



**Identificación de patrones temporales asociados al desembarque de especies altamente migratorias y su relación con variables ambientales en Chile a través de herramientas de ciencias de datos.**

**Ljubitzta Alejandra Clavijo Gorostiaga**

Marzo de 2025

Profesor Guía

Dr. Francisco Plaza Vega

Tesis para optar al

grado académico de Magíster en Estadística

# Índice general

1	Introducción .....	1
2	Antecedentes .....	4
3	Metodología .....	6
3.1	Descripción del conjunto de datos.....	6
3.2	Esquema Metodológico.....	7
3.3	Análisis descriptivo .....	8
3.4	Análisis de Series Temporales y Autocorrelación .....	10
3.5	Análisis de Agrupamiento.....	11
3.6	Implementación de modelos .....	13
4	Resultados .....	16
4.1	Esquema metodológico .....	16
4.2	Análisis descriptivo .....	17
4.2.1	Datos pesqueros .....	17
4.2.2	Datos ambientales.....	22
4.3	Análisis de Componentes Principales.....	27
4.4	Análisis de Series de Tiempo y Autocorrelación .....	29
4.5	Análisis de Agrupamiento.....	36
4.6	Implementación de modelo para desembarque de pez espada, tiburones y dorado de altura. ...	53
5	Discusión .....	60
6	Conclusiones.....	66
7	Bibliografía.....	68

## Abreviaciones

ACF:	Función de Autocorrelación
APC:	Análisis de Componentes Principales
Anom N12:	Anomalía del Índice Climático en la región Niño 1+2
Anom N34:	Anomalía del Índice Climático en la región Niño 3.4
CPUE:	Captura por unida de esfuerzo
DANC:	Desembarques de anchoveta ( <i>Engraulis ringens</i> )
DCD:	Descubrimiento del Conocimiento en los Datos
DDORA:	Desembarques de dorado de altura ( <i>Coryphaena hippurus</i> )
DSAR:	Desembarques de sardina ( <i>Strangomera bentincki</i> )
DTIB:	Desembarques de tiburón azul y marrajo ( <i>Prionace glauca</i> y <i>Isurus oxyrinchus</i> )
ENSO:	Fenómeno El Niño-Oscilación del Sur
HCI:	Índice de Corriente de Humboldt
IOS:	Índice de Oscilación del Sur
MEI:	Índice Multivariado de El Niño/Oscilación del Sur
N12:	Índice Climático de la región Niño 1+2
N34:	Índice Climático de la región Niño 3.4
PACF:	Función de Autocorrelación Parcial
PC:	<i>Principal Components</i> , Componentes Principales
PCA:	<i>Principal Components Analysis</i> , Análisis de Componentes Principales
PDO:	<i>Pacific Decadal Oscillation</i> , Oscilación Decadal del Pacífico
SCH:	Sistema de la Corriente de Humboldt
TPI:	<i>Tripole Index for the Interdecadal Pacific Oscillation</i> , Índice Tripolar del Pacífico
TSM:	Temperatura Superficial del Mar

## Resumen

Existe una necesidad urgente de estudiar las capturas de especies altamente migratorias en Chile considerando variables ambientales, con el objetivo de desarrollar un enfoque inclusivo que utilice herramientas estadísticas avanzadas. En este contexto, los análisis multivariados y de series de tiempo, en conjunto con un enfoque de ciencia de datos, juegan un rol clave para identificar patrones subyacentes que expliquen la interacción entre los desembarques y las dinámicas ambientales.

El presente estudio se centra en especies de alta relevancia para las pesquerías del norte de Chile, como el pez espada (*Xiphias gladius*), el tiburón azul (*Prionace glauca*), el marrajo (*Isurus oxyrinchus*) y el dorado de altura (*Coryphaena hippurus*), analizando su relación con variables ambientales entre 1980 y 2022. Para ello, se utilizaron datos de desembarques mensuales y anuales, junto con información ambiental sobre la temperatura superficial del mar (TSM), fenómenos climáticos globales como El Niño-Oscilación del Sur (ENSO), la dinámica del Sistema de la Corriente de Humboldt (SCH) y diversos índices climáticos, como la Oscilación Decadal del Pacífico (PDO), entre otros.

Desde una perspectiva metodológica, se adoptó un enfoque basado en seis etapas: i) enmarcar el problema, ii) entendimiento de los datos, iii) extracción de características, iv) modelación y análisis, v) implementación práctica, y vi) comunicación efectiva de los resultados; siguiendo los lineamientos propuestos por Cady, 2017; Chapman et al., 2000; Fayyad et al., 1996; Plaza et al., 2017). Se emplearon herramientas específicas de programación en R para realizar análisis exploratorios, agrupamiento, análisis de componentes principales y análisis de series de tiempo.

Los resultados revelaron fluctuaciones significativas en los desembarques, influenciadas por eventos climáticos como El Niño 1982-1983, 1997-1998 y 2015-2016, los cuales podrían haber alterado los patrones de captura de las especies de interés. Para analizar estas variaciones, se aplicó Análisis de Componentes Principales (PCA) con el objetivo de reducir la dimensionalidad de los datos. Las primeras cinco componentes explicaron el 73 % de la variabilidad, permitiendo identificar las principales tendencias en la serie temporal. Adicionalmente, se implementó un análisis de agrupamiento mediante *k-means clustering*, el cual permitió segmentar la serie de tiempo en tres grupos, evidenciando cuatro periodos diferenciados: 1980-1983, 1983-1996, 1996-2008 y 2008-2022. Se observó que las condiciones al inicio del período analizado presentaban similitudes con las del tramo final, sugiriendo un patrón cíclico en la interacción entre los desembarques y las variables ambientales.

Asimismo, se identificaron relaciones clave entre los índices climáticos y las capturas, destacando la influencia del Índice Multivariado de El Niño (MEI) y el Índice Niño 3.4 (N34) en la variabilidad pesquera, así como la correlación entre el desembarque de tiburones y el Índice de Corriente de Humboldt (HCI), y entre la Temperatura Superficial del Mar (TSM) con el dorado de altura y el tiburón marrajo. Además, se evidenciaron cambios sostenidos en las capturas de tiburón azul y marrajo a partir del año 2000, mientras que el dorado de altura mostró un patrón estacional vinculado a los picos de TSM.

Para evaluar estas relaciones, se implementaron modelos de aprendizaje automático, específicamente Random Forest y XGBoost, con el fin de estimar la influencia de las variables ambientales en la predicción de los desembarques. Estos modelos permitieron identificar la

importancia relativa de cada variable climática en la dinámica pesquera, proporcionando información clave para la interpretación de los patrones observados.

Este enfoque metodológico demuestra el potencial de la estadística y la ciencia de datos para abordar problemas complejos en el manejo pesquero, al integrar análisis multivariados, series de tiempo y modelos. Al incorporar factores ambientales a macro escala, incluyendo tanto al SCH como el fenómeno de El Niño, el estudio no solo proporciona herramientas para comprender los cambios en los ecosistemas marinos relacionados a índices climáticos y a la temperatura del mar, sino que también enfatiza la importancia de optimizar estos análisis para una mejor comprensión de la dinámica de los desembarques.

## 1 Introducción

---

La minería de datos se ha convertido en una herramienta esencial para la extracción de información útil y patrones subyacentes en grandes conjuntos de datos, integrando algoritmos y técnicas que permiten realizar predicciones basadas en datos históricos (Witten et al., 2011; Silva et al., 2015). Estas metodologías, ampliamente aplicadas en disciplinas como la economía, la salud y las ciencias naturales (Liao et al., 2012), han mostrado un gran potencial en el ámbito de las pesquerías, al ser combinadas con herramientas estadísticas avanzadas, especialmente para modelar la relación espacio-temporal entre factores ambientales y la distribución de recursos marinos (Bedriñana et al., 2022; Cope y Punt, 2009; Naranjo et al., 2015; Su et al., 2004), así como la identificación de patrones ecosistémicos en sistemas marinos, destacando su aplicabilidad para optimizar la gestión de recursos pesqueros (Enomoto et al., 2011; Huettmann et al., 2011).

En Chile, investigaciones previas han demostrado la utilidad de estas metodologías para analizar series de tiempo de desembarques pesqueros y su relación con variables ambientales, permitiendo identificar patrones de distribución asociados a eventos climáticos como El Niño y a dinámicas ecosistémicas propias del Sistema de la Corriente de Humboldt (Espino, 2013; Naranjo et al., 2015; Plaza Vega, 2017; Plaza et al., 2023; Silva et al. 2015; Yáñez et al., 2012; Yáñez et al., 2018). En este sentido, los trabajos de Ocaña-Riola (2017) y Contreras y Molina-Portillo (2019) refuerzan el papel fundamental de la estadística como ciencia interdisciplinaria, capaz de analizar la variabilidad, modelar procesos aleatorios y sustentar decisiones informadas en diversos campos.

Las especies altamente migratorias, como el pez espada (*Xiphias gladius*), los tiburones marrajo (*Isurus oxyrinchus*) y azul (*Prionace glauca*), y el dorado de altura (*Coryphaena hippurus*), representan una relevancia económica y social para las pesquerías del norte de Chile (Barría et al., 2020; Bustamante et al., 2023). Clasificadas bajo el Anexo I de la Convención de las Naciones Unidas sobre el Derecho del Mar (CMS, 2007), estas especies tienen la capacidad de recorrer largas distancias, distribuyéndose tanto en zonas económicas exclusivas como en alta mar. Por lo que su relevancia no solo radica en su valor ecológico, sino también en su importancia económica y cultural para las comunidades costeras.

El Memorándum de Entendimiento sobre la Conservación de Tiburones Migratorios, desarrollado por la Convención sobre la Conservación de las Especies Migratorias de Animales Silvestres (CMS, 2007), busca alcanzar un estado de conservación favorable para estas especies, promoviendo alternativas sostenibles para su explotación y reconociendo su relevancia cultural y económica. No obstante, las poblaciones de tiburones enfrentan un alto riesgo de sobrepesca debido a características como lento crecimiento, bajo potencial reproductivo y limitada resiliencia (Lamilla et al., 2012; Bustamante et al., 2023). En Chile, estas especies suelen capturarse como fauna acompañante del pez espada y la flota de la pesquería del dorado de altura se alterna con la pesquería del tiburón, dependiendo de la disponibilidad del recurso, oferta y demanda.

La regulación pesquera en Chile se rige por la Ley General de Pesca y Acuicultura (LGPA, Ley N° 18.892, 1989), cuyo objetivo principal es la conservación y uso sustentable de los recursos hidrobiológicos mediante un enfoque precautorio y ecosistémico. La normativa establece que el

uso sustentable implica garantizar beneficios sociales y económicos sin comprometer las oportunidades de las generaciones futuras. La Ley 20.657 (2013) refuerza estos principios, destacando la necesidad de objetivos de largo plazo, la aplicación del principio precautorio y un enfoque ecosistémico para la administración sostenible de los recursos pesqueros.

En Chile, las migraciones de estas especies tienen una marcada estacionalidad influenciada por variables como la temperatura superficial del mar (TSM) y fenómenos climáticos globales, como El Niño-Oscilación del Sur (ENSO). El pez espada migra de sur a norte en Chile según las estaciones, mientras que los tiburones se concentran principalmente en el norte del país debido a su preferencia por aguas templadas, y por su parte, el dorado de altura es capturado exclusivamente en las cuatro primeras regiones del país durante la primavera y el verano (Barria et al. 2019 y Zárate et al., 2021). Estas dinámicas destacan la importancia de identificar patrones de interacción entre las condiciones ambientales y los desembarques pesqueros para un manejo más efectivo.

En este marco, el Sistema de la Corriente de Humboldt (SCH), reconocido como una de las surgencias costeras más productivas del mundo, es un elemento clave para la dinámica de los ecosistemas marinos de Perú y Chile, sustentando una amplia variedad de especies, incluidas aquellas de interés en este estudio. Sin embargo, se prevé que los cambios climáticos globales afecten su productividad, con potenciales impactos en las pesquerías asociadas (Bakun & Weeks, 2008; Yáñez et al 2012; Yáñez et al 2016; Silva et al 2015). Este escenario resalta la necesidad de comprender cómo interactúan las condiciones ambientales y los desembarques pesqueros para adaptar las estrategias de manejo frente a estos cambios.

En este contexto, las herramientas de ciencia de datos, como el análisis de series de tiempo y la minería de datos, han demostrado ser valiosas para analizar grandes volúmenes de información y extraer conocimiento relevante (Naranjo et al., 2015; Plaza et al., 2017; Witten et al., 2011; Yáñez et al., 2018). En pesquerías, estas técnicas han permitido identificar relaciones significativas entre variables ambientales y recursos pesqueros, contribuyendo al desarrollo de estrategias de manejo sostenible (Naranjo et al 2015; Silva et al 2015; Su et al., 2004; Yang et al., 2008; Yáñez et al 2012). Integrar estos enfoques en el análisis de especies altamente migratorias podría proporcionar información clave para enfrentar los desafíos actuales en la gestión de estos recursos.

El presente estudio tiene como objetivo explorar la interacción entre las pesquerías de recursos altamente migratorios y las condiciones ambientales en el norte de Chile, con un enfoque en el pez espada, los tiburones y el dorado de altura. A través de herramientas de ciencia de datos, se pretende identificar patrones en estas relaciones que permitan comprender mejor las dinámicas de las especies, generando información de utilidad para contribuir al desarrollo de políticas de manejo sostenible e informado.

## **1. Objetivo general**

Identificar patrones de interacción entre el ambiente y los desembarques en series de tiempo de las pesquerías de especies altamente migratorias en Chile entre 1980 al 2022, a través de herramientas de ciencia de datos.

### **1.1.1 Objetivos específicos**

1. Desarrollar un análisis exploratorio de las series de tiempo para las variables ambientales y pesqueras.
2. Identificar herramientas de ciencia de datos para el análisis de las series de tiempo ambientales y pesqueras.
3. Implementar procedimiento de análisis para la identificación de patrones en la interacción de las pesquerías de recursos altamente migratorios con variables ambientales.

## **2. Hipótesis**

La implementación de un procedimiento que involucra herramientas de ciencia de datos permitirá identificar patrones de interacción entre el ambiente y el desembarque de especies altamente migratorias en Chile.

### 2 Antecedentes

---

El Sistema de la Corriente de Humboldt (SCH) es ampliamente reconocido como uno de los ecosistemas marinos más productivos del mundo (Daneri et al. 2000; Montecino & Lange, 2009; Weidberg et al 2021; Zarate et al 2020). Este sistema de corriente fría y de baja salinidad fluye hacia el norte a lo largo de la costa occidental de América del Sur, desde el sur de Chile hasta el norte de Perú, y se caracteriza por un intenso proceso de surgencia que enriquece las aguas superficiales con nutrientes, fomentando una alta productividad biológica. Este ecosistema influye significativamente en la distribución y abundancia de especies marinas debido a su interacción con factores ambientales locales y remotos, como los vientos costeros, la temperatura superficial del mar (TSM) y los eventos climáticos globales, particularmente El Niño-Oscilación del Sur (ENSO) (Hormazábal et al., 2001; Hormazábal et al., 2002). Durante El Niño, la reducción de los vientos alisios debilita las surgencias, elevando la TSM y disminuyendo la productividad, mientras que La Niña genera el efecto contrario, intensificando las surgencias y favoreciendo la productividad (Bakun & Weeks, 2008). Estas variaciones impactan directamente en las dinámicas tróficas del SCH, afectando desde la productividad primaria, el zooplancton, hasta la biomasa de especies clave como la anchoveta (*Engraulis ringens*) y la sardina (*Strangomera bentincki*), que constituyen la base alimenticia de depredadores de pelágicos como el pez espada y los tiburones (Gatica, 2009; Yáñez et al., 2016).

