta 2 kg/m³ (Aumassane, 2019). Atraviesa una diagonal árida, siendo este río el principal recurso para el abastecimiento para uso urbano, agropecuario, minería e industrial.

La descripción detallada de la cuenca del río Hua Hum se presentó en el apartado A1.1.1.

A1.2.2. METODOLOGÍA DE MUESTREO

Los lugares de muestreo se seleccionaron teniendo en cuenta las actividades humanas; la viabilidad de que la especie pueda llegar, instalarse y desarrollarse; la fácil accesibilidad y el valor sociocultural del sitio. Se asumió que los sitios con el mayor riesgo de contaminación son los puntos de acceso público y relacionados con las actividades pesqueras.

En cada sitio se procedió al registro de parámetros fisicoquímicos mediante el empleo de medidores multiparamétricos marca HACH, modelo HQ 40d, con sus electrodos correspondientes: conductividad eléctrica ($\mu\text{S}/\text{cm}$) “CE”, oxígeno disuelto (mg/L) “OD”, nivel de acidez o alcalinidad “PH”, temperatura del agua ($^{\circ}\text{C}$) “TAG”.

Para la obtención de las muestras de agua para análisis químico de iones y nutrientes se utilizaron bidones de 1 y 2 L de capacidad de polietileno de alta densidad. Se mantuvieron congelados a -18°C hasta su envío al laboratorio GECARA del CRUB-UNCo

Las muestras multihabitat (M) se extrajeron mediante el raspado y enjuague de cantos rodados, restos leñosos y vegetación acuática. Luego de su recolección, fueron fijadas en el campo con formol al 40% y posteriormente remitidas al Laboratorio GECARA (CRUB – UNCo), para su análisis. Mayor detalle del procedimiento se encuentra en el protocolo publicado en el año 2012 (AIC, 2012) y en Beamud *et al.* (2019).

A2. Tablas

Tabla I. Detalle de los 21 sitios muestreados, años 2015 y 2016, incluyendo: subcuenca (TQ: Trabunco-Quitrahue, P: Pocahullo, L: Lacar-Hua Hum), sitio, nombre, código de muestra e índice BMPS según año de muestreo. Monitoreo LHH.

Sub cuenca	Sitio	Nombre	Código Año 2015	Código Año 2016	BMPS Año 2015	BMPS Año 2016
TQ	PP0	Vertiente del Mocho. Punto de referencia o blanco.		01PP0		103
TQ	PP2	Arroyo Pradera del Puma. Base cuádruple.	02PP2	03PP2	84	95
TQ	PP3	Arroyo Pradera del Puma. Toma Las Pendientes.		04PP3		129
TQ	Ga3	Arroyo la Gamela debajo de AB5. En Paso Los Coihues.	05Ga3		128	
TQ	Ga2	Arroyo de La Gamela. Puesto Vialidad.		06Ga2		110
TQ	Q5	Arroyo de la Cascada. Escuela.	07Q5	08Q5	122	138
TQ	AB1	Arroyo Blanco. Casa Fidel.		09AB1		130

TQ	T0	Arroyo Trabunco. Toma Valle escondido.	11T0	10T0	125	134
TQ	Q1	Arroyo Quitrahue, nn Panquehue.		13Q1		120
P	ACH1	Arroyo Chapelco Chico en Golf.	14ACH		91	
P	ACH2	Arroyo Chapelco Chico en Pio Proto arriba.	15ACH		79	
P	ACH3	Arroyo Chapelco Chico en Pio Proto.		16ACH		116
P	CQ	Canal derivado del rio Quilquihue.		17CQ		105
P	PM2	Arroyo Maipu. En Callejón de Bello.	18PM2	19PM2	73	113
P	A5	Arroyo Calbuco. Puente camino Hua-Hum.	20A5	21A5	87	46
P	A5	Arroyo Calbuco. Puente camino Hua-Hum. (*remuestreo)		22A5*		49
TQ	T3	Arroyo Trabunco. Desembocadura en Pocahullo.	23T3	24T3	66	51
P	A10	Arroyo Pocahullo. Calle Juez del Valle.		25A10		27
P	A11	Arroyo Pocahullo, desembocadura en lago Lacar.	26A11	27A11	22	18
L	CAT	Arroyo Tren Tren. Desembocadura Catritre.		28CAT		97
L	QQ	Arroyo Grande. Desembocadura Quila Quina.		29QQ		111
L	HH	Rio Hua Hum. Puente.		30HH		168

Tabla II. Escala de calidad de agua correspondiente al Índice BMPS (Miserendino y Pizzolón, 1999).

Clase	Calidad	Valor	Significado	Color
I	Buena	Mayor a 150	Aguas muy limpias	Azul
		Mayor a 101	Aguas no contaminadas	Azul
II	Aceptable	61-100	Probablemente contaminación incipiente u otra clase de perturbación	Verde
III	Dudosa	36-60	Probablemente aguas contaminadas	Amarillo
IV	Crítica	16-35	Aguas contaminadas	Naranja
V	Muy crítica	Menor que 15	Aguas fuertemente contaminadas	Rojo

Tabla III. Listado completo de familias de macroinvertebrados incluyendo grupo, taxón, abreviatura utilizada en gráficos, puntuación del grado de sensibilidad a la contaminación (índice BMPS) y frecuencia absoluta. Monitoreo LHH.