Estudios recientes han destacado que la interacción entre condiciones ambientales como la TSM, la concentración de clorofila-a y los gradientes térmicos desempeña un rol clave en la distribución y abundancia de especies marinas. Trabajos como los de Bedriñana et al. (2021, 2022) han utilizado modelos avanzados, como los modelos *N-mixture* bayesianos, para identificar áreas críticas para especies como las ballenas azules en la Patagonia chilena y su superposición con actividades humanas como la pesca y el tráfico marítimo. Estos estudios resaltan el como la estadística ayuda a la comprensión del medio marino y señala como estas metodologías son directamente aplicables al análisis de especies altamente migratorias en el SCH, como es el caso del pez espada, el tiburón azul y el dorado de altura.

La minería de datos, definida como un proceso de descubrimiento de patrones válidos y novedosos en grandes volúmenes de datos, permite identificar relaciones complejas mediante técnicas como análisis de clústeres, redes neuronales y modelos predictivos (Han et al., 2011; Fayyad et al., 1996; Plaza et al., 2017). En el contexto pesquero, estas herramientas han demostrado ser efectivas para modelar la respuesta de los recursos frente a la variabilidad ambiental, facilitando la toma de decisiones en un marco de manejo sostenible (Cios, 2007; Naranjo et al 2015; Plaza et al., 2017; Yáñez et al 2012; Silva et al 2015). Por ejemplo, el uso de redes neuronales ha permitido identificar patrones no lineales en los desembarques, destacando su utilidad frente a escenarios de cambio climático (Yáñez et al., 2017).

La influencia de ENSO y otros eventos climáticos globales también ha sido estudiada mediante herramientas estadísticas como el análisis de componentes principales (PCA) y los modelos aditivos generalizados (GAM). Estas técnicas han permitido identificar patrones temporales en las capturas y su relación con variables como la TSM, los índices de surgencia y las oscilaciones climáticas (Plaza et al., 2017; Zárate et al., 2021). Por ejemplo, el índice de El Niño 1+2 ha

demostrado ser un predictor eficaz para modelar la distribución espacial y la abundancia relativa del pez espada en el Pacífico suroriental (Espíndola et al., 2011; Nan Jay Su et al., 2020).

El enfoque ecosistémico en las pesquerías del SCH ha integrado análisis biológicos, pesqueros y oceanográficos, destacando asociaciones significativas entre la variabilidad ambiental y las capturas de tiburón azul y marrajo. Este enfoque, respaldado por informes como el de Pesquerías de Recursos Altamente Migratorios (Zarate et al., 2020; Zarate et al., 2021; Zarate et al., 2022; Zarate et al., 2023), enfatiza la necesidad de incorporar modelos predictivos en la gestión sostenible de estas especies (Zárate et al., 2021; Zarate et al., 2022).

## 3 Metodología

### 3.1 Descripción del conjunto de datos

El presente estudio se centra en el análisis de una base de datos, cuya recopilación abarca el período desde enero de 1980 hasta diciembre de 2022 y se centra en la zona norte, que va desde la región de Arica y Parinacota hasta la región de Atacama. La información proviene de los anuarios estadísticos del Servicio Nacional de Pesca y Acuicultura (SERNAPESCA).

La base de datos incluye información pesquera y ambiental (**Tabla 1**). La información pesquera abarca los desembarques mensuales de 6 especies, entre las cuales se encuentran 4 especies altamente migratorias: tiburón azulejo (*Prionace glauca*), tiburón marrajo (*Isurus oxyrinchus*), pez espada (*Xiphias gladius*), dorado de altura (*Coryphaena hippurus*). Además, se consideran dos especies que forman parte de la dieta de las especies mencionadas: anchoveta (*Engraulis ringens*) y sardina española (*Sardinops sagax*).

Para los datos ambientales, se consideran 16 variables ambientales, entre ellas 8 variables provienen del repositorio de la Administración Nacional de Océanos y Atmósfera (NOAA) para el periodo mensual comprendido entre enero de 1980 y diciembre de 2022. Las variables incluyen el Índice Climático de la Oscilación Decadal del Pacífico (PDO), el Índice de Oscilación del Sur (IOS), el Índice Multivariado Bimestral de El Niño/Oscilación del Sur (MEI), el Índice Climático de la región Niño 1+2 (N12), la Anomalía del Índice Climático en la región Niño 1+2 (Anom N12), el Índice Climático de la región Niño 3.4 (N34), la Anomalía del Índice Climático en la región Niño 3.4 (Anom N34), y el Índice Tripolar de la Oscilación Interdecadal del Pacífico (TPI).

También como variable ambiental se utilizó la temperatura superficial del mar (TSM), expresada en grados Celsius (°C), fue obtenida a partir de los registros de la Dirección Meteorológica de Chile. Además, los datos del Índice de Corriente de Humboldt (HCL) para el periodo de enero de 1980 a diciembre de 2022 se obtuvieron de Bluewater ([www.bluewater.cl/HCI/index.htm](http://www.bluewater.cl/HCI/index.htm)).

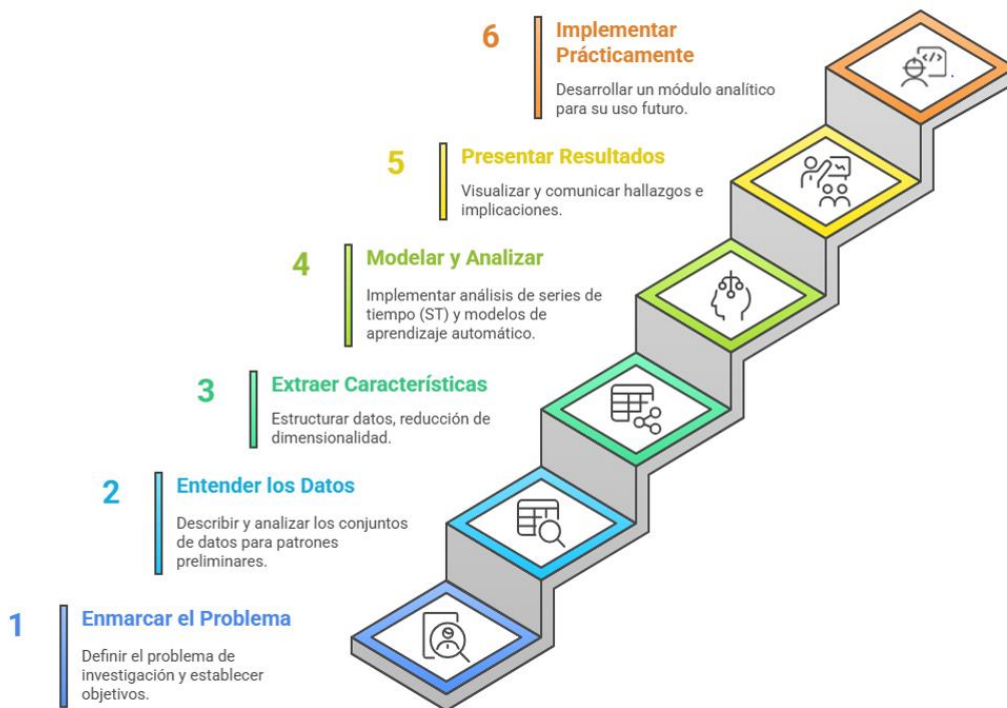
**Tabla 1.** Variables pesqueras y ambientales

Variables	Descripción	Unidad
<b>Pesqueros</b>		
DESPA	Desembarque de pez espada ( <i>Xiphias gladius</i> )	ton
DAZUL	Desembarque de tiburón azulejo ( <i>Prionace glauca</i> )	ton
DMARR	Desembarque de tiburón marrajo ( <i>Isurus oxyrinchus</i> )	ton
DTIB	Desembarque de tiburón azulejo y marrajo	
DDORA	Desembarque de dorado de altura ( <i>Coryphaena hippurus</i> )	ton
DANC	Desembarques anchoveta ( <i>Engraulis ringens</i> )	ton
DSAR	Desembarques sardina española ( <i>Sardinops sagax</i> )	ton
<b>Ambientales</b>		
PDO	Índice Climático de la Oscilación Decadal del Pacífico	°C
IOS	Índice Oscilación del Sur	mbar

HCI	Índice de Corriente de Humboldt (Isla de Pascua-Antofagasta)	
TSM	Temperatura superficial del mar de estación oceanográfica costera Antofagasta	°C
MEI	El índice bimestral multivariado de El Niño/Oscilación del Sur (ENSO)	°C
N12	Índice Climático de la región Niño 1+2	°C
Anomalia 1+2	Anomalía del Índice Climático en la región Niño 1+2	°C
N34	Índice Climático de la región Niño 3.4	°C
Anomalia 3.4	Anomalía del Índice Climático en la región Niño 3.4	°C
TPI	Índice Tripolar de la Oscilación Inter Decadal del Pacífico	°C

### 3.2 Esquema Metodológico

Se utilizó el esquema metodológico propuesto por Cady (2017) como guía general (**Figura 1**), abordando el trabajo en seis etapas. El estudio se centra en identificar patrones en la interacción entre las pesquerías de especies altamente migratorias y las variables ambientales en el norte de Chile, aplicando herramientas de ciencia de datos para analizar series de tiempo y detectar variabilidad en los desembarques. Para ello, se trabajó con datos de desembarques y variables ambientales desde 1980 hasta 2022, realizando un análisis exploratorio mediante estadísticas descriptivas, visualización de series de tiempo y detección de patrones preliminares. Posteriormente, se estructuran los datos en formato tabular y se generan nuevas variables relevantes y se aplica Análisis de Componentes Principales (PCA) para reducir la dimensionalidad y captar las principales fuentes de variabilidad. En la fase de modelamiento, se implementaron análisis de series de tiempo, autocorrelación (ACF y PACF) y modelos de aprendizaje automático, incluyendo métodos de agrupamiento para identificar patrones temporales y algoritmos como Random Forest y *XGBoost* para evaluar la relación entre desembarques y variables ambientales. Los resultados fueron presentados a través de figuras y tablas, evidenciando la evolución de los desembarques y su relación con las variables ambientales, además de identificar periodos diferenciados en las pesquerías. Finalmente, se desarrolló un módulo analítico en R con código reproducible, donde se analizan los patrones identificados y los grupos detectados en relación con las variables pesqueras de interés y las condiciones ambientales.



**Figura 1.** Esquema modificado de metodológico propuesto por Cady (2017)

### 3.3 Análisis descriptivo

El análisis de series temporales se aplicó a cada grupo con el objetivo de identificar patrones y tendencias en los desembarques de especies. Esta metodología se centró en el cálculo de medias móviles (MA), una técnica fundamental para suavizar la variabilidad de los datos y destacar tendencias subyacentes. Se calcularon dos tipos principales de medias móviles: una de 12 meses, diseñada para resaltar patrones estacionales y ciclos anuales, y otra de 60 meses, utilizada para capturar tendencias a largo plazo y posibles cambios estructurales en los desembarques. Estas herramientas permiten identificar de forma efectiva la influencia de factores estacionales y dinámicas más amplias que afectan a la abundancia y captura de las especies.

$$MA = \frac{P_0 + P_1 + \dots + P_{n-1}}{N}$$

Donde  $P_1 + P_2 + \dots + P_{n-1}$  representan los valores del conjunto de datos en cada período.

La media móvil de 12 meses permite visualizar fluctuaciones de corto plazo y ciclos anuales, reflejando variaciones estacionales en los desembarques. Por otro lado, la media móvil de 60 meses proporciona una visión de los cambios prolongados que pueden estar relacionados con alteraciones en los ecosistemas o variaciones en el esfuerzo pesquero. Ambas medias móviles fueron representadas junto con la serie original, facilitando una comparación directa entre los datos brutos y las versiones suavizadas. Esta visualización conjunta resulta clave para comprender tanto la estacionalidad como las tendencias prolongadas, lo que ofrece una perspectiva más amplia del comportamiento de los grupos analizados.

Para analizar la estructura temporal, se generaron gráficos de líneas que superponen las series originales con sus respectivas medias móviles. Se utilizó el software R, junto con las bibliotecas

zoo para el manejo de series temporales, *ggplot2* para la creación de gráficos y *gridExtra* para la organización de paneles gráficos.

Además, se realizó un análisis de correlación de Spearman ( $\rho$ ) para explorar las relaciones entre las variables pesqueras y ambientales.

$$P = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Donde:

$d_i$  es la diferencia entre los rangos de cada par de observaciones ( $X_i$  y  $Y_i$ ).

$n$  es el número de observaciones.

Este método fue seleccionado debido a que es una herramienta fundamental en el análisis de datos multivariados, ya que permite de manera visual observar relaciones entre variables, a través de una matriz de correlación (Closas et al. 2013; Dillon & Goldstein, 1984), la matriz de correlación se entrega en una tabla que muestra los coeficientes de correlación entre múltiples las variables, facilitando la identificación de patrones y relaciones en los datos. Se siguió el criterio propuesto por Hinkle et al. (2003) para interpretar la magnitud de las correlaciones, destacando aquellas con valores superiores a 0,7 como de alta correlación positiva o negativa.

### **Análisis de componentes Principales (PCA)**

El análisis de componentes principales (PCA) es una técnica estadística utilizada para reducir la dimensionalidad de un conjunto de datos, conservando la mayor cantidad de varianza posible. Desarrollada por Pearson (1901), esta técnica se aplica a través de la matriz de correlaciones, proyectando variables en un nuevo espacio. El PCA transforma las variables originales en componentes principales ortogonales, facilitando el análisis y visualización de las relaciones entre variables.

Dado un conjunto de datos con  $n$  observaciones y  $p$  variables, las variables deben ser estandarizadas para garantizar que cada una contribuya equitativamente al análisis. La estandarización se realiza mediante:

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j}$$

Donde  $Z_{ij}$  es el dato estandarizado de la variable  $j$  en la observación  $i$

$X_{ij}$  es el valor original de la variable  $j$  en la observación  $i$

$\mu_j$  es la media de la variable  $j$ , definida como  $\mu_j = \sum_{i=1}^n X_{ij}$

$\sigma_j$  es la desviación estándar de la variable  $j$ , calculada como:

$$\sigma_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \mu_j)^2}$$

La estandarización garantiza que todas las variables tengan media cero y desviación estándar uno.

$$C = \frac{1}{n-1} Z^T Z$$

Los valores propios  $\lambda$  y los vectores propios  $v$  de la matriz de covarianza se obtienen resolviendo:

$$Cv = \lambda v$$

Donde:

$v_i$  es el vector propio asociado, que define la dirección del nuevo eje de la componente principal.

$\lambda_i$  representa la varianza explicada por la  $i$ -ésima componente principal

Las componentes principales se ordenan en función de sus valores propios de mayor a menor, seleccionando aquellas que explican una proporción significativa de la varianza total. La proporción de varianza explicada por la  $i$ -ésima componente se calcula como:

$$\text{Varianza explicada} = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j}$$

Se proyectarán los datos originales en el nuevo espacio definido por los componentes principales seleccionados. La proyección se calculará utilizando la fórmula:

$$Y = ZV$$

Donde  $Y$  es la matriz de datos transformados,  $Z$  es la matriz de datos estandarizados y  $V$  es la matriz de vectores propios de los componentes seleccionados.

Para la interpretación de resultados, los componentes principales obtenidos se analizarán para identificar patrones y relaciones significativas en los datos. Se generarán gráficos y visualizaciones, como gráficos de dispersión, a pares, entre otros.

El gráfico de codo se utiliza para seleccionar el número óptimo de componentes, basado en el punto donde se observa una disminución marcada en la varianza explicada.

Además, el índice de Kaiser-Guttman (Kaiser, 1960) se emplea para descartar componentes con valores propios menores a 1, lo que asegura que solo se retengan componentes significativos para el análisis.

### 3.4 Análisis de Series Temporales y Autocorrelación

Para el análisis de las series temporales, se aplicó el Análisis de Componentes Principales (PCA) con el objetivo de reducir la dimensionalidad y resaltar las principales fuentes de variabilidad en los datos pesqueros y ambientales durante el periodo 1980-2022. Las series temporales de estas componentes se modelaron y analizaron para identificar patrones de estacionalidad, tendencias y variabilidad de corto plazo.

Se aplicó la descomposición aditiva de series temporales es una técnica que permite separar una serie en componentes estructurales para analizar su comportamiento. Se basa en el supuesto de que la serie observada  $Y_t$  es la suma de tres componentes principales:

$$Y_t = T_t + S_t + R_t$$

donde:

$T_t$  es la tendencia, que representa la evolución a largo plazo de la serie.

$S_t$  es la estacionalidad, que captura los patrones repetitivos en intervalos fijos de tiempo.

$R_t$  es el residuo, que recoge la variabilidad no explicada por las otras dos componentes.

Se calcularon las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) con un desfase de hasta 36 meses (3 años) para cada componente principal. Este criterio se estableció considerando la variabilidad ambiental a escalas interanuales y la influencia de eventos climáticos de mediano plazo, en particular el fenómeno El Niño-Oscilación del Sur (ENSO). ENSO presenta variaciones en escalas temporales de 2 a 7 años, con impactos significativos en las condiciones oceanográficas y atmosféricas que afectan las dinámicas pesqueras. Un desfase de 36 meses permite capturar tanto los efectos directos de eventos de El Niño y La Niña como las posibles respuestas retardadas en los desembarques pesqueros debido a cambios en la disponibilidad de los recursos y en la estructura trófica del ecosistema marino (Chavez et al., 2003; Wang & Fiedler, 2006). El cálculo de la ACF se realizó utilizando la fórmula:

$$ACF(k) = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2}$$

Donde:

$X_t$  es el valor de la serie en el tiempo  $t$

$\bar{X}$  es la media de la serie

$k$  representa el desfase (lag)

$N$  es el número total de observaciones

La PACF se calculó eliminando los efectos de los desfases intermedios utilizando la ecuación recursiva de Yule-Walker:

$$PACF(k) = \phi_k$$

Donde:

$\phi_k$  es el coeficiente autorregresivo de orden  $k$

### 3.5 Análisis de Agrupamiento

Como herramienta de minería de datos, el análisis de clústeres puede emplearse de manera independiente para mejorar la percepción de la distribución de los datos, observar las características de cada clúster y enfocar el análisis en un grupo específico de clústeres. Alternativamente, también puede utilizarse como un paso de preprocesamiento para otros análisis, combinando diferentes técnicas (Jain, 2010; Everitt y Hothorn, 2011; Han et al., 2003).

Para determinar el número óptimo de clústeres, se implementó la metodología propuesta por Charrad et al. (2014) mediante el paquete NbClust en R, el cual evalúa múltiples índices de validación y selecciona el número óptimo de clústeres según la regla de la mayoría. El procedimiento consideró un rango de 2 a 9 clústeres, aplicando tanto métodos gráficos como cuantitativos, destacando:

- Índice de Hubert, que evalúa la estabilidad del clustering y sugiere el número óptimo a partir de un pico en la segunda diferencia del índice.
- Índice D (Dunn Index), que mide la compacidad y separación de los clústeres, identificando el número óptimo a través de un máximo en su segunda diferencia.
- Regla de la mayoría en NbClust, que selecciona el número de clústeres con mayor respaldo entre los métodos evaluados.

Para la construcción de los clústeres, se utilizó el método de agrupamiento *Complete Linkage*, el cual define la distancia entre dos clústeres como la mayor distancia entre cualquier par de observaciones de diferentes grupos, favoreciendo la formación de clústeres compactos y bien separados.

Una vez determinado el número óptimo de clústeres, se validó la estructura de agrupamiento mediante el índice de Silhouette (Rousseeuw, 1987) y el método del Gap Statistic (Tibshirani et al., 2001), los cuales permiten evaluar la calidad de la segmentación considerando la cohesión interna y la separación entre grupos.

Finalmente, la selección del número óptimo de clústeres se realizó considerando la convergencia entre los diferentes criterios de validación, asegurando una segmentación representativa de los datos analizados. Para la agrupación, se empleó el algoritmo K-Medias, uno de los métodos más utilizados en la literatura para la clasificación de datos multivariados (Jain, 2010; Nurdin et al, 2023). Su popularidad radica en su capacidad para encontrar estructuras en grandes volúmenes de datos al minimizar la variabilidad dentro de cada grupo, permitiendo identificar patrones en los desembarques de especies altamente migratorias.

El análisis de clústeres se basa en minimizar la suma del error cuadrático entre los puntos y sus respectivos centroides. La función objetivo del algoritmo K-Medias se expresa como:

$$J(C) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

Donde:

$J(C)$  es la suma de los errores cuadráticos dentro de los clústeres.

$x_i$  representa los puntos de datos.

$\mu_k$  es el centroide del clúster k.

$c_k$  representa los puntos que pertenecen al clúster k.

El procedimiento se repite hasta que la distribución de clústeres se estabilice.

Se analizó el comportamiento del análisis multivariado sin incluir determinadas especies de interés, se realizaron análisis adicionales excluyendo una especie a la vez. Se llevaron a cabo cuatro análisis independientes: uno sin tiburón azulejo, otro sin tiburón marrajo, otro sin pez espada y el

último sin dorado de altura. Posteriormente, estos resultados fueron comparados con el análisis general y con la serie temporal correspondiente a cada especie excluida. Este enfoque permitió evaluar el impacto de cada especie en la formación de clústeres y en la trayectoria general del análisis.

### 3.6 Implementación de modelos

#### 3.6.1 *Random Forest*

Para finalizar, se utilizó Random Forest, un método de aprendizaje supervisado basado en la combinación de múltiples árboles de decisión, cuya principal ventaja es la reducción del sobreajuste en comparación con un solo árbol, mejorando así la capacidad predictiva. El algoritmo emplea *bagging* (*bootstrap aggregation*) para generar diferentes subconjuntos de datos y construir árboles de decisión independientes. Luego, los resultados se combinan mediante un proceso de agregación, donde en clasificación se selecciona la clase más votada y en regresión se promedian las predicciones de los árboles. Esta metodología está basada en Hastie, Tibshirani y Friedman (2009).

Dado un conjunto de datos de entrenamiento con  $n$  observaciones y  $p$  variables, el algoritmo de Random Forest puede ser descrito como se menciona a continuación:

1. Se selecciona un subconjunto aleatorio de los datos mediante muestreo con reemplazo.
2. Se construyen árboles de decisión profundos sin poda.
3. En cada nodo de división, se selecciona aleatoriamente un subconjunto de  $m$  variables (con  $m < p$ ).
4. Se repite el proceso hasta construir  $B$  árboles.
5. Se obtiene la predicción final promediando los resultados.

Para modelos de Random Forest enfocados en regresión, se calcula el promedio de las predicciones de todos los árboles

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

Donde:

$\hat{y}$  es el valor predicho final obtenido del modelo de regresión

$B$  es el número total de árboles en el bosque aleatorio (Random Forest).

$T_b(x)$  es la predicción del árbol  $b$  para una entrada  $x$ .

$x$  es el vector de características (valores de las variables predictoras).

$\sum_{b=1}^B T_b(x)$  es la suma de las predicciones de todos los árboles en el modelo.

### 3.6.2 XGBoost

XGBoost (*Extreme Gradient Boosting*) es un algoritmo de aprendizaje supervisado basado en *boosting* que optimiza el rendimiento de modelos de árboles de decisión mediante la minimización de una función de pérdida y la adición de árboles secuenciales. Se utilizó como método complementario a *Random Forest* para evaluar la capacidad predictiva de los desembarques pesqueros en función de variables ambientales y presas.

El Algoritmo de XGBoost puede ser descrito a partir de un conjunto de datos con  $n$  observaciones y  $p$  variables predictoras, XGBoost minimiza la siguiente función de pérdida regularizada:

$$L(\Theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

Donde

$l(y_i, \hat{y}_i)$  es la función de pérdida (error cuadrático medio en regresión).

$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum w_j^2$  es el término de regularización, que penaliza la complejidad del modelo para evitar sobreajuste.

$\gamma$  controla la penalización por cantidad de hojas en los árboles ( $T$ ), y

$\lambda$  regula la magnitud de los pesos ( $w_j$ ).

### 3.6.3 Evaluación de los modelos

Para la evaluación de los modelos propuestos se utilizarán las siguientes medidas de desempeño:

**Error cuadrático medio (MSE) para regresión:**

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

**Coefficiente de determinación ( $R^2$ ) para regresión:**

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Donde:

$y_i$  = valor real de la variable respuesta.

$\hat{y}_i$  = valor predicho por el modelo.

$\bar{y}$  = media de los valores reales.

$n$  = número total de observaciones.

### 3.6.4 Implementación

La implementación de los modelos *Random Forest* y *XGBoost* se llevó a cabo para evaluar el impacto de las variables ambientales en los desembarques pesqueros. Se integraron variables pesqueras, incluyendo el desembarque de pez espada, tiburones y dorado de altura, junto con variables ambientales como PDO, IOS, HCI, MEI, N12, N34, sus anomalías y TPI, además de las variables tróficas DANC y DSAR.

Se aplicó un preprocesamiento riguroso de los datos, incluyendo normalización y tratamiento de valores perdidos, con el objetivo de mejorar la calidad del modelo. Posteriormente, los datos fueron divididos en un conjunto de entrenamiento (80%) y prueba (20%), asegurando que la evaluación se realizara con datos no utilizados en el ajuste del modelo. La separación se llevó a cabo utilizando la función *createDataPartition()* de la librería *caret*, garantizando una distribución balanceada de la variable respuesta en ambos conjuntos.

En la etapa de optimización de hiperparámetros, se ajustaron el número de árboles (B), la profundidad máxima de los árboles, la cantidad de variables consideradas en cada nodo (*mtry*) para *Random Forest*, y para *XGBoost*, se optimizaron *nrounds*, *max\_depth*, la tasa de aprendizaje (*eta*) y la fracción de datos utilizada en cada iteración (*subsample*). Finalmente, se compararon el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación ( $R^2$ ) de ambos modelos.

# 4 Resultados

---

## 4.1 Esquema metodológico

Se utilizó el esquema metodológico basado en el modelo propuesto por Cady (2017; **Figura 2**), abordando el trabajo en seis etapas.

### 1. Enmarcar el problema

El estudio identifica patrones en la interacción entre los desembarques de especies altamente migratorias y las variables ambientales en Chile, aplicando herramientas de ciencia de datos para el análisis de series de tiempo. Se consideran las especies de interés: pez espada, tiburón azul, marrajo, tiburones y dorado de altura, junto con variables ambientales clave como PDO, IOS, HCI, MEI, N12, N34, sus anomalías y TPI. El enfoque metodológico permite evaluar la influencia de las variables ambientales en la variabilidad de los desembarques y determinar patrones (Ver Objetivos)

### 2. Entendimiento de los datos

Los datos analizados comprenden series de tiempo de desembarques pesqueros y variables ambientales desde 1980 hasta 2022. Se realiza un análisis exploratorio mediante estadísticas descriptivas, visualización de tendencias temporales y detección de patrones preliminares. La evaluación de correlaciones entre desembarques y variables ambientales permite identificar relaciones clave (Ver Sección 4.2).

### 3. Extracción de características

Los datos se estructuran en formato tabular y se generan nuevas variables relevantes para el análisis. Se aplica Análisis de Componentes Principales (PCA) para reducir la dimensionalidad y capturar las principales fuentes de variabilidad en las series temporales. Además, se emplean técnicas de agrupamiento para segmentar la serie de tiempo en distintos períodos con características diferenciadas (Ver Sección 4.3)

### 4. Modelamiento y análisis

Se implementan modelos de series de tiempo para evaluar la evolución de los desembarques y su relación con los índices climáticos. La aplicación de funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) permite detectar dependencias temporales y estructuras recurrentes en los datos. Se utilizan métodos de agrupamiento, como k-Medias y análisis jerárquico, para identificar cambios en los desembarques. Además, se aplican modelos de aprendizaje automático, como *Random Forest* y *XGBoost*, para evaluar la relación entre desembarques y variables ambientales. (Ver Secciones 4.4, 4.5 y 4.6)

## 5. Presentación de resultados

Los patrones detectados en la dinámica pesquera se presentan mediante gráficos y tablas, evidenciando la evolución de los desembarques y su relación con las variables ambientales. Se identifican periodos diferenciados en la serie temporal. (Ver Sección 4)

## 6. Implementación práctica

Se desarrolla un módulo analítico en R con código reproducible para facilitar la replicabilidad de los análisis (Anexo 1). La implementación de estos modelos permite evaluar la influencia de los patrones ambientales en los desembarques, proporcionando herramientas para conocer de mejor manera la pesquería. (Ver Secciones 7 y 8)

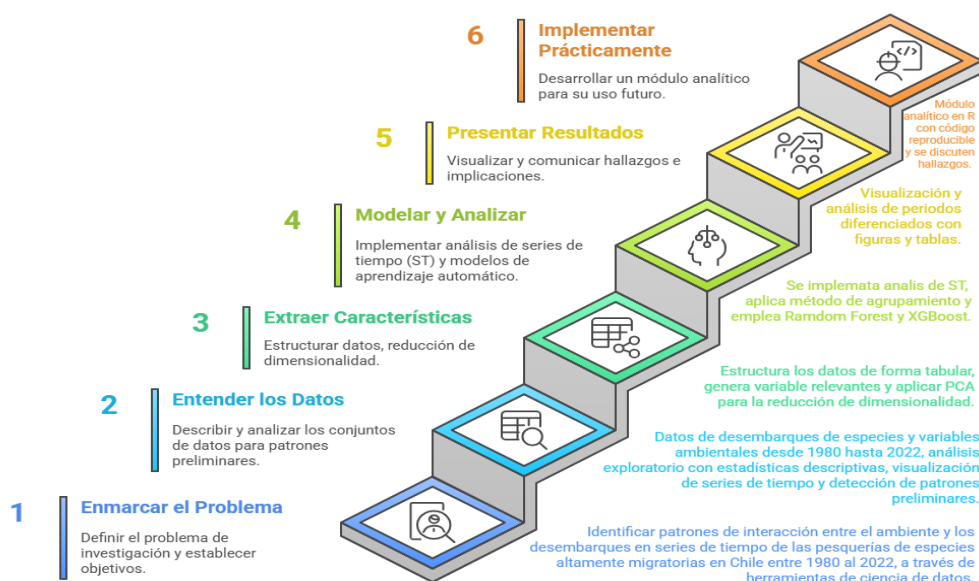


Figura 2. Esquema metodológico modificado, propuesto por Cady (2017)

### 4.2 Análisis descriptivo

Para llevar a cabo el análisis descriptivo de las variables pesqueras y ambientales, se recopilaron y organizaron datos históricos de desembarques de especies clave y series temporales de variables ambientales. Se aplicaron estadísticas descriptivas básicas, como medias, medianas, desviaciones estándar y coeficientes de variación, con el fin de evaluar la dispersión y el comportamiento de las series a lo largo del tiempo.