Grupo	Taxón (familia)	Abreviatura	Índice BMPS	Abundancia (frecuencia absoluta)
Diptera	Blepharacerae	"Bleph"	10	5
Plecoptera	Eustheniidae	"Eusth"	10	10
Plecoptera	Diamphipnoidae	"Diamp"	10	18
Ephemeroptera	Ameletopsidae	"Amele"	10	58
Trichoptera	Leptoceridae	"Leptoc"	10	61
Trichoptera	Sericostomidae	"Seric"	10	119
Plecoptera	Notoneumoridae	"Noton"	10	132
Diptera	Athericidae	"Ather"	10	684
Plecoptera	Austroperlidae	"Austr"	10	1025
Ephemeroptera	Leptophlebiidae	"Leptop"	10	1316
Plecoptera	Gripopterygidae	"Gripo"	10	1721
Coleoptera	Psephenidae	"Pseph"	7	9
Odonata	Gomphidae	"Gomph"	7	13
Trichoptera	Limnephilidae	"Limne"	7	75
Trichoptera	Hydrobiosidae	"Hydrob"	7	251
Megastropoda	Chiliniidae	"Chili"	6	95
Amphipoda	Hyalellidae	"Hyale"	6	101
Ephemeroptera	Baetidae	"Baeti"	6	471
Megaloptera	Corydalidae	"Coryd"	5	3
Decapoda	Aegliidae	"Aegli"	5	19
Coleoptera	Scirtidae	"Scirt"	5	25
Turbellaria	Dugesidae	"Duges"	5	56
Trichoptera	Hydroptilidae	"Hydropt"	5	57
Diptera	Tipulidae	"Tipul"	5	109
Diptera	Simuliidae	"Simul"	5	864
Coleoptera	Elmidae	"Elmid"	5	1312
Trichoptera	Hydropsychidae	"Hydrops"	5	2961
Arachnida	Hydrachnidia	"Hydra"	4	2
Diptera	Ceratopogonidae	"Cerat"	4	13

Diptera	Empididae	"Empid"	4	49
Diptera	Tabanidae	"Taban"	4	131
Heteroptera	Corixidae	"Corix"	3	1
Coleoptera	Dytiscidae	"Dytis"	3	2
Coleoptera	Hydrophilidae	"Hydroph"	3	4
Gastropoda	Physidae	"Physi"	3	10
Hemiptera	Saldidae	"Saldi"	3	10
Mollusca	Sphaeriidae	"Sphae"	3	14
Mollusca	Limnaeidae	"Limej"	3	27
Caenogastropoda	Cochliopidae	"Cochl"	3	36
Coleoptera	Gyrinidae	"Gyrin"	3	81
Hirudinea	Glossiphoniidae	"Gloss"	3	134
Diptera	Ephydriidae	"Ephyd"	2	4
Diptera	Chironomidae	"Chiro"	2	4668
Annelida	Naididae	"Naidi"	1	132
Annelida	Lumbriculidae	"Lumbr"	1	231
Annelida	Tubificidae	"Tubif"	1	647

Tabla IV. Estadísticos resumen de variables ambientales. E.E.: error estándar, Mín: mínimo, Máx.: máximo. Monitoreo LHH.

Variable	n	Media	E.E.	Mín	Máx
CE	29	120,61	9,68	26	257
OD	29	9,14	0,14	7,03	10,48
PH	29	7,72	0,06	6,91	8,31
TAG	29	11,41	0,57	4,7	16,7
ALT	29	902,14	62,59	642	1687
NT	29	371,38	90,91	33	1879
PRS	29	41,39	13,67	0,9	268
SST	29	5,15	0,83	0,1	20,3
FE	29	0,09	0,02	0,01	0,37
SI	29	19,7	0,85	10,4	27,8
DU	29	40,97	3,05	10,3	64,6
ALC	29	44,1	3,02	13	70
PEH	29	120,52	5,66	55	166

Tabla V. Valores observados de variables ambientales por sitio. Monitoreo LHH.

Sitio	CE	OD	PH	TAG	ALT	NT	PRS	SST	FE	SI	DU	ALC	PEH
01PP0	36,70	9,78	8,31	4,70	1687	70,00	17,70	0,20	0,01	11,80	11,20	13,00	138,00
02PP2	26,00	8,27	7,95	9,00	1666	33,00	15,30	2,40	0,02	12,90	10,30	14,00	130,00
03PP2	38,20	9,48	7,57	5,70	1666	66,00	6,30	20,30	0,07	12,40	11,50	14,00	119,00
04PP3	46,70	9,47	6,91	4,90	1570	83,00	2,70	0,10	0,01	11,30	12,70	16,00	148,00
05Ga3	159,30	8,79	7,93	10,00	1064	1862,00	8,70	7,10	0,01	21,60	64,60	66,00	165,00
06Ga2	114,40	7,03	8,16	11,00	1299	157,00	1,30	6,50	0,09	20,80	36,50	40,50	90,00
07Q5	102,50	8,81	7,21	15,00	860	293,00	1,40	8,50	0,08	20,20	44,60	50,00	125,00
08Q5	110,20	9,88	7,64	9,10	860	271,00	0,90	1,00	0,06	19,10	50,50	45,00	88,00
09AB1	153,00	9,47	7,89	12,10	805	171,00	2,70	10,80	0,02	18,70	52,10	43,50	138,00
10T0	105,50	9,64	7,75	9,70	980	95,00	14,40	0,70	0,01	19,60	28,30	38,00	129,00
11T0	77,40	9,58	7,89	14,00	980	88,00	15,50	1,10	0,01	20,80	29,50	39,50	165,00
13Q1	161,90	9,73	8,25	10,00	784	547,00	1,30	6,20	0,05	21,30	53,60	51,00	144,00
14ACH	144,20	8,63	7,69	11,10	819	161,00	11,10	10,40	0,02	23,60	64,60	70,00	100,00
15ACH	114,10	8,49	7,89	11,50	772	39,00	15,30	2,70	0,10	24,40	57,40	65,50	166,00
16ACH	149,50	9,23	7,73	13,00	769	80,00	19,90	9,90	0,10	22,50	49,00	49,00	115,00
17CQ	75,60	8,83	7,38	15,20	769	55,00	4,40	4,40	0,07	14,90	18,70	25,00	137,00
18PM2	108,10	8,33	7,32	12,00	725	138,00	7,50	3,10	0,27	23,40	43,50	52,50	109,00
19PM2	147,50	8,28	7,64	14,60	725	127,00	4,20	3,30	0,37	21,10	45,40	49,00	104,00
20A5	108,60	8,53	7,41	13,50	719	801,00	207,00	2,50	0,22	27,80	48,30	68,00	113,00
21A5	140,70	9,76	7,64	9,80	719	313,00	85,90	1,70	0,18	23,90	45,00	45,00	122,00
22A5	131,90	9,47	7,47	10,60	719	431,00	78,10	6,40	0,24	24,60	47,00	45,50	118,00
23T3	126,10	9,66	7,34	16,00	664	589,00	10,20	6,40	0,05	21,20	45,20	54,00	89,00
24T3	181,60	10,42	8,16	11,10	664	536,00	5,90	6,80	0,04	21,70	54,20	50,00	114,00
25A10	195,10	9,26	7,52	12,90	647	684,00	186,80	9,10	0,19	23,70	52,70	48,00	66,00
26A11	192,80	8,20	7,40	14,00	642	1879,00	189,80	4,50	0,15	23,50	53,90	58,50	55,00
27A11	257,00	8,69	7,97	13,60	642	1028,00	268,00	9,10	0,16	21,10	59,40	60,00	60,00
28CAT	120,70	10,48	8,27	9,50	646	65,00	16,10	3,30	0,05	19,00	37,50	43,00	138,00
29QQ	101,40	10,25	7,94	10,60	643	63,00	0,90	0,80	0,01	14,00	40,90	41,00	152,00
30HH	71,10	8,57	7,78	16,70	657	45,00	0,90	0,10	0,03	10,40	19,90	24,50	158,00