#### 4.2.1 Datos pesqueros

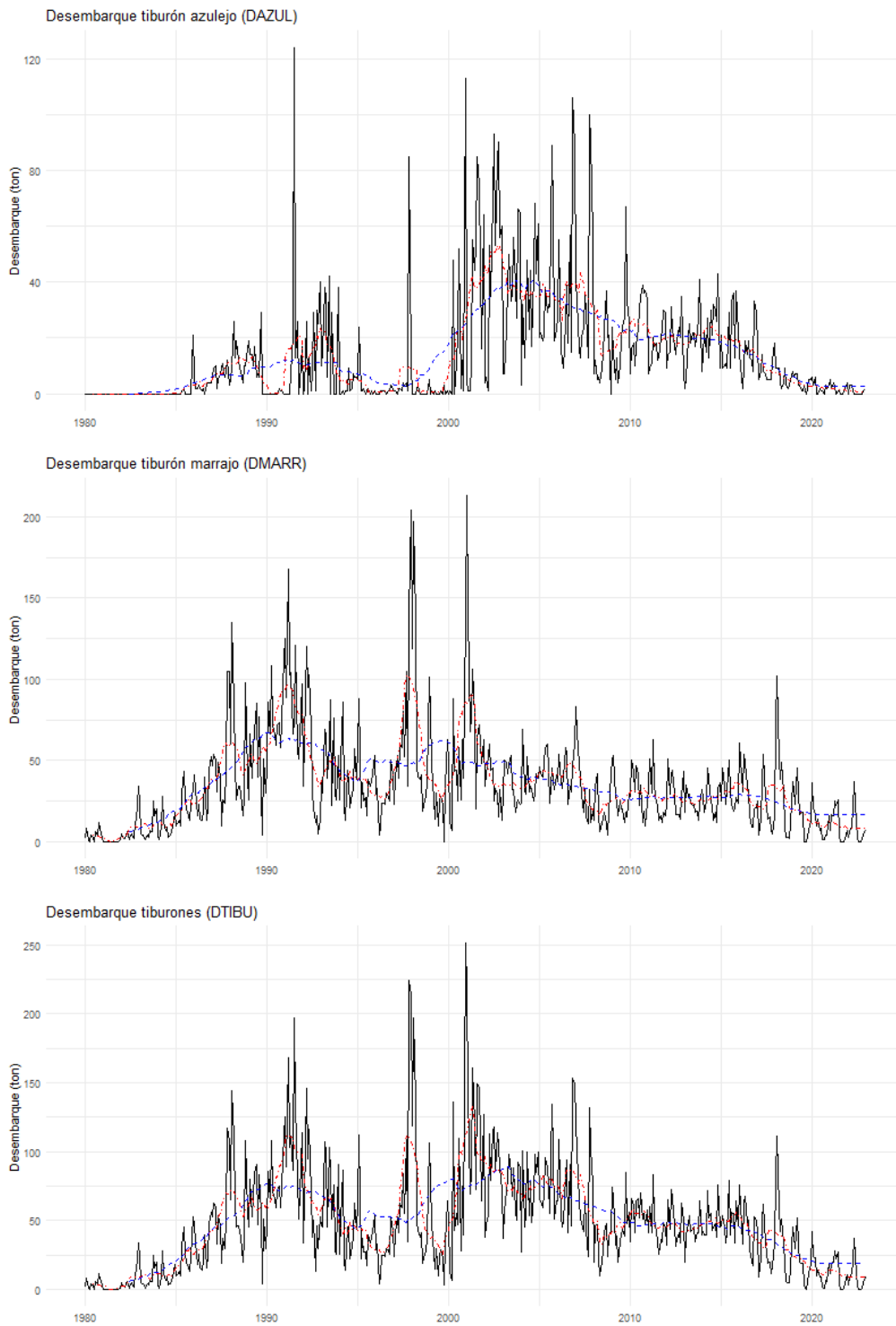
A continuación, se muestran los resultados del análisis del desembarque de las series de tiempo (desde 1980 hasta 2022) de las especies de interés como peces cartilaginosos altamente migratorias: tiburón azulejo (*Prionace glauca*, DAZUL), tiburón marrajo (*Isurus oxyrinchus*, DMARR), grupo de tiburones (sumatoria de tiburón azulejo y marrajo, DTIBU); peces óseos altamente migratorios: pez espada (*Xiphias gladius*, DESPA) y dorado de altura (*Coryphaena*

*hippurus*, DDORA) y especies que forman parte de la dieta de todas ellas, tales como anchoveta (*Engraulis ringens*, DANC) y sardina española (*Sardinops sagax*, DSAR). Se presentan las tendencias generales de las fluctuaciones estacionales en los desembarques de todas las especies en conjunto.

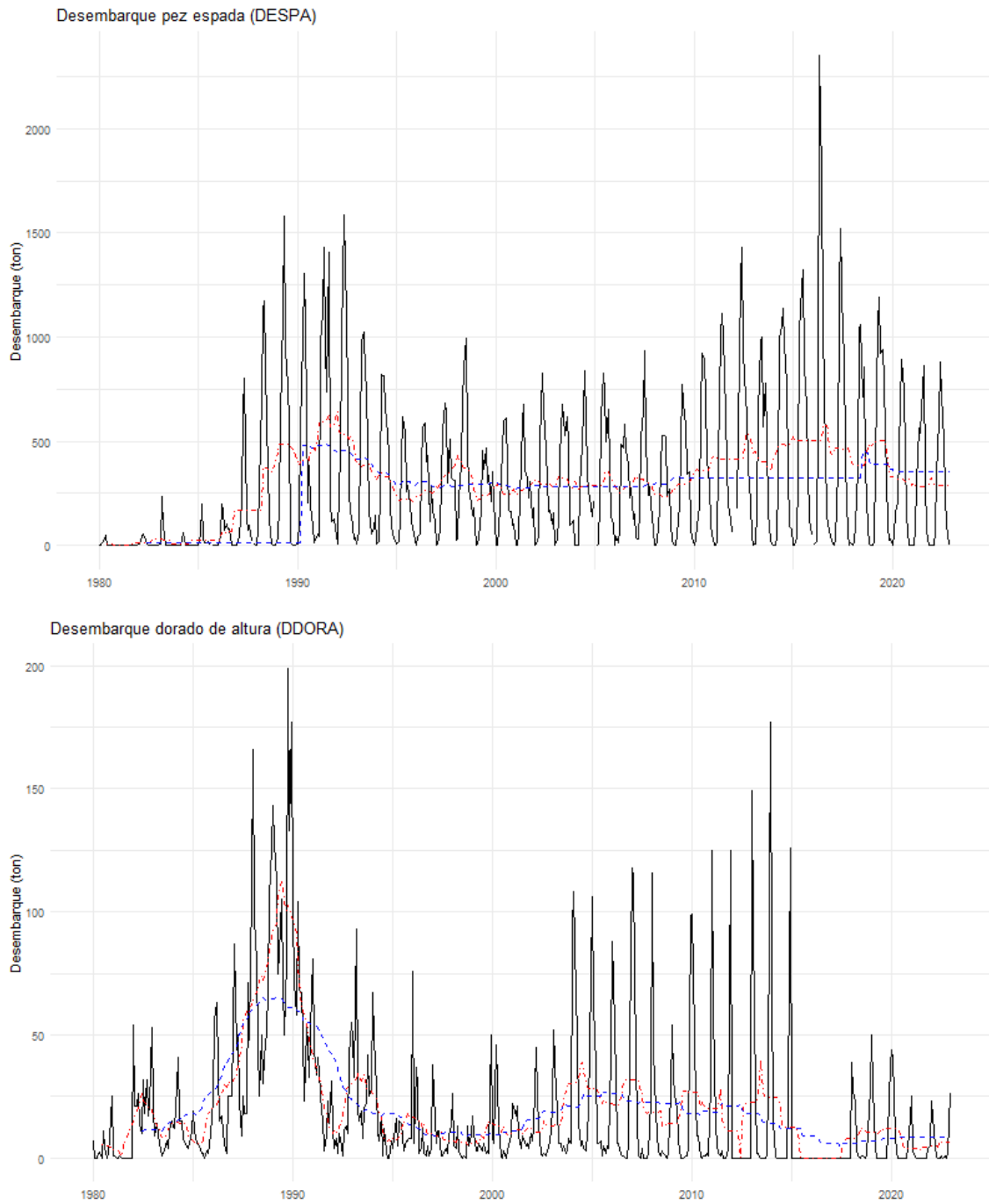
La **Figura 3**, presenta las series de tiempo de desembarques de tiburón azul, tiburón marrajo y el grupo de tiburones. Se observa un incremento progresivo en los desembarques durante los años 80 y 90, con picos significativos alrededor de finales de los 90 y principios de los 2000, seguido de una marcada disminución en las décadas posteriores. Las fluctuaciones estacionales son evidentes, con ciclos recurrentes de desembarques altos en todos los casos, especialmente en DMARR y DTIBU. Las medias móviles (representadas por líneas de colores) muestran la tendencia a largo plazo. En los tres casos, las medias móviles reflejan un ascenso sostenido hasta finales de los 90, seguido de una disminución constante a partir de los 2000. DMARR y DTIBU presentan medias móviles más altas que DAZUL, lo que indica mayores volúmenes de desembarque, pero con una tendencia decreciente más pronunciada. La estabilización de las medias móviles hacia 2020 sugiere una reducción en la variabilidad de los desembarques.

La **Figura 4** muestra las series de tiempo de desembarques del pez espada y el dorado de altura, donde se observa un patrón común en ambas especies, con picos significativos de desembarque durante finales de los años 80 y principios de los 90, seguidos de una marcada disminución a partir del 2000. El pez espada exhibe mayores volúmenes de desembarque en comparación con el dorado de altura, aunque ambos presentan fluctuaciones estacionales bien definidas a lo largo del tiempo. Las medias móviles (representadas en líneas de colores) reflejan una tendencia decreciente sostenida, más pronunciada en DESPA, mientras que DDORA muestra una disminución menos acentuada pero persistente. Estos resultados sugieren que, aunque ambas especies siguen ciclos estacionales, las capturas han disminuido a lo largo de las décadas.

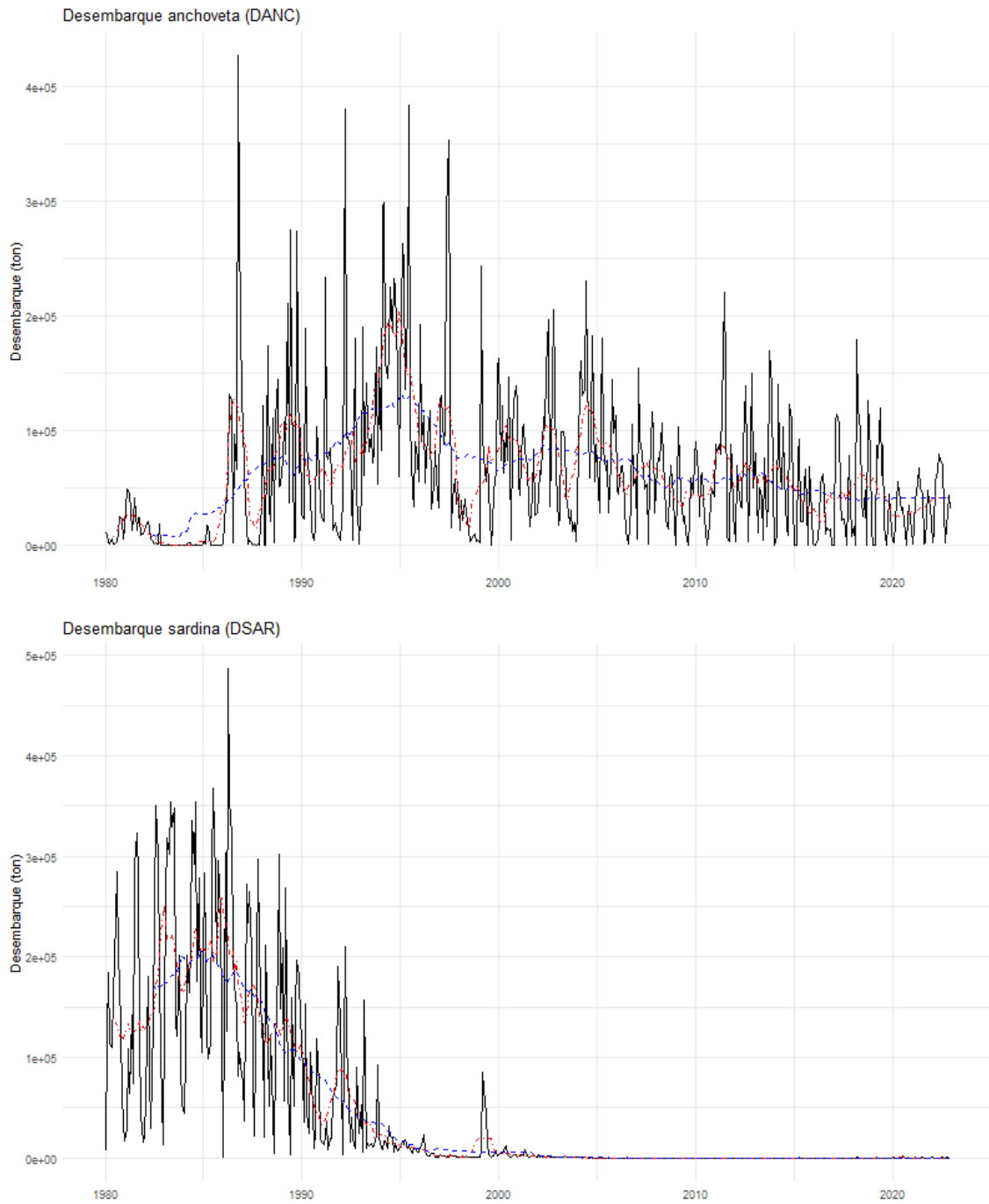
La **Figura 5** muestra las series de tiempo de desembarques de anchoveta y sardina española. Se observa que ambas especies presentan un incremento significativo en los desembarques durante los años 80 y principios de los 90, con picos pronunciados, seguidos de una disminución abrupta y sostenida a partir de finales de los 90 y 2000. Las fluctuaciones estacionales son evidentes en ambos casos, aunque más pronunciadas en la anchoveta. Las medias móviles, representadas por líneas de colores, revelan la tendencia a largo plazo de los desembarques. En la anchoveta, la media móvil muestra un ascenso hasta mediados de los 90, reflejando un periodo de alta productividad, seguido de una fuerte caída que se mantiene estable en niveles bajos a partir de los 2000. En el caso de la sardina, la media móvil destaca una disminución más marcada desde principios de los 90, con desembarques que se reducen casi a cero hacia el 2000, y permanecen en niveles mínimos hasta el final del periodo analizado. Estos resultados indican que, aunque ambas especies experimentaron periodos de alta abundancia, la sardina ha sufrido una reducción más drástica y sostenida en comparación con la anchoveta.



**Figura 3.** Desembarque de tiburón azulejo (DAZUL), tiburón marrajo (DMARR) y grupo de tiburones (DTIBU). Media móvil de 12 meses (línea punteada en rojo) y de 60 meses (línea punteada en azul).



**Figura 4.** Desembarque de pez espada (DESPA) y dorado de altura (DDORA). Media móvil de 12 meses (línea punteada en rojo) y de 60 meses (línea punteada en azul).



**Figura 5.** Desembarque anchoveta (DANC) y sardina (DSAR). Media móvil de 12 meses (línea punteada en rojo) y de 60 meses (línea punteada en azul)

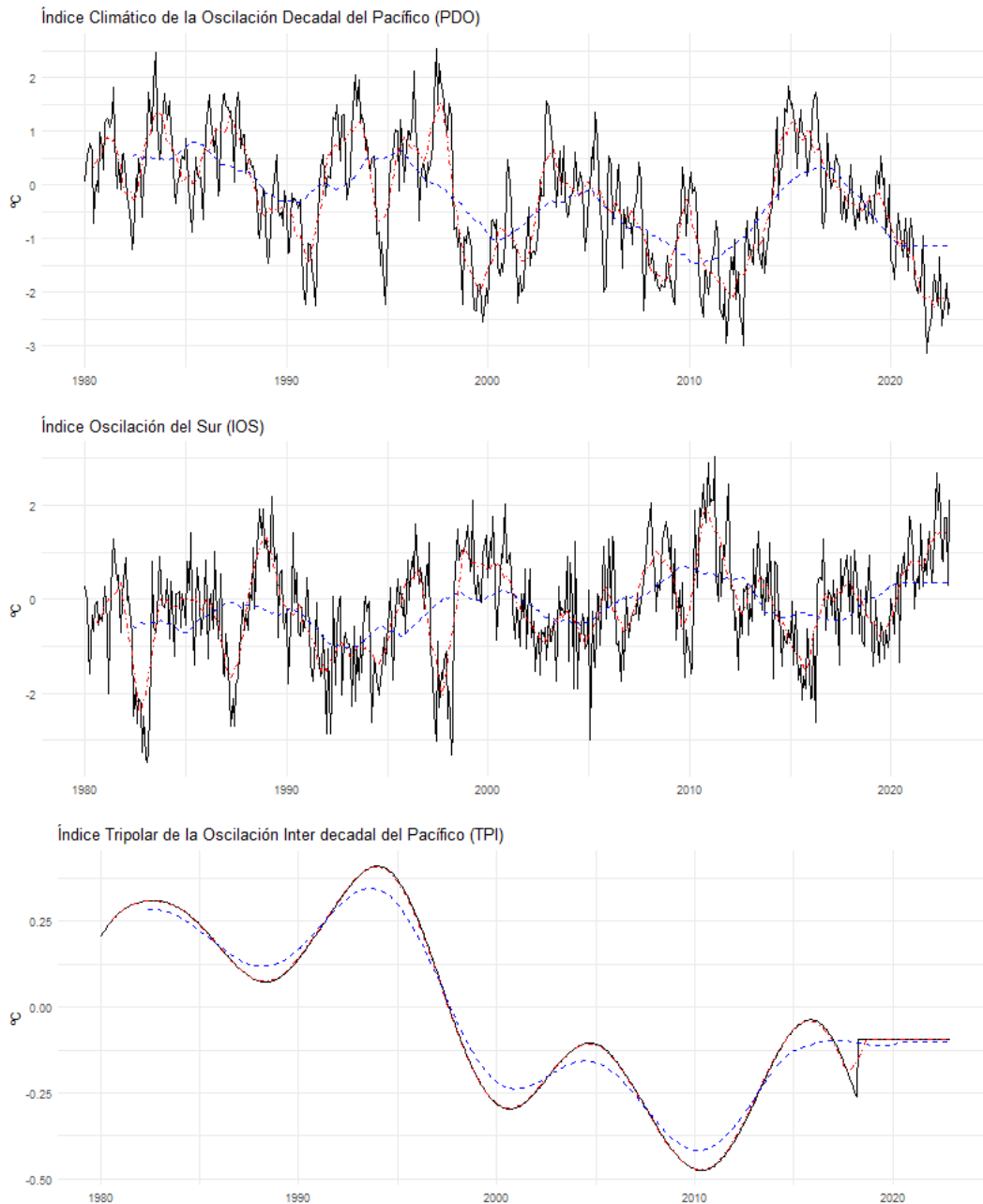
#### 4.2.2 Datos ambientales

El análisis de los índices climáticos revela patrones significativos de variabilidad y tendencias a largo plazo en los principales fenómenos que afectan la costa oriental del océano Pacífico (**Figura 6**). El Índice Climático de la Oscilación Decadal del Pacífico (PDO) muestra fluctuaciones decenales claras, con cambios cada 10-15 años. A lo largo del periodo analizado, se observan fases alternas de valores positivos y negativos, lo que indica ciclos de calentamiento y enfriamiento. Desde 2010, el PDO ha experimentado una tendencia negativa, con una leve recuperación hacia 2020. Las medias móviles a 12 meses suavizan las fluctuaciones estacionales, mientras que las medias a 60 meses resaltan las tendencias a largo plazo, revelando cambios estructurales en la dinámica climática del océano. Por su parte, el Índice de Oscilación del Sur (IOS) presenta oscilaciones regulares con mayor amplitud en comparación con el PDO, especialmente entre 1980 y 2000. Este índice refleja episodios de El Niño y La Niña, con variaciones más extremas durante ese periodo. A partir de 2000, la media móvil a 60 meses muestra una recuperación gradual de los valores negativos, indicando una estabilización del sistema. Las fluctuaciones cíclicas del IOS, aunque más irregulares, siguen patrones similares a los del PDO, lo que sugiere una interacción entre ambos fenómenos en la regulación del clima del Pacífico. El Índice Tripolar de la Oscilación Interdecadal del Pacífico (TPI) muestra un patrón cíclico con oscilaciones que reflejan variabilidad interdecadal. A diferencia del PDO y el IOS, el TPI presenta fluctuaciones más suaves, con fases bien definidas de ascenso y descenso a lo largo del periodo analizado. Las medias móviles a 12 y 60 meses reflejan este comportamiento, resaltando tres ciclos principales desde 1980 hasta 2023. En la última década, se observa una fase negativa con una leve recuperación hacia el final del periodo. Este comportamiento sugiere que el TPI podría desempeñar un papel importante en la modulación de otros índices climáticos, reflejando cambios en la circulación oceánica a gran escala.

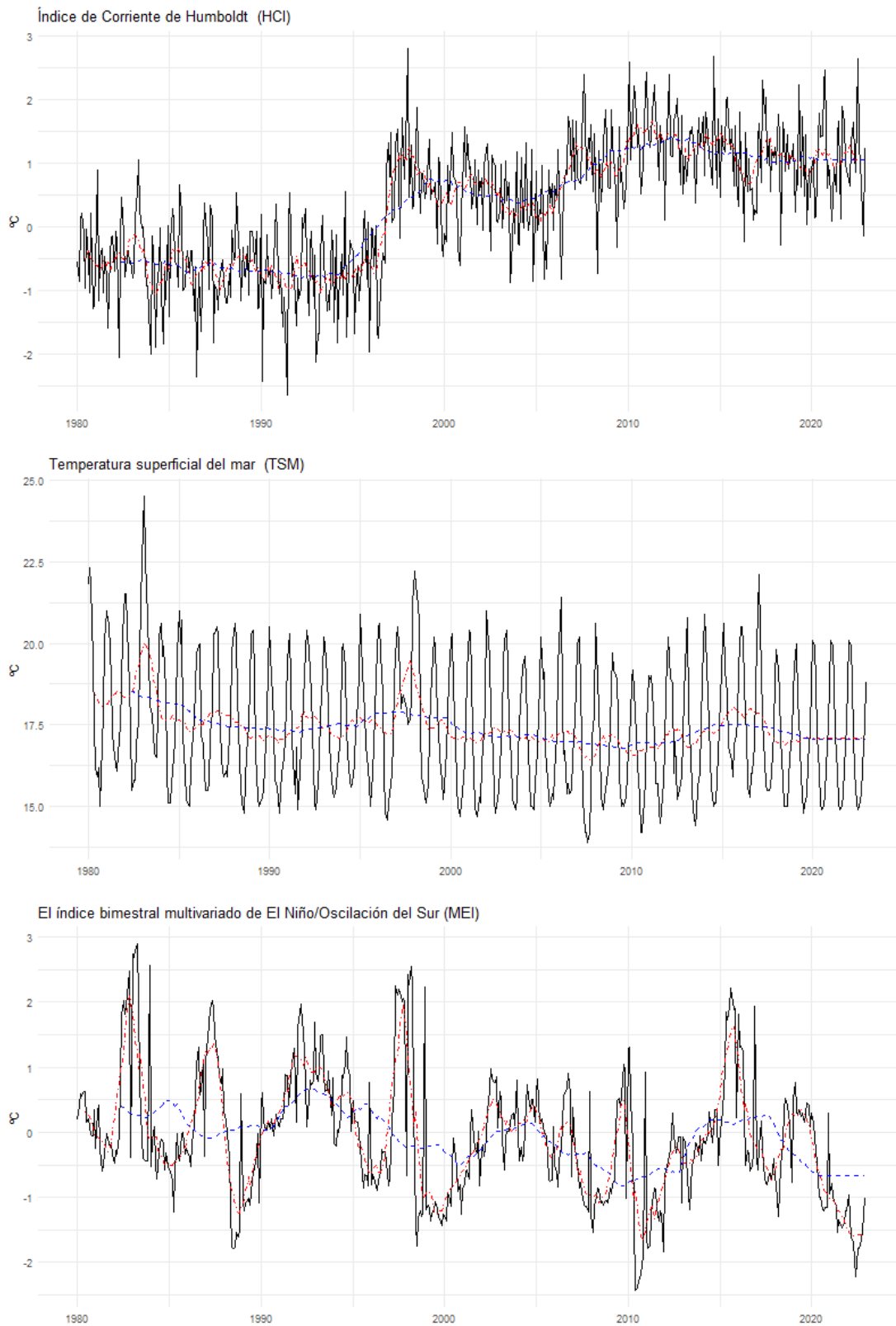
La **Figura 7** muestra el análisis del Índice de Corriente de Humboldt (HCI) y de la Temperatura Superficial del Mar (TSM), la cual refleja una estrecha relación con los patrones climáticos globales. El HCI muestra picos durante los años 90 y una disminución progresiva desde 2000, con una ligera recuperación reciente. Las fluctuaciones estacionales son evidentes, pero las medias móviles a 60 meses indican una tendencia decreciente en el último periodo. En cuanto a la TSM, se observa una tendencia creciente desde principios de los 2000, con picos asociados a eventos de El Niño. La media móvil a 60 meses destaca un incremento sostenido de las temperaturas desde 2010 hasta fines unos años antes del 2020, lo que sugiere un posible efecto del cambio climático sobre el calentamiento oceánico y para los últimos años desde 2019 se observa una estabilización de la temperatura. El Índice Multivariado de El Niño (MEI) ofrece una visión integral de los fenómenos de El Niño y La Niña, reflejando oscilaciones bien definidas con picos significativos en años clave como 1997-1998 y 2015-2016. Las fluctuaciones estacionales son evidentes, y la media móvil a 12 meses resalta estas variaciones, vinculadas a los episodios de calentamiento y enfriamiento en el Pacífico. La media móvil a 60 meses muestra una recuperación parcial en los últimos años, aunque con una tendencia general hacia el enfriamiento reciente. Esto sugiere que, aunque los eventos extremos persisten, hay indicios de estabilización en los valores de MEI hacia el final del periodo analizado.

Finalmente, en la **Figura 8** se visualizan, los índices de la región Niño (N12 y N34) y sus anomalías (Anom N12 y Anom N34) que reflejan patrones característicos de los eventos de El Niño y La Niña. Las fluctuaciones estacionales en el índice N12 son marcadas, con picos significativos en años

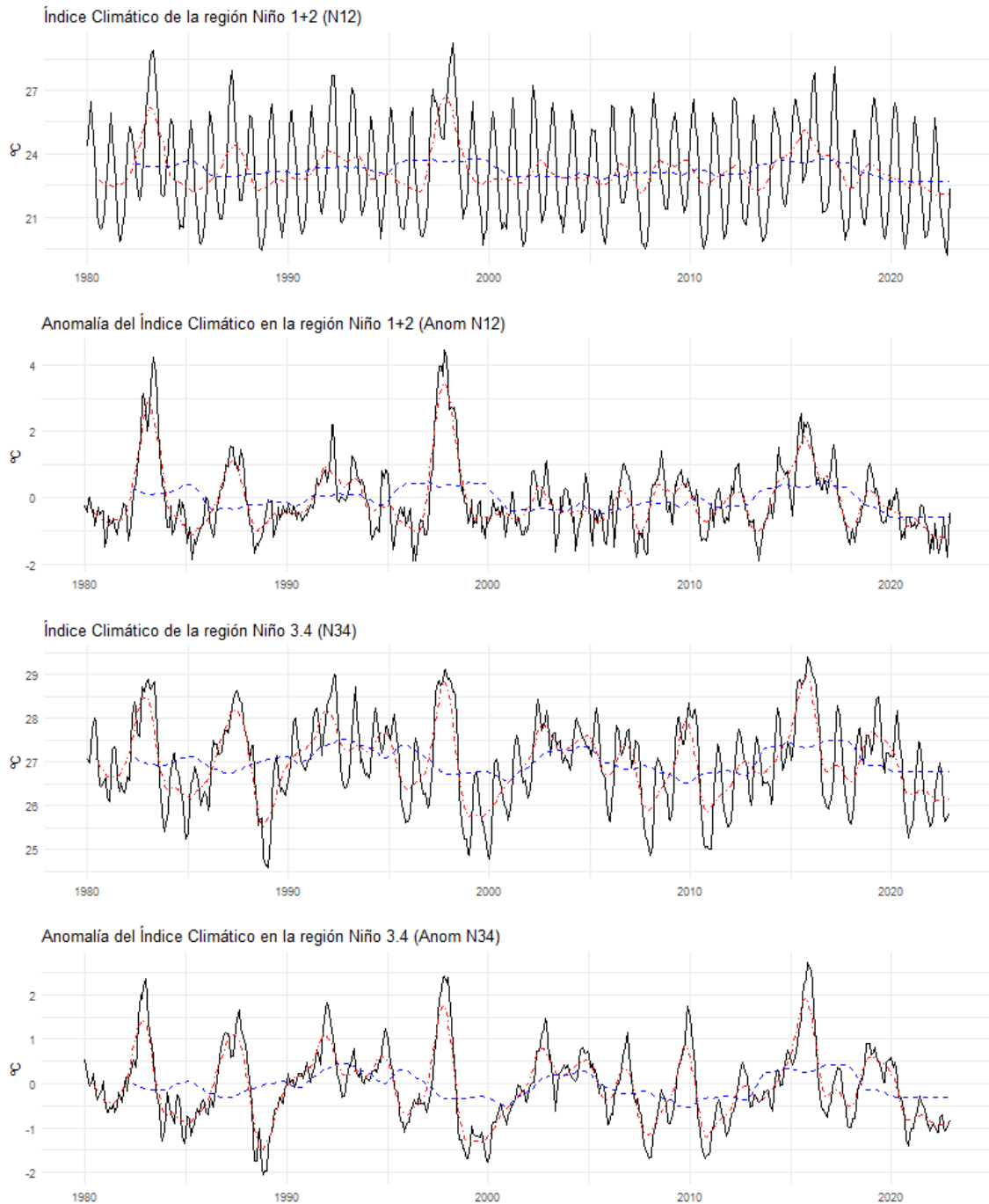
como 1982-1983 y 1997-1998. Las anomalías siguen patrones similares, mostrando picos más pronunciados durante eventos extremos. La media móvil a 60 meses indica que, aunque las fluctuaciones estacionales son notorias, los valores generales han permanecido relativamente estables desde 2000, con algunas oscilaciones hacia la baja en años recientes. La estabilidad en los índices Niño sugiere una fase de menor intensidad en comparación con décadas anteriores, aunque persiste la influencia de El Niño y La Niña en la variabilidad climática del Pacífico.