Tabla VI. Listado completo de sitios incluyendo código de sitio (utilizado en gráficos), cuenca y nombre. Monitoreo DIDYMO.

Código de sitio	Cuenca	Nombre del sitio
01LHH	LHH	Arroyo Pocahullo. En desembocadura.
02LHH	LHH	Río Hua Hum. Puente.
04RC	RC	Río Colorado. En El Porton.
05RC	RC	Río Colorado en Rincón de los Sauces.

03RL	RL	Arroyo El Once. Desembocadura en lago Nahuel Huapi.
06RL	RL	Río Caleufu. En hito.
07RL	RL	Río Chimehuín. Junín (caño EPAS).
08RL	RL	Río Chimehuín boca.
10RL	RL	Río Currhue. Naciente.
11RL	RL	Río Filo Hua Hum Confluencia Meliquina
12RL	RL	Río Hermoso. Inferior.
13RL	RL	Río Hermoso. Superior (naciente).
14RL	RL	Río Limay toma de agua de Arroyito.
16RL	RL	Río Limay naciente (lado NQN).
17RL	RL	Río Malleo puente RP 23.
18RL	RL	Río Malleo superior.
19RL	RL	Río Meliquina (naciente).
20RL	RL	Río Pulmarí superior.
21RL	RL	Río Quillén inferior.
22RL	RL	Río Quillén superior.
23RL	RL	Río Quilquihue. Confluencia con río Chimehuin.
24RL	RL	Río Quilquihue naciente.
25RL	RL	Río Ruca Choroí inferior (desembocadura).
26RL	RL	Río Ruca Choroí superior.
27RL	RL	Río Traful (debajo de Cullin Manzano).
28RL	RL	Río Villarino en Falkner.
09RN	RN	Río Curí Leuvú. Puente en Chos Malal.
15RN	RN	Río Neuquén pre-confluencia con Limay.
29RN	RN	Río Agrío aguas arriba puente Las Lajas.
30RN	RN	Río Agrío en Loncopue.
31RN	RN	Río Guañacos en Guañacos.
32RN	RN	Río Lileo.
33RN	RN	Río Nahueve superior.
34RN	RN	Río Neuquén. Chacra El Sauce.
35RN	RN	Río Neuquén. Aguas abajo Dique Chañar.
36RN	RN	Río Neuquén. Chos Malal. Puente RN 40.
37RN	RN	Río Neuquén. Aguas arriba Dique Chañar.
38RN	RN	Río Neuquén. En paraje Cuatro Esquinas.
39RN	RN	Río Neuquén. En calle Figueroa de ciudad de Neuquén.
40RN	RN	Río Limay. Pre-confluencia con Neuquén.

41RN	RN	Río Neuquén. En Sauzal Bonito.
42RN	RN	Río Neuquén. En toma de agua Añelo.
43RN	RN	Río Reñileuvú.
44RN	RN	Río Varvarco. Aguas arriba de localidad Varvarco.

Tabla VII. Listado completo de especies de algas incluyendo grupo taxonómico, taxón, abreviatura utilizada en gráficos y frecuencia relativa porcentual.

Monitoreo DIDYMO.

Grupo taxonómico (clase)	Nombre completo (especie)	Abreviatura	Abundancia (frecuencia relativa porcentual)
Bacillariophyceae	Achnanthes	"Achnan"	7,26
Bacillariophyceae	Achnanthidium	"Achnanthid"	58,84
Bacillariophyceae	Asterionella	"Asteri"	5,46
Bacillariophyceae	Aulacoseira	"Aulaco"	281,27
Bacillariophyceae	Cocconeis	"Coccon"	123,48
Bacillariophyceae	Cyclotella	"Cyclot"	31,23
Bacillariophyceae	Cymatopleura	"Cymato"	0,12
Bacillariophyceae	Cymbella	"Cymbel"	175,99
Bacillariophyceae	Diatoma	"Diatom"	265,2
Bacillariophyceae	Didymosphenia	"Didymo"	0,88
Bacillariophyceae	Diploneis	"Diplon"	8,17
Bacillariophyceae	Epithemia	"Epithe"	254,82
Bacillariophyceae	Fragilaria	"Fragil"	331,11
Bacillariophyceae	Frustulia	"Frustu"	0,35
Bacillariophyceae	Gomphoneis	"Gomphoneis"	83,14
Bacillariophyceae	Gomphonema	"Gomphonema"	333,18
Bacillariophyceae	Gyrosigma	"Gyrosi"	0,89
Bacillariophyceae	Hannaea	"Hannae"	9,80
Bacillariophyceae	Navicula	"Navicu"	321,18
Bacillariophyceae	Nitzschia	"Nitzsc"	105,37
Bacillariophyceae	Pinnularia	"Pinnul"	10,17
Bacillariophyceae	Rhoicosphenia	"Rhoico"	22,54
Bacillariophyceae	Rhopalodia	"Rhopal"	93,74
Bacillariophyceae	Surirella	"Surire"	0,12
Bacillariophyceae	Synedra	"Synedr"	89,98