**Figura 6.** Serie de tiempo de las variables ambientales PDO, IOS y TPI. Media móvil de 12 meses (línea punteada en rojo) y de 60 meses (línea punteada en azul)



**Figura 7.** Serie de tiempo de las variables ambientales HCI, TSM y MEI. Media móvil de 12 meses (línea punteada en rojo) y de 60 meses (línea punteada en azul)



**Figura 8.** Serie de tiempo de las variables ambientales N2, Anom N2, N34 y Anom N34. Media móvil de 12 meses (línea punteada en rojo) y de 60 meses (línea punteada en azul)

### **Análisis descriptivo de las variables pesqueras y ambientales**

Se realizó un análisis descriptivo de las variables pesqueras (**Tabla 2**), las cuales presentaron una alta variabilidad en los desembarques, destacando DANC y DSAR como las especies con mayor dispersión, mientras que las variables ambientales, como TSM y N12, muestran fluctuaciones estacionales moderadas. En general, la desviación estándar indica que las variables pesqueras tienen una mayor dispersión en comparación con las variables ambientales. En términos de menor

variación, DAZUL tiene la menor variabilidad en los desembarques, con una desviación estándar de 20, lo que indica que los valores de desembarque de esta especie son más constantes en comparación con otras. En cuanto a las variables ambientales, el TPI muestra la menor variabilidad, con una desviación estándar de 0,2; lo que sugiere que los valores de este índice son más estables a lo largo del tiempo en comparación con otras variables climáticas y oceánicas.

El análisis de correlación de Spearman (**Tabla 3**), destacó por valores de R superiores a 0,7, interpretadas como de alta correlación positiva, como en los casos de DMARR-DTIBU ( $R = 0,87$ ), N34-Anom 34 ( $R = 0,86$ ), MEI-Anom 34 ( $R = 0,79$ ), y una alta correlación negativa entre IOS-MEI ( $R = -0,72$ ). Las correlaciones entre las variables pesqueras y ambientales muestran valores positivos moderados entre DSAR-TPI ( $R = 0,70$ ) y una correlación negativa moderada entre DESPA-TSM ( $R = -0,59$ ), DAZUL-TPI ( $R = -0,52$ ) y DSAR-HCI ( $R = -0,70$ ). Además, se identificaron correlaciones bajas pero positivas entre DESPA-N34 ( $R = 0,36$ ), DAZUL-HCI ( $R = 0,31$ ), DMARR-N12 ( $R = 0,32$ ), DDORA-TSM ( $R = 0,42$ ) y DSAR-PDO ( $R = 0,34$ ), así como una baja correlación negativa entre DDORA ( $R = -0,37$ ).

**Tabla 2.** Análisis descriptivo de las variables de la serie histórica desde el año 1980 al 2022.

	Variables	Media	Mínimo	Máximo	Desviación estándar
Variables pesqueras	DESPA	312,5	0,0	2350	374
	DAZUL	13,8	0,0	124	20
	DMARR	34,4	0,0	213	30
	DTIBU	48,2	0,0	251	39
	DDORA	20,2	0,0	199	32
	DANC	61.551,3	0,0	426.727	65.783
	DSAR	45.203,7	0,0	487.151	86.583
Variables ambientales	PDO	-0,3	-3,1	2,6	1,2
	IOS	-0,2	-3,5	3,0	1,2
	HCI	0,3	-2,6	2,8	1,0
	TSM	17,5	13,9	24,5	2,0
	MEI	-0,1	-2,4	2,9	1,0
	N12	23,2	19,2	29,2	2,3
	Anom N12	-0,1	-1,9	4,5	1,1
	N34	27,0	24,6	29,4	1,0
	Anom N34	-0,1	-2,1	2,7	0,9
	TPI	0,0	-0,5	0,4	0,2

**Tabla 3.** Matriz de correlación de Spearman entre variables. Los valores de correlación en color verde claro mayor a 0,30; en verde oscuro entre 0,50 a 0,70 y azul mayor a 0,70.

	DESPA	DAZUL	DMARR	DTIBU	DDORA	DANC	DSAR	PDO	IOS	HCI	TSM	MEI	N12	Anom N12	N34	Anom N34	TPI
DESPA	1																
DAZUL	0,32	1															
DMARR	0,24	0,30	1														
DTIBU	0,31	0,64	0,87	1													
DDORA	-0,32	-0,05	0,37	0,26	1												
DANC	0,27	0,22	0,31	0,30	0,12	1											
DSAR	-0,29	-0,49	0,02	-0,14	0,39	-0,09	1										
PDO	-0,11	-0,12	0,06	0,00	0,09	-0,09	0,34	1									
IOS	-0,09	-0,03	-0,18	-0,17	-0,03	0,02	-0,18	-0,42	1								
HCI	0,23	0,31	-0,04	0,07	-0,37	0,06	-0,70	-0,32	0,22	1							
TSM	-0,59	-0,18	0,15	0,04	0,42	-0,03	0,14	0,22	-0,09	-0,11	1						
MEI	0,00	0,06	0,17	0,20	0,07	-0,05	0,22	0,50	-0,72	-0,27	0,15	1					
N12	-0,09	-0,06	0,32	0,18	0,29	0,10	0,02	0,26	-0,24	0,03	0,67	0,24	1				
Anom N12	0,05	0,10	0,19	0,19	-0,02	-0,08	-0,05	0,29	-0,45	0,05	0,19	0,53	0,44	1			
N34	0,36	0,12	0,24	0,23	-0,05	0,14	0,04	0,39	-0,63	-0,03	-0,04	0,66	0,43	0,51	1		
Anom N34	0,13	0,15	0,22	0,25	0,02	0,04	0,02	0,40	-0,69	-0,08	0,10	0,79	0,29	0,61	0,86	1	
TPI	-0,20	-0,52	-0,09	-0,26	0,19	-0,11	0,70	0,54	-0,34	-0,66	0,16	0,41	0,06	0,09	0,20	0,25	1

### 4.3 Análisis de Componentes Principales

La **Tabla 4a** presenta los porcentajes de varianza explicada por cada componente principal (PC) en el Análisis de Componentes Principales (PCA). El PC1 explica el 28% de la varianza total, seguido por el PC2, que representa el 17%. A medida que se avanza en las componentes, la varianza explicada disminuye, con los componentes finales (PC16 y PC17) explicando únicamente el 0,01% cada uno. La varianza acumulada alcanza el 73% con las primeras cinco componentes, lo que sugiere que estas componentes explican la mayor parte de la variabilidad en los datos. El promedio de varianza explicada por componente es 0,06%, lo que indica que las componentes menos relevantes aportan de manera marginal al total de la varianza.

Por otro lado, la **Tabla 4b** presenta los resultados del PCA tomando como referencia las correlaciones superiores a  $\pm 0,40$ . En este análisis, el PC1 está principalmente explicado por variables ambientales, como PDO, IOS, MEI, N12, Anom N2, N34, Anom N34 y TPI. El PC2 está predominantemente influenciado por variables pesqueras, junto con dos variables ambientales (DESPA, DAZUL, DMARR, DTIBU, DSAR, HCI y TPI). El PC3 es explicado por una combinación de variables pesqueras y ambientales (DESPA, DMARR, DTIBU, DDORA, TSM y N12). Finalmente, los PC4 y PC5 están explicados principalmente por variables ambientales y pesqueras, respectivamente.

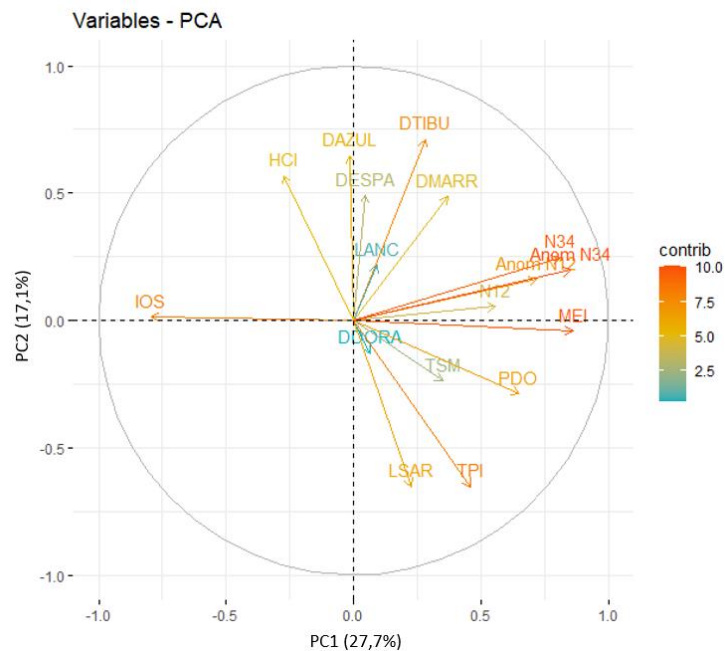
El gráfico de análisis de componentes principales (PCA, **Figura 9**) muestra la relación entre las variables climáticas y pesqueras, explicando el 27,7 % de la variabilidad en el eje del PC1 y el 17,1 % en el eje del PC2. Las variables relacionadas con el fenómeno de El Niño, como N34, Anom N34 y MEI, se encuentran altamente correlacionadas y contribuyen significativamente a la PC1. Además, el Índice de Oscilación Decadal del Pacífico (PDO) y el Índice Tripolar de la Oscilación Interdecadal del Pacífico (TPI) también muestran una fuerte influencia en este eje. Por otro lado, el Índice de Oscilación del Sur (IOS) se proyecta sobre PC2, indicando que captura una variabilidad distinta a la explicada por los otros índices.

En cuanto a las especies pesqueras, el desembarque de tiburones (DTIBU), tiburón marrajo (DMARR) y tiburón azul (DAZUL) presentan una mayor asociación con el PDO y el HCI, lo que

sugiere una relación entre estas especies y las fluctuaciones climáticas de largo plazo. Por el contrario, especies como la anchoveta (LANC) y el dorado de altura (DDORA) se agrupan más cerca del centro, indicando una menor variabilidad explicada por las dimensiones principales. El gráfico resalta cómo las variables climáticas tienen un papel clave en la dinámica pesquera, especialmente aquellas vinculadas al fenómeno de El Niño y las oscilaciones del Pacífico.

**Tabla 4.** a. Resumen transformación PCA, en negrilla los componentes seleccionados. b. Correlación entre las componentes principales seleccionadas y las variables explicativas. En plomo correlaciones mayores a 0,40.

a.			b.					
PC	% de varianza	% de varianza acumulada	Variables	PC1	PC2	PC3	PC4	PC5
PC1	<b>0,28</b>	<b>0,28</b>	DESPA	0,05	0,49	-0,43	0,23	0,44
PC2	<b>0,17</b>	<b>0,45</b>	DAZUL	-0,01	0,65	0,02	0,27	-0,50
PC3	<b>0,12</b>	<b>0,57</b>	DMARR	0,37	0,49	0,53	0,32	0,05
PC4	<b>0,10</b>	<b>0,67</b>	DTIBU	0,28	0,71	0,43	0,39	-0,22
PC5	<b>0,07</b>	<b>0,73</b>	DDORA	0,06	-0,13	0,67	0,14	0,04
PC6	0,05	0,78	DANC	0,09	0,22	0,17	0,33	<b>0,63</b>
PC7	0,04	0,82	DSAR	0,23	-0,65	-0,01	0,21	-0,21
PC8	0,04	0,86	PDO	0,65	-0,29	-0,06	0,11	0,02
PC9	0,03	0,89	IOS	-0,79	0,01	0,22	-0,06	0,08
PC10	0,03	0,92	HCI	-0,27	0,57	-0,15	-0,58	0,01
PC11	0,02	0,94	TSM	0,35	-0,24	0,70	-0,44	-0,02
PC12	0,02	0,96	MEI	0,86	-0,04	-0,17	0,01	-0,16
PC13	0,01	0,98	N12	0,56	0,05	0,44	-0,47	0,30
PC14	0,01	0,99	Anom N12	0,72	0,16	-0,09	-0,34	-0,09
PC15	0,01	1,00	N34	0,81	0,25	-0,33	-0,05	0,16
PC16	0,00	1,00	Anom N34	0,85	0,20	-0,22	-0,10	-0,10
PC17	0,00	1,00	TPI	0,46	-0,65	-0,03	0,40	0,11
Pormedio	0,06							



**Figura 9.** Valor de Contribución de las Variables en PCA (PC1 y PC2).

#### 4.4 Análisis de Series de Tiempo y Autocorrelación

##### Descomposición de las series temporales para pez espada, tiburón y dorado de altura.

Las **Figura 10**, **Figura 12** y **Figura 14** muestran la descomposición aditiva de los desembarques para pez espada, tiburones y dorado de altura. Se observa una tendencia creciente en ciertas décadas, seguida de estabilidad o caída en los desembarques. Además, la componente estacional es fuerte, lo que indica que las capturas tienen patrones recurrentes anuales. La componente irregular es variable, sugiriendo fluctuaciones no explicadas por la tendencia o estacionalidad. Las capturas presentan patrones predecibles de estacionalidad, pero la tendencia y la variabilidad requieren un modelo para capturar dinámicas subyacentes.

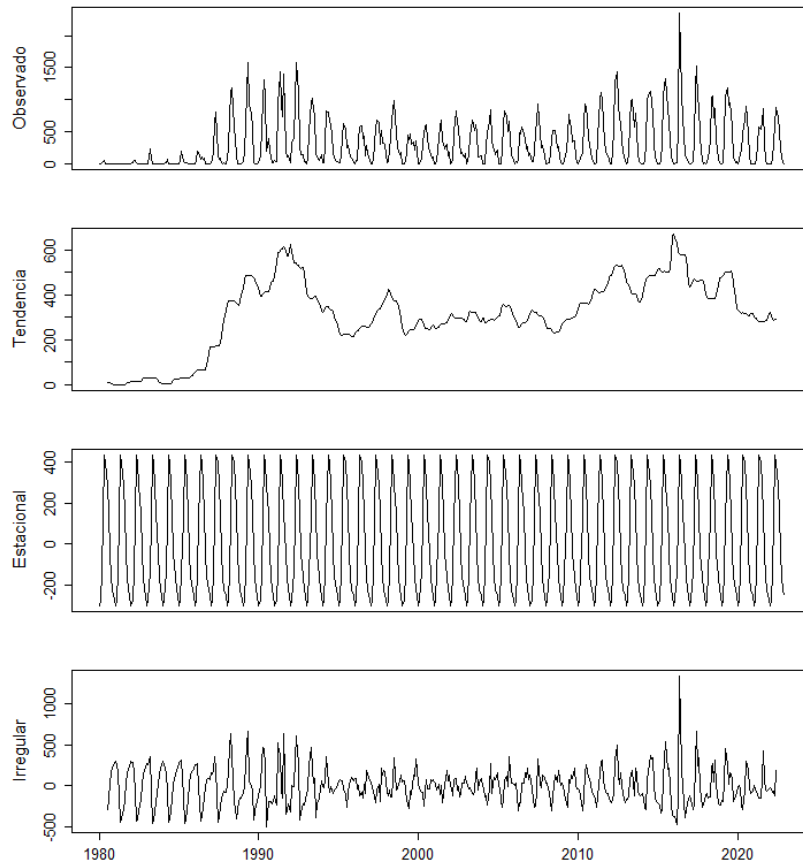
##### Autocorrelación y Pruebas Ljung-Box para pez espada, tiburón y dorado de altura

Las **Figura 11**, **Figura 13** y **Figura 15** muestran las ACF y PACF de los desembarques de pez espada, tiburones y dorado de altura, considerando un desfase de 3 años. Los resultados del Test de Ljung-Box (**Tabla 5**) indican que todas las especies presentan autocorrelación significativa en los desembarques ( $p$ -valor  $< 0.05$ ), lo que confirma que los valores pasados influyen en los futuros.