Bacillariophyceae	Tabellaria	"Tabell"	263,09
Chlorophyceae	Chaetophora	"Chaeto"	3,72
Chlorophyceae	Closterium	"Closte"	2,21
Chlorophyceae	Cosmarium	"Cosmar"	14,09
Chlorophyceae	Monoraphidium	"Monora"	3,74
Chlorophyceae	Oedogonium	"Oedogo"	0,95
Chlorophyceae	Pediastrum	"Pedias"	4,44
Chlorophyceae	Scenedesmus	"Scened"	31,04
Chlorophyceae	Sphaerocystis	"Sphaer"	7,58
Chlorophyceae	Ulothrix	"Ulothr"	17,05
Cyanophyceae	Calothrix	"Caloth"	595,76
Cyanophyceae	Dolichospermum	"Dolich"	321,55
Cyanophyceae	Lygnbya	"Lygnby"	287,8
Cyanophyceae	Merismopedia	"Merism"	12,75
Cyanophyceae	Nostoc	"Nostoc"	1,26
Cyanophyceae	Oscillatoria	"Oscill"	139,85
Trebouxiophyceae	Oocystis	"Oocyst"	1,70
Zygnematophyceae	Mougeotia	"Mougeo"	48,63
Zygnematophyceae	Zygnema	"Zygnem"	28,67

(*) Los valores de abundancia para cada especie representan la sumatoria de la frecuencia relativa porcentual en todos los sitios donde estuvo presente. Guiry, G.M. 2020. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. <http://www.algaebase.org>.

Tabla VIII. Estadísticos resumen de variables ambientales. E.E.: error estándar, Mín: mínimo, Máx.: máximo. Monitoreo DIDYMO.

Variable	n	Media	E.E.	Mín	Máx
CE	46	161,46	38,48	30,50	1219,00
OD	46	9,41	0,13	7,45	11,94
PH	46	7,71	0,07	6,52	8,79
TAG	46	12,89	0,59	3,40	21,00
PT	46	48,29	7,02	9,70	184,00
SST	46	33,83	9,35	0,20	303,00
ALC	46	38,95	4,16	13,00	144,50
PEH	46	125,20	4,00	66,00	170,00

Tabla IX. Estadísticos resumen de variables ambientales por cuencas. E.E.: error estándar, Mín: mínimo, Máx.: máximo. Monitoreo DIDYMO.

	CE			
Cuenca	LHH	RC	RL	RN
Media	74,85	1207,50	64,25	191,55
E.E.	26,45	11,50	10,36	44,72
Mín	48,40	1196,00	30,50	30,50
Máx	101,30	1219,00	273,00	679,00
n	2	2	25	17
	OD			
Cuenca	LHH	RC	RL	RN
Media	10,62	8,57	9,61	9,07
E.E.	1,32	0,11	0,12	0,22
Mín	9,29	8,46	8,62	7,45
Máx	11,94	8,69	10,76	11,85
n	2	2	25	17
	PH			
Cuenca	LHH	RC	RL	RN
Media	7,56	7,92	7,53	7,96
E.E.	0,05	0,03	0,08	0,13
Mín	7,51	7,89	6,52	6,91
Máx	7,61	7,94	8,37	8,79
n	2	2	25	17
	TAG			
Cuenca	LHH	RC	RL	RN
Media	8,15	13,50	11,65	15,19
E.E.	4,75	1,50	0,66	0,93
Mín	3,40	12,00	6,80	9,00
Máx	12,90	15,00	18,10	21,00
n	2	2	25	17
	PT			
Cuenca	LHH	RC	RL	RN
Media	42,90	130,95	21,30	78,88
E.E.	31,30	44,15	2,71	12,69
Mín	11,60	86,80	9,70	18,50
Máx	74,20	175,10	62,60	184,00
n	2	2	25	17
	SST			
Cuenca	LHH	RC	RL	RN
Media	6,45	150,35	13,97	52,55
E.E.	5,55	51,65	9,10	17,30
Mín	0,90	98,70	0,20	9,40

Máx	12,00	202,00	230,80	303,00
n	2	2	25	17
	ALC			
Cuenca	LHH	RC	RL	RN
Media	30,5	66,25	25,74	56,15
E.E.	9,50	1,25	4,56	6,65
Mín	21,00	65,00	13,00	23,00
Máx	40,00	67,50	131,50	144,50
n	2	2	25	17
	PEH			
Cuenca	LHH	RC	RL	RN
Media	113	96	135	116
E.E.	44,5	11,5	5,0	5,4
Mín	68	85	66	71
Máx	157	108	170	158
n	2	2	25	17

Tabla X. Valores observados de variables ambientales por sitio. Monitoreo DIDYMO.

Sitio	Cuenca	CE	OD	PH	TAG	PT	SST	PEH
1LHH	LHH	101,3	11,94	7,51	3,4	74,2	12	68
2LHH	LHH	48,4	9,29	7,61	12,9	11,6	0,9	157
4RC	RC	1.196,00	8,69	7,89	12	86,8	98,7	108
5RC	RC	1.219,00	8,46	7,94	15	175,1	202	85
3RL	RL	80,8	9,97	6,52	6,8	30,1	26,1	68
6RL	RL	49,7	10,14	7,83	11,2	19,8	9,6	139
7RL	RL	44,6	9,91	7,57	10,4	12,3	1	129
8RL	RL	42,3	9,63	7,80	12	13,4	4,4	151
9RL	RL	273	9,63	8,37	16	47,3	230,8	66
10RL	RL	70,5	10,16	7,98	7,4	12,9	4,7	140
11RL	RL	44,2	10,14	7,26	9,6	15,8	13,3	170
12RL	RL	38,8	10,76	7,71	7,8	15,8	2,1	125
13RL	RL	33,1	9,83	7,19	9,1	12,6	0,6	156
14RL	RL	37,4	9,18	7,17	8	20	5,2	147
15RL	RL	38	9,18	7,46	15,6	22,2	7,1	122
16RL	RL	30,5	9,89	7,67	11	9,7	0,7	141
17RL	RL	56,7	9,77	7,95	13,3	62,6	2,7	139
18RL	RL	36,1	10,17	8,12	10,7	48,1	3	163
19RL	RL	43,1	9,5	6,90	10,1	15,5	4,4	139

20RL	RL	48,5	8,79	7,26	18,1	13,9	1	126
21RL	RL	159,3	8,79	7,93	10	11,7	1,3	159
22RL	RL	91,2	8,81	7,51	12	13,6	0,7	152
23RL	RL	102,5	8,81	7,21	15	14,5	0,6	107
24RL	RL	61,6	8,62	7,09	16,2	22,8	3,2	158
25RL	RL	42,6	9,26	7,54	14,6	21	7,5	139
26RL	RL	47	8,76	7,38	17,5	37	9,9	150
27RL	RL	54	10,67	7,96	8	10,5	2	128
28RL	RL	47,9	10,03	7,17	11,3	14,5	7,2	136
29RL	RL	32,8	9,83	7,68	9,6	14,9	0,2	121
30RN	RN	160,4	9,27	8,32	12	85	27,6	90
31RN	RN	176,5	8,95	8,68	13	85	9,4	123
32RN	RN	34,7	8,84	8,04	9	68,2	17,1	108
33RN	RN	31,3	8,84	7,29	10	118,6	70,6	128
34RN	RN	30,5	8,87	7,83	17,1	29,3	10,5	136
35RN	RN	63,4	8,86	8,25	10	160,5	107,7	96
36RN	RN	394	11,85	8,79	21	76,1	56,1	99
37RN	RN	253	10,28	8,78	20	18,5	10,7	134
38RN	RN	72,2	8,89	7,42	17,5	184	84,1	101
39RN	RN	679	7,45	7,77	18	70	48,2	129
40RN	RN	49,1	9,22	8,04	15,7	37,6	16	132
41RN	RN	87,3	9,16	7,18	12,3	21	13,7	128
42RN	RN	96	9,16	8,00	20,8	31,5	9,4	128
43RN	RN	359	8,42	8,10	17,8	174,2	67	71
44RN	RN	370	8,48	8,31	17	70,2	30,5	121
45RN	RN	363	8,86	7,58	14	38,5	11,7	158
46RN	RN	36,9	8,87	6,91	13,1	72,8	303	88

Tabla XI. Planilla de campo para elaboración del índice cualitativo “Puntaje de evaluación del hábitat – PEH”. Adaptado de Barbour *et al.*, 1999.

Parámetro del hábitat	Categoría			
	ÓPTIMA	SUBÓPTIMA	MARGINAL	POBRE
1.ESTABILIDAD DEL HÁBITAT Y DESARROLLO DE LA BIOTA	Más del 70 % del hábitat es estable y favorece el asentamiento de la biota, incluyendo áreas potenciales para el refugio de peces.	Hasta un 40 % del hábitat es una mezcla de un hábitat estable mixto con un sustrato adicional reciente, aún no adecuado para la colonización por la biota (el porcentaje puede estar en el límite superior de la escala)	Hasta un 20 % del hábitat es una mezcla de un hábitat estable con un sustrato frecuentemente perturbado o removido	Menos del 20 % del hábitat es estable. La falta de hábitat es obvia, el sustrato es inestable o falta.
PUNTUACIÓN	Presencia de sustratos y espacios variados y diversos (hábitat mixto).			

	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
2.FIJACIÓN DEL SUSTRATO (componentes inorgánicos del lecho)	Tamaños 1, 2 y 3 rodeados como máximo por un 25% de sedimentos finos.					Tamaños 1, 2 y 3 ocupan del 50 al 75 % y los sedimentos finos que los rodean ocupan del 25 al 50 %					Tamaños 1, 2 y 3 ocupan entre 25 y 50 % y los sedimentos finos que los rodean ocupan del 50 al 75 %.					Tamaños 1, 2 y 3 ocupan menos del 25 % y los sedimentos finos más del 75 %.					
PUNTUACIÓN	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
3.VELOCIDAD/ PROFUNDIDAD	Todas las velocidades/ profundidades (lento-profundo, lento-somero, rápido-profundo, rápido-somero) (lento es menor a 0,3 m/seg., profundo es mayor a 0,5 m)					Sólo 3 de los 4 regímenes presentes (si el rápido-somero falta el valor baja más que si faltan los otros regímenes)					Sólo 2 de los 4 regímenes presentes (si rápido-somero o lento-somero faltan, el valor baja)					Dominado por 1 régimen de velocidad/profund. (usualmente lento-profundo)					
PUNTUACIÓN	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
4.DEPOSITACIÓN DE SEDIMENTOS	Menos del 5% del fondo afectado por depositación de sedimentos. Sin o pocas islas o barras puntuales					Del 5 al 30 % del fondo afectado por depositación de sedimentos. Ligera depositación en los pozones. Indicios de formación de barras, principalmente de tamaño 3 y menores (sedimentos finos).					Del 30 al 50 % del fondo con sedimentos; depósitos en curvas, obstrucciones, en los pozones predomina una depositación moderada. Presencia evidente de depósitos nuevos de sedimentos finos sobre barras viejas o nuevas.					Más del 50% del fondo cambiando frecuentemente. Fuertes depósitos de material fino. Aumento del desarrollo de las barras. Pozones casi ausentes debido a una depositación sustancial de sedimento.					
PUNTUACIÓN	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
5. ESTADO ACTUAL DEL NIVEL DEL AGUA	El nivel del agua alcanza la base de ambas riberas y una mínima cantidad de los sustratos están expuestos					El área mojada ocupa más del 75% del cauce disponible o menos del 25% de los sustratos están expuestos.					El área mojada comprende entre un 25 y el 75% del cauce disponible y los sustratos están expuestos en su mayoría.					Muy poco agua en el cauce. Presencia mayoritaria de pozones quietos.					
PUNTUACIÓN	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
6.ALTERACIONES ANTRÓPICAS DEL CAUCE	Obras (p/ej. canalización y/o defensas) ausentes o mínimas. Cursos con un desarrollo normal. Sin tomas de agua o las que existen no interfieren con el normal escurrimiento del curso					Algunas obras, usualmente en áreas de puentes y/o evidencias de obras pasadas.					Obras extensivas con estructuras en ambas orillas; de un 40 a 80% del tramo modificado.					Orillas cementadas. Más del 80% del tramo del curso de agua modificado. Los hábitats están muy alterados o han sido totalmente eliminados.					
PUNTUACIÓN	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