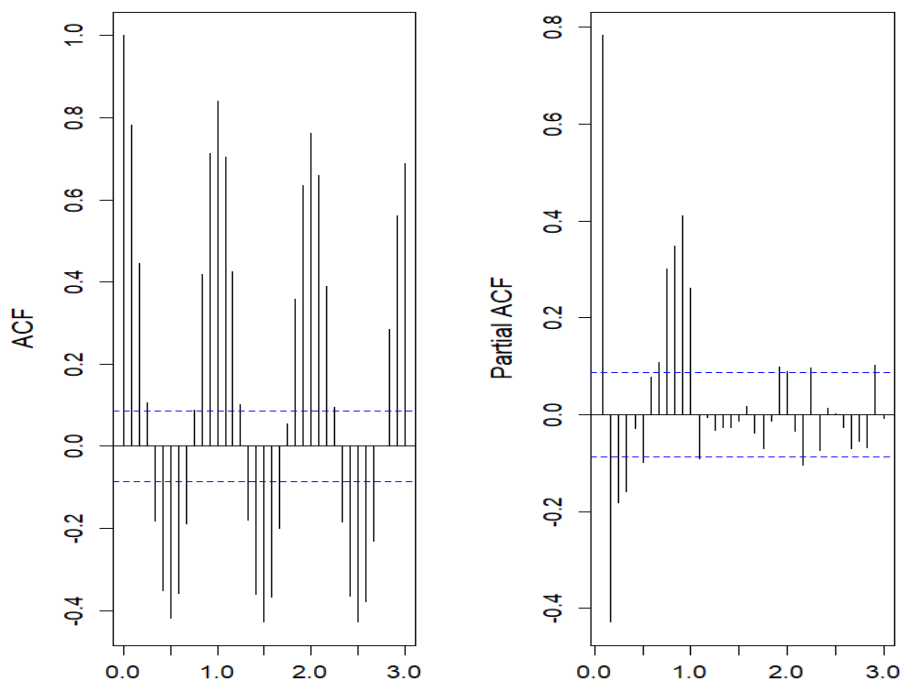
El PACF para DESPA muestra retardos significativos en T-1, T-2, T-3, T-4, T-6, T-8, T-9, T-10, T-11 y T-12, mientras que para DTIBU se observan en T-1, T-2, T-3 y T-7. En el caso de DDORA, los retardos identificados corresponden a T-1, T-8, T-9, T-10 y T-11. Esta información es fundamental para la construcción de los modelos *Random Forest* y *XGBoost*, ya que permite incorporar la estructura temporal de los desembarques y mejorar la capacidad predictiva. La presencia de autocorrelación significativa refuerza la necesidad de emplear enfoques avanzados que capturen relaciones no lineales entre las capturas y las variables ambientales

**Tabla 5. a.** Resultado del test Ljung-Box para VARIABLES PESQUERA DE INTERES

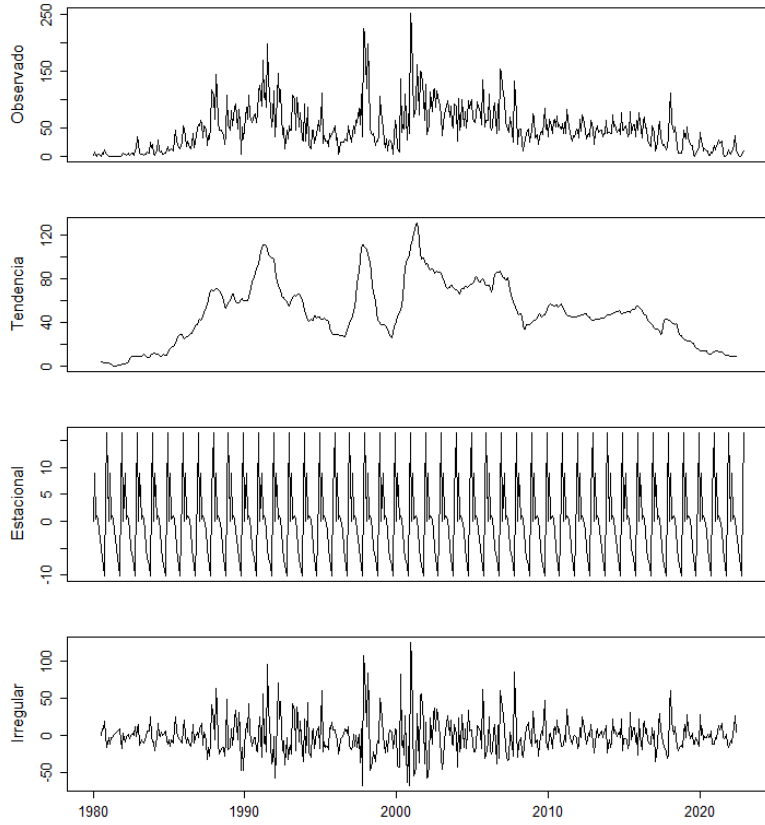
Especies	$\chi^2$	df	Valor-p
Pez espada	2073.3	20	$< 2.2e-16$
Tiburones	1646.4	20	$< 2.2e-16$
Dorado de altura	1191.1	20	$< 2.2e-16$



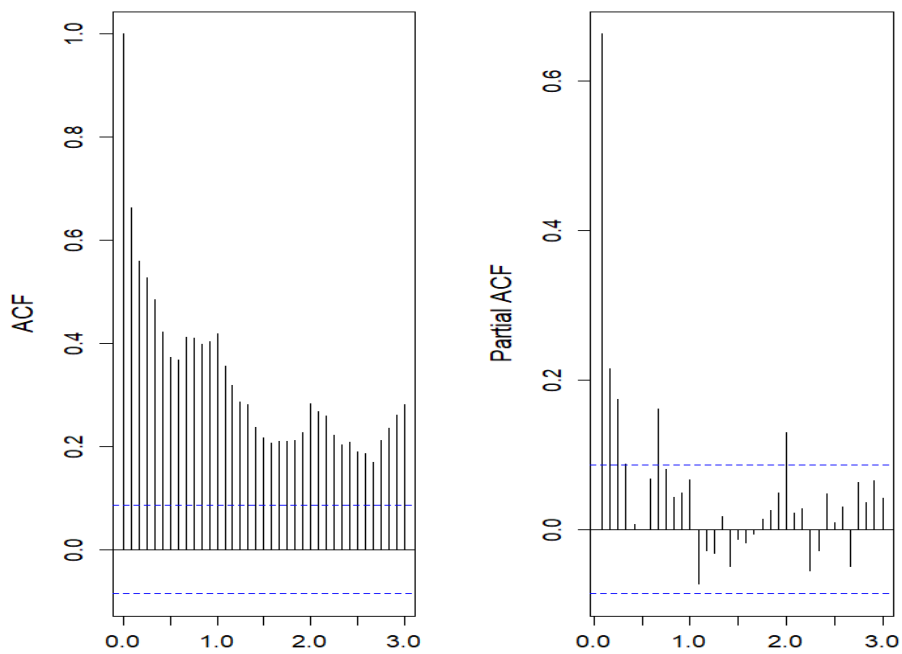
**Figura 10.** Descomposición del desembarque de pez espada para la ACF y PACF.



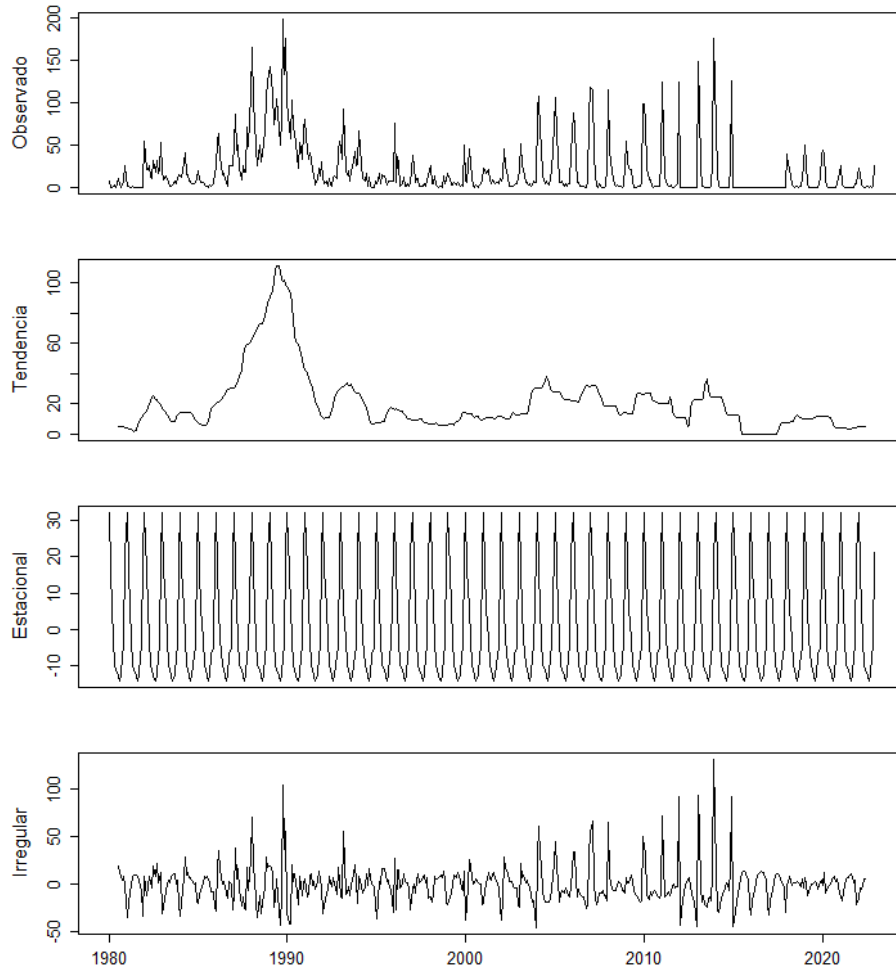
**Figura 11.** Autocorrelación d del desembarque de pez espada para la ACF y PACF. Considerando un desfase de 3 años.



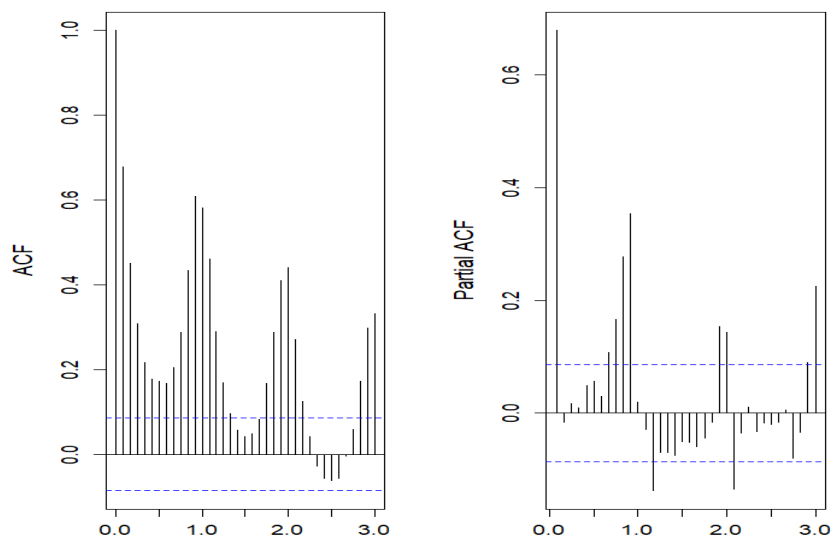
**Figura 12.** Descomposición del desembarque de tiburones para la ACF y PACF.



**Figura 13.** Autocorrelación d del desembarque de tiburones para la ACF y PACF. Considerando un desfase de 3 años.



**Figura 14.** Descomposición del desembarque de dorado de altura para la ACF y PACF.



**Figura 15.** Autocorrelación d del desembarque de dorado de altura para la ACF y PACF. Considerando un desfase de 3 años.

### Descomposición de las series temporales de las componentes principales.

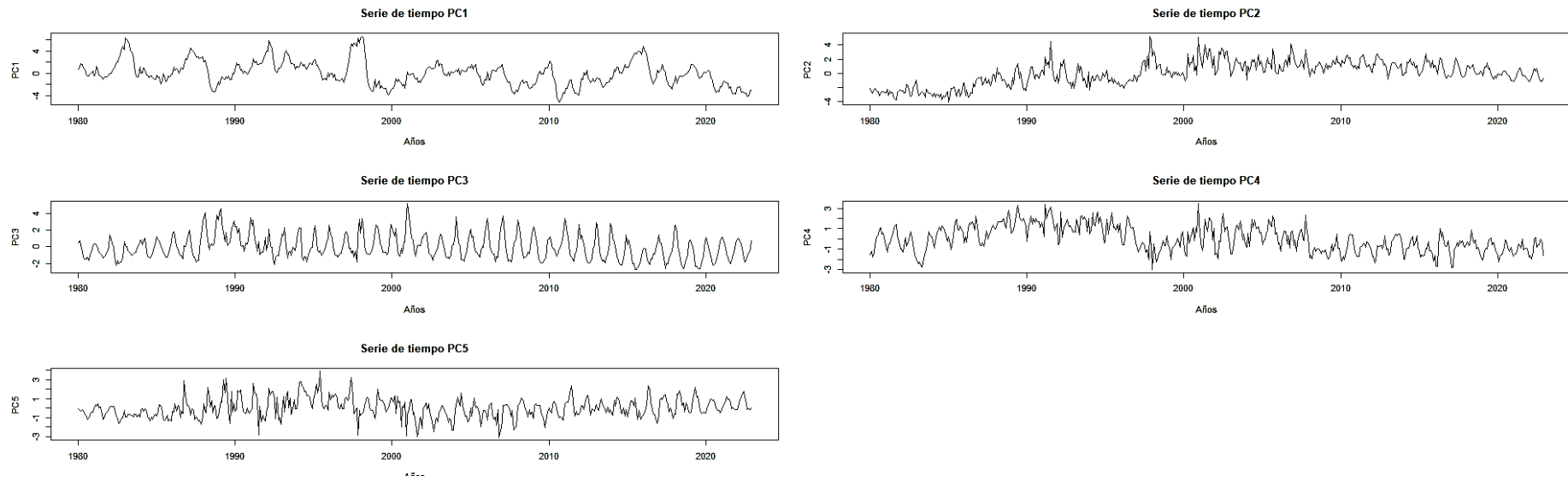
El análisis de las series temporales de las cinco primeras Componentes Principales (PC1 a PC5; **Figura 16**) permitió identificar patrones diferenciados en términos de tendencia, estacionalidad y aleatoriedad. Las gráficas muestran que PC1 y PC2 presentan fluctuaciones a largo plazo con fases de crecimiento y decrecimiento, lo que sugiere la influencia de eventos climáticos a gran escala sobre estas componentes. Por otro lado, PC3 y PC4 exhiben una marcada estacionalidad, con picos regulares que indican ciclos anuales o semestrales, reflejando posibles variaciones en respuesta a factores ambientales periódicos. La serie temporal de PC5 muestra menor varianza, pero mantiene cierto nivel de estacionalidad, aunque con fluctuaciones más suaves.