7.FRECUENCIA DE RÁPIDOS PUNTUACIÓN	Rápidos relativamente frecuentes, distancia entre rápidos/ancho del arroyo<7:1 (generalmente 5-7). La variedad de hábitats es la clave. En arroyos donde los rápidos son continuos es importante la ubicación de los bloques u otros obstáculos grandes.	Rápidos poco frecuentes, distancia entre rápidos/ancho del arroyo entre 7 y 15.	Rápidos o curvas ocasionales. Los contornos del fondo proveen algunos hábitats. Distancia entre rápidos/ancho del arroyo entre 15 y 25.	Generalmente todo es somero o los rápidos son someros. Hábitat pobre. Distancia entre rápidos/ancho del arroyo mayor a 25																
	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
8.ESTABILIDAD DE LAS RIBERAS (determinar derecha o izquierda mirando aguas abajo) PUNTUACIÓN	Orillas estables. Ninguna o mínima evidencia de erosión. Poca probabilidad de futuros problemas de erosión (<5% de las orillas afectadas)	Orillas moderadamente estables Pequeñas áreas de erosión. . 5-30% de las riberas con áreas de erosión.	Orillas moderadamente inestables. 30-60% de las riberas con áreas de erosión. Alta erosión potencial durante las crecidas.	Orillas inestables. Muchas áreas erodadas. 60-100% de las riberas tienen marcas de erosión.																
	Orilla derecha	10	9	8	7	6	5	4	3	2	1	0								
	Orilla izquierda	10	9	8	7	6	5	4	3	2	1	0								
9.PROTECCIÓN VEGETAL PUNTUACIÓN	Más del 90% de las riberas y de la zona riparia inmediata cubierta por vegetación nativa, incluyendo árboles, arbustos o macrófitas, no modificada o poco modificada (pastoreo, movimientos mínimos)	70-90% de las riberas cubiertas por vegetación nativa, pero una clase de plantas no está bien representada. Alguna disrupción evidente que no afecta completamente el crecimiento potencial de las plantas.	50-70% de las riberas cubiertas por vegetación. Disrupción obvia. Parches de suelo desnudo. Menos de la mitad del crecimiento potencial de las plantas se mantiene.	Menos del 50% de las riberas cubiertas por vegetación. La vegetación ribereña está muy modificada e incluso puede faltar.																
	Orilla derecha	10	9	8	7	6	5	4	3	2	1	0								
	Orilla izquierda	10	9	8	7	6	5	4	3	2	1	0								
10.ALTERACIÓN DE LA ZONA CON VEGETACIÓN RIPARIA PUNTUACIÓN	Las actividades humanas no han impactado la zona	Las actividades humanas han impactado mínimamente en esta zona.	Las actividades humanas han impactado en gran medida en esta zona.	Debido a las actividades humanas la zona riparia es pequeña o no existe.																
	Orilla derecha	10	9	8	7	6	5	4	3	2	1	0								
	Orilla izquierda	10	9	8	7	6	5	4	3	2	1	0								

A3. Recursos Web

<http://adn.biol.umontreal.ca/~numericalecology/Rcode/>

<http://ordination.okstate.edu/>

<https://labscn-unalmed.github.io/ecologia-numerica/guiones/resumen.nb.html>

https://wiki.qcbs.ca/r_workshop10

<https://www.davidzeleny.net/anadat-r/doku.php/en:start>

https://github.com/ajsmit/Quantitative_Ecology/blob/master/Num_Ecol_R_book_ed1/chap6.R

<https://github.com/vegandevs/vegan>

<https://r-coder.com/plot-en-r/>

https://www.youtube.com/watch?v=ffCM2RHccqk&ab_channel=Fran%C3%A7ois_Husson

A4. Sentencias de R

A continuación, se presentan las sentencias específicas en R, asociadas a los métodos AFM, RDA y ACC.

```
#===== LIBRERIAS =====#
library(vegan)
library(FactoMineR)

#==ANALISIS DE DATOS =====#

# 1. MONITOREO LHH

lhht <- read.table("lhhepa.csv",sep=";",dec="," ,row.names=1,header=TRUE)
lhh.frec <- lhht[,21:66] # matriz Y (composición de la comunidad bentónica)
lhh.frec.hel <- decostand(lhh.frec, method="hellinger")
lhh.cuantit <- lhht[,2:20] # matriz X
lhh.cuanti <- lhh.cuantit[,c("CE", "OD", "PH", "TAG", "ALT", "NT", "PRS", "SST",
,"FE", "SI", "DU", "ALC", "PEH" )]
# Estandarización con el uso de la función decostand () en vegan:
lhh.cuanti.z <- decostand(lhh.cuanti, method="standardize")
# Proporción de ceros:
```

```
sum(lhh.frec==0)/(nrow(lhh.frec)*ncol(lhh.frec))
```

AFM LHH

```
library(FactoMineR) #AFM
lhh<-cbind(lhh.cuanti,lhh.frec.hel)
lhh.AFM<-MFA(lhh,group=c(13,46),type=c("s","f"),
ncp=2,name.group=c("Cuanti.lhh","Frec.hel.lhh"),graph=TRUE)
summary(lhh.AFM) #Resumen de resultados
lhh.AFM$inertia.ratio # Inercia de cada uno de los ejes
round(lhh.AFM$group$Lg,digits = 3) # Coeficiente Lg
round(lhh.AFM$group$RV, digits=3) # Coeficiente RV
lhh.AFM$group$contrib #Contribucion de los grupos a la construcción del eje
lhh.AFM$group$cos2 #Representacion de la calidad de cada grupo en el
subespacio (cos2 entre el punto y su proyección)
dimdesc(lhh.AFM,axes = 1:2, proba=0.0001) #eleccion de var mas
características con cada eje

# @@Gráficos AFM LHH@@
plot(lhh.AFM)
plot(lhh.AFM,choix="ind",partial="all",habillage="group",cex=1.2) # Grafico de
Individuos con puntos parciales (Individual factor map)
plot(lhh.AFM,choix="ind",partial="all",habillage="group",cex=1.2, select = "cos2
0.22") #Elijo los sitios con cos2 > 0,3
plot(lhh.AFM,choix="var",habillage="group",cex=0.8, shadow=TRUE, select =
"contrib 6") #Circulo de correlación con 6 var mas contributivas
plot(lhh.AFM,choix="var",habillage="group",cex=1.2, shadow=TRUE, label.pos =
-2)
plot(lhh.AFM,invisible="ind",choix="freq",habillage="group",cex=0.8) #Grafico de
familias
plot(lhh.AFM,choix="freq",habillage="group",cex=0.8) #Factor Map
#Grafico Representación conjunta AFM LHH
Eje1<-lhh.AFM$ind$coord[,1]
Eje2<-lhh.AFM$ind$coord[,2]
SITIOS <- rownames(lhh.AFM$ind$coord) # etiqueta individuos
Eje3<-lhh.AFM$freq$coord[,1]
Eje4<-lhh.AFM$freq$coord[,2]
MACROINVERTEBRADOS<-rownames(lhh.AFM$freq$coord) # etiqueta
familias
COLOR<-ifelse(lhht$scuencia=="TQ",'green4',
ifelse(lhht$scuencia=="L",'blue','darkmagenta'))#
B2<-
read.table("COLORBMPS2.csv",row.names=1,sep=";",dec="," ,header=TRUE,na.
strings = "-1") #para poner etiquetas sitios en blanco
BM2<-B2[,1]
```

```

COLORBMPS2<-as.character(BM2)
P<-
read.table("POSLHH.csv",row.names=1,sep=";",dec=".",header=TRUE,na.strings
="-1") #para poner etiquetas sitios en blanco
PP<-P[,1]
POSLHH<-as.character(PP) #para darles distintas posiciones a las etiquetas
plot(Eje1,Eje2, xlim=c(-2,3.8), ylim=c(-1.5,1.4), type="n",xlab='Dim 1(25.74%)',
ylab='Dim 2 (9.77%)',pch=19)
abline(h=0,lty=3) # linea en x=0
abline(v=0,lty=3) # linea en y=0
points(Eje3,Eje4, pch=16, col=COLORBMPS2,cex=1)
text(Eje3, Eje4, MACROINVERTEBRADOS, cex=1.1,col="black", pos =
POSLHH)
text(Eje1, Eje2, SITIOS, cex=1.2,col = COLOR, font=2 )

# RDA o ACC? LHH
library(vegan)
lhh.DCA <- decorana (log1p (lhh.frec))

# RDA LHH
# RDA de datos de familias restringido a 13 variables ambientales
lhh.RDA <- rda(lhh.frec.hel~., data=lhh.cuanti.z)
summary(lhh.RDA) # Por defecto: scaling=2, scale=FALSE
# Test Global del RDA:
anova.cca(lhh.RDA, step=999) # RDA modelo significativo (p=0.001)
# Test de los ejes canónicos:
anova.cca(lhh.RDA, step=999, by="axis")
# R2 - Relacion entre X e Y:
R2 <- RsquareAdj(lhh.RDA)$r.squared
R2adj <- RsquareAdj(lhh.RDA)$adj.r.squared
# coeficientes canonicos (=coef de regresion para cada var en cada eje
canonico)
coef(lhh.RDA)
# Estadisticos diagnosticos:
goodness(lhh.RDA)
fam.good <- goodness(lhh.RDA)
fam.good
sel.fam <- which(fam.good[,1]>=0.15 | fam.good[,2]>=0.15) # 17 familias mas
importantes
sel.fam
inertcomp(lhh.RDA)# ajuste de especies
spenvcor(lhh.RDA)
vif.cca(lhh.RDA)
intersetcor(lhh.RDA) # correlaciones de las var amb con los ejes

```

```

sit.good.lhh <- goodness(lhh.RDA, display = "sites") # Bondad del ajuste de los
sitios
sel.sit.lhh <- which(sit.good.lhh[,1]>=0.25 | sit.good.lhh[,2]>=0.25) # sitios mejor
representados
sel.sit.lhh
# Elección del mejor modelo reducido (ordiR2step - Utiliza el criterio de AIC):
ordi.sel.lhh <- ordiR2step(rda(lhh.frec.hel~1., data=lhh.cuanti.z), scope=
formula(lhh.RDA), direction= "forward", R2scope=TRUE, pstep=1000)
ordi.sel.lhh$call

# @@ GRAFICOS RDA LHH @@
COLORL<-ifelse(lhht$cuenca=="TQ",'forestgreen',
              ifelse(lhht$cuenca=="L",'deepskyblue2','deeppink2'))
#BIPLOTS LHH
# 2 panels
layout(matrix(1:2,1,2))
#1# Plot Sitios
#RDA BIPILOT - sitio x amb - Escalado 1 without species (using lc site scores)
plot(lhh.RDA, display=c("lc", "cn"), type="n", main="Sitios", scaling=1)
text(lhh.RDA, display="lc", col=COLOR, scaling=1) #sitios
#points(lhh.RDA, display = "lc", choices = c(1, 2),scaling = 2, axis.bp = FALSE,
correlation = FALSE, hill = FALSE,pch=21, col=1, bg=COLOR)
text(lhh.RDA, display="cn", col="red") #var ambientales
#2# Plot Familias
#RDA BIPILOT - fam x amb - Escalado 2 without sites
plot(lhh.RDA, display=c("species", "bp"), type="n", ylab="", main="Familias",
scaling=2)
text(lhh.RDA, display="species",arrow=TRUE, col="blue", scaling=2, select =
sel.fam)
spe.sc <- scores(lhh.RDA, choices=1:2, scaling=2, display="sp")
arrows(0,0,spe.sc[,1]*0.92, spe.sc[,2]*0.92, length=0.05, lty=1, col='blue')
text(lhh.RDA, display="bp", col="red")

# TRIPILOT LHH
#3#
plot(lhh.RDA, type="n", scaling=2, xlim=c(-0.8,0.8), ylim=c(-0.8,0.8))
text(lhh.RDA,dis="cn", scaling=2,col="red")
points(lhh.RDA,dis="lc", pch=21, col=COLORL, bg=COLORL, cex=1.2,
scaling=2) #Sitios default wa
text(lhh.RDA, dis="sp", col="blue", cex=0.8, scaling=2, correlation=TRUE, select
= sel.fam) #solo familias representativas
spe.sc <- scores(lhh.RDA, choices=1:2, scaling=2, display="sp")
arrows(0,0,spe.sc[,1]*0.92, spe.sc[,2]*0.92, length=0.05, lty=1, col='blue')

```



```

Eje3<-didy.AFM$freq$coord[,1]
Eje4<-didy.AFM$freq$coord[,2]
ALGAS <- rownames(didy.AFM$freq$coord) # etiqueta algas

COLORD<-ifelse(didyt$Cuenca=="RN",'goldenrod',
              ifelse(didyt$Cuenca=="RL",'dodgerblue',
                    ifelse(didyt$Cuenca=="RC",'darkmagenta','forestgreen'))))

#para cambiar la posicion de las etiquetas:
P2 <-
read.table("POSDISIT.csv",row.names=1,sep=";",dec=";",header=TRUE,na.string
s = "-1") #para poner etiquetas sitios en blanco
PO<-P2[,3]
POSDISIT<-as.character(PO) #posicion etiquetas sitios
P3 <-
read.table("POSDISP.csv",row.names=1,sep=";",dec=";",header=TRUE,na.string
s = "-1") #para poner etiquetas especies en blanco
POS<-P3[,1]
POSDISP<-as.character(POS) # posicion etiquetas especies

plot(Eje1,Eje2, xlim=c(-1.9,3.2), ylim=c(-2.3,2.5), type="n",xlab='Dim 1(18.66%)',
ylab='Dim 2 (9.17%)')
abline(h=0,lty=3) # linea en x=0
abline(v=0,lty=3) # linea en y=0
points(Eje1,Eje2, pch=16, col=COLORD,cex=0.5)
text(Eje1, Eje2, SITIOSD, cex=0.8,col = COLORD,pos=POSDISIT,font=2 )
points(Eje3,Eje4, pch=16, col="black",cex=0.5)
text(Eje3, Eje4, ALGAS, cex=0.7,col="black", pos=POSDISP)
## RDA o ACC DIDYMO?
didy.DCA <- decorana (log1p (didy.frec))

# ACC DIDYMO
# no puede haber datos faltantes
didy.ACC <- cca(didy.frec~., data=didy.cuanti)
didy.ACC
summary(didy.ACC)
summary(didy.ACC, scaling = "species", axes = 6,
        display = c("sp", "wa", "lc", "bp", "cn"),
        digits = max(3, getOption("digits") - 3),
        correlation = FALSE, hill = FALSE)
constrained_eig <- didy.ACC$CCA$eig/didy.ACC$CA$tot.chi*100
unconstrained_eig <- didy.ACC$CA$eig/didy.ACC$CA$tot.chi*100
barplot (c(constrained_eig, unconstrained_eig), col = c(rep ('red',
length (constrained_eig)), rep ('black', length (unconstrained_eig))),

```

```

    las = 2, ylab = '% variation')
#Test de permutacion del analisis global:
anova.cca(didy.ACC, step=999) # Significativo p=0.002
#Test de permutacion de cada eje:
anova.cca(didy.ACC, by="axis", step=999)
# Unadjusted and adjusted R^2 - like statistics
RsquareAdj(didy.ACC)
# Selección del mejor modelo
ACC.ordi.sel.didy <- ordistep(cca(didy.frec~1, data = didy.cuanti), scope =
formula(didy.ACC),direction = "forward",pstep=1000)
ACC.ordi.sel.didy$call
# Estadísticos diagnósticos
esp.good <- goodness(didy.ACC) # Bondad del ajuste# - eleccion de especies
mas representativas
esp.good
sel.esp <- which(esp.good[,1]>=0.05 | esp.good[,2]>=0.05) # 25 especies
sel.esp
select.spe=sel.esp # podria usarse para graficar
sel.esp15 <- which(esp.good[,1]>=0.1 | esp.good[,2]>=0.1) # 15 especies mas
representativas en los dos ejes
sel.esp15
inertcomp(didy.ACC) # ajuste de especies
spenvcor(didy.ACC)
tolerance(didy.ACC)
sit.good <- goodness(didy.ACC, display = "sites") # Bondad del ajuste de los
sitios
sel.sit <- which(sit.good[,1]>=0.15 | sit.good[,2]>=0.15) # 16 sitios mejor
representados
sel.sit
intersetcor(didy.ACC) # correlaciones de las var amb con los ejes
vif.cca(didy.ACC) # VIF de las var ambientales

# @@ Gráficos ACC DIDYMO @@ = )
COLORD1<-ifelse(didy.Cuenca=="LNH", "blueviolet",
               ifelse(didy.Cuenca=="RN", 'goldenrod2',
                     ifelse(didy.Cuenca=="RL", 'dodgerblue',
                           ifelse(didy.Cuenca=="RC", 'deeppink2', 'forestgreen'))))# Ind x
cuenca
# BIPLOTS DIDYMO
# 2 panels
layout(matrix(1:2, 1,2))
#1# Plot Sitios
#BIPLOT - sitio x amb - Escalado 2 without species (using lc site scores)
plot(didy.ACC, display=c("lc", "cn"), type="n", main="Sitios", scaling=1)

```