El análisis de autocorrelación (ACF) y autocorrelación parcial (PACF) mostró que las componentes principales presentan correlación significativa a distintos retardos, especialmente en PC3 y PC4, donde se observa una fuerte autocorrelación durante varios períodos (**Figura 17**). El test de Box-Ljung (**Tabla 6**) confirmó la presencia de autocorrelación significativa en todas las componentes ( $p < 2.2e-16$ ), lo que indica que las series no son independientes y presentan una estructura de dependencia temporal. El test de Box-Ljung, realizado para evaluar la independencia de las series temporales a 20 retardos indica que todas las series presentan autocorrelación significativa, rechazando la hipótesis nula de independencia. Esto podría reflejar la influencia de factores ambientales persistentes o interacciones entre variables pesqueras y climáticas a lo largo del tiempo

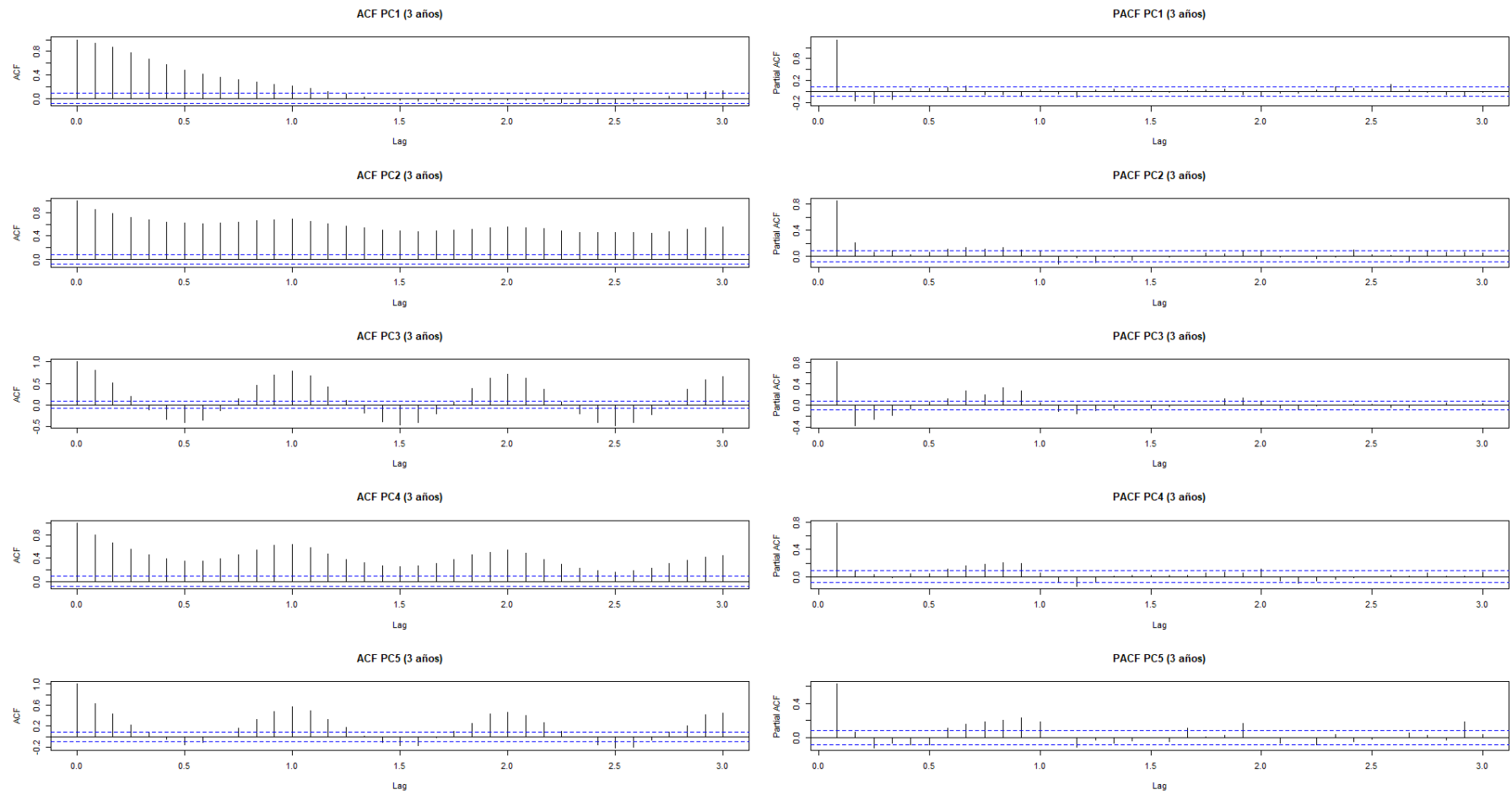
**Tabla 6.** Resultado del test Ljung-Box para los componentes de PC1, PC2, PC3, PC4 y PC5.

Componentes principales	$\chi^2$	df	Valor-p
PC1	2037.1	20	< 2.2e-16
PC2	4223	20	< 2.2e-16
PC3	2099.5	20	< 2.2e-16
PC4	2368	20	< 2.2e-16
PC5	944.65	20	< 2.2e-16

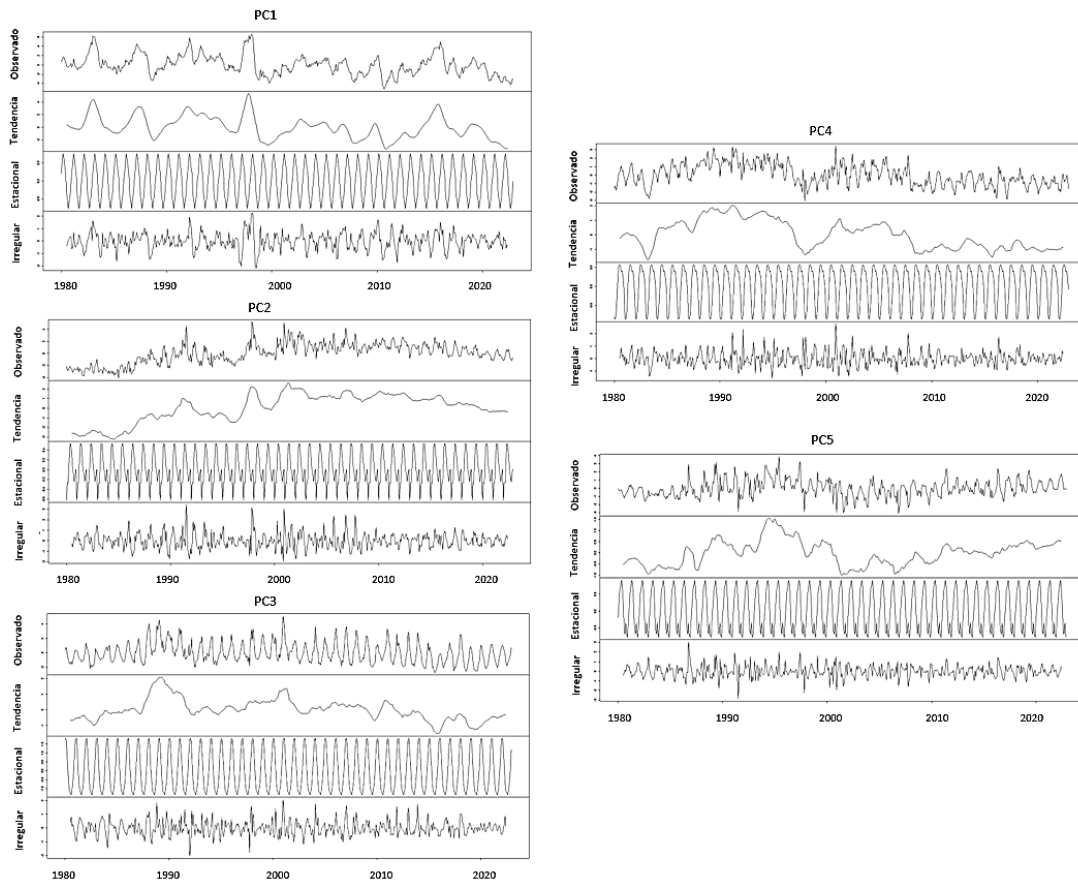
La descomposición aditiva de las series temporales (**Figura 18**), reveló que las tendencias de PC1 y PC2 reflejan cambios a largo plazo, mientras que PC3, PC4 y PC5 presentan tendencias más suaves con algunos cambios notorios hacia finales de los años 2000. La estacionalidad es particularmente evidente en PC3 y PC4, mostrando patrones cíclicos repetidos cada año. El componente aleatorio o residual es más destacado en PC1 y PC2, lo que indica que una proporción significativa de la variabilidad en estas componentes no puede ser explicada únicamente por la tendencia o estacionalidad, sugiriendo la presencia de factores exógenos.



**Figura 16.** Serie de tiempo de las 5 primeras componentes principales seleccionadas para el ACF y PACF.



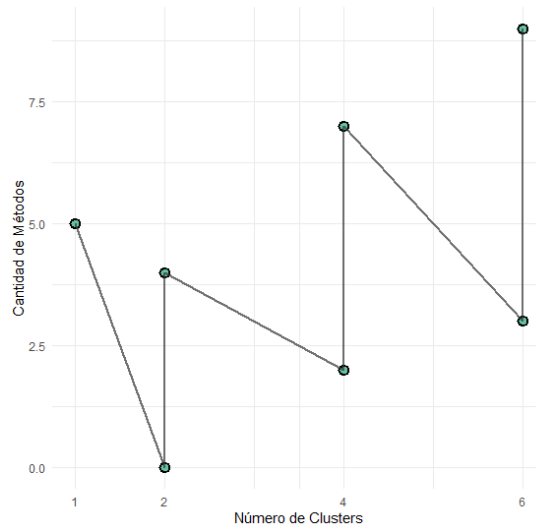
**Figura 17.** Autocorrelación de las primeras 5 componentes principales seleccionadas para la ACF y PACF. Considerando un desfase de 3 años.



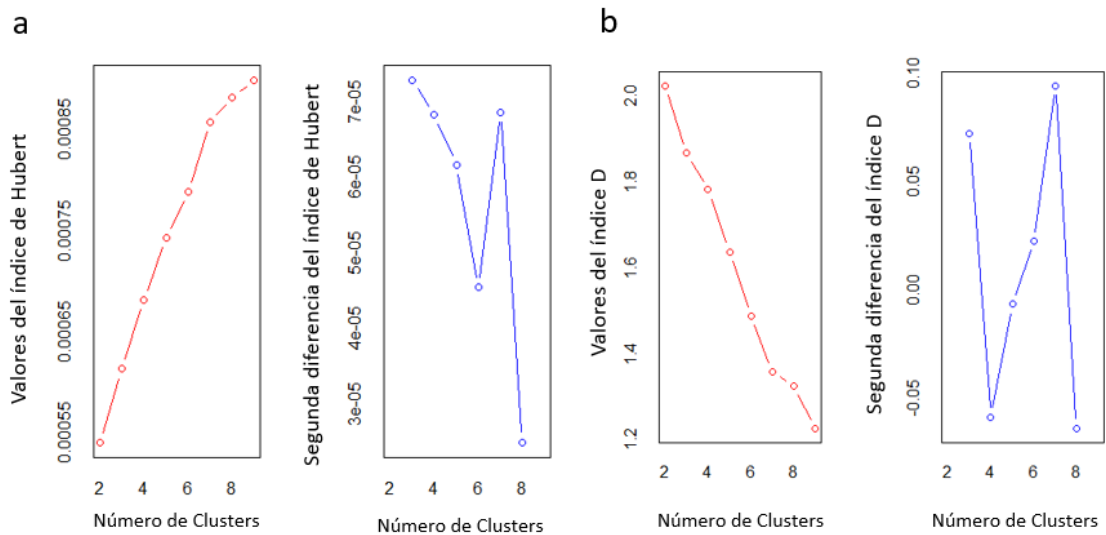
**Figura 18.** Descomposición de las primeras 5 componentes seleccionadas para la ACF y PACF.

#### 4.5 Análisis de Agrupamiento

La selección del número óptimo de clústers se realizó utilizando el paquete *NbClust*, el cual aplicó diversos criterios de validación. Según la regla de la mayoría, tres clústers fueron identificados como la mejor opción, siendo respaldados por seis métodos distintos (**Figura 19**). Adicionalmente, la evaluación gráfica de los índices Hubert y D (Dunn Index) reveló picos en sus diferencias de segundo orden para tres y cuatro clústers (**Figura 20 a y b**), lo que sugiere que ambas configuraciones son estructuralmente relevantes. Aunque algunos métodos propusieron valores alternativos (2, 4, 7 y 9 clústers), la convergencia de múltiples criterios refuerza la elección de tres clústers como la solución más robusta para la segmentación de los datos.



**Figura 19.** Distribución de métodos según el número óptimo de clústers.

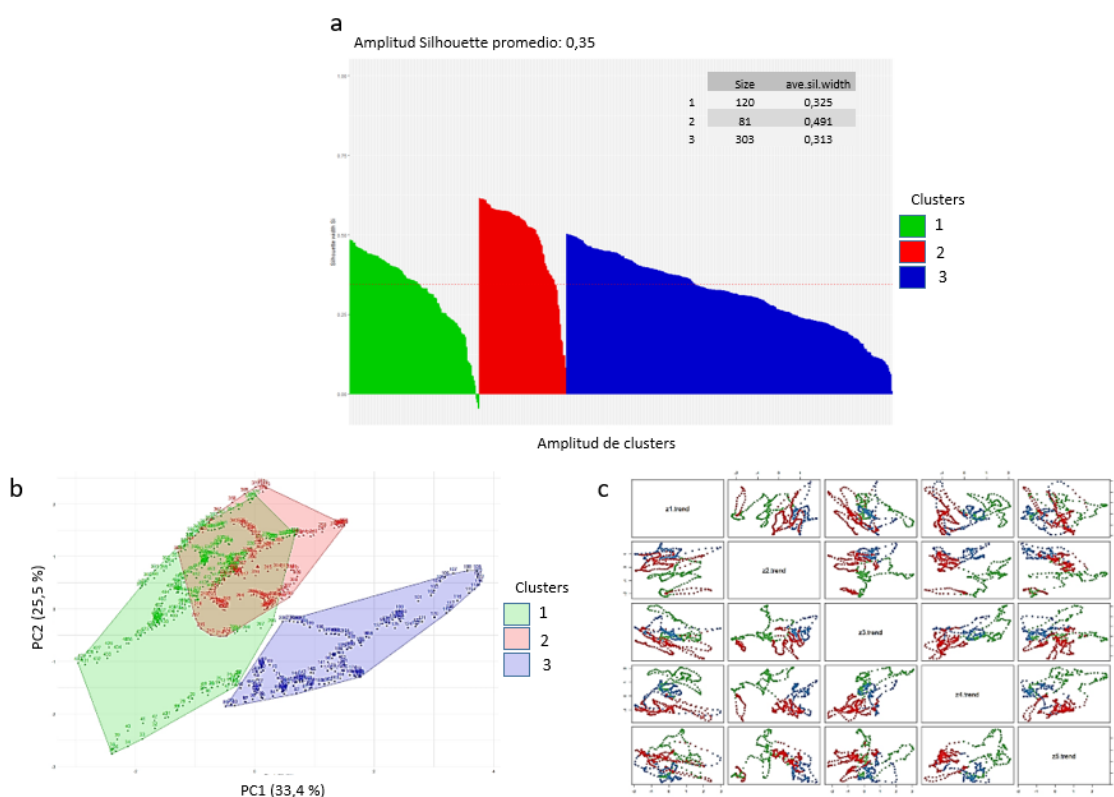


**Figura 20.** a Evaluación del número óptimo de clústers mediante el índice de Hubert y b. Evaluación del número óptimo de clústers mediante el índice D.

El gráfico de silueta (**Figura 21a**) muestra un índice promedio de amplitud de silueta de 0,35, lo que indica una calidad de agrupamiento moderada. Se identifican tres clústers, donde el clúster 2 presenta la mayor cohesión con un valor promedio de 0,49, seguido del clúster 1 con 0,33, mientras que el clúster 3 exhibe un valor más bajo de 0,31, lo que sugiere una menor separación y mayor variabilidad interna.

En el gráfico de distribución de clústers (**Figura 21b**), los clústers se distribuyen sobre las dos primeras dimensiones principales (Dim1: 33,4 % y Dim2: 25 %). Los Clústers 1 y 3 muestran una mayor diferenciación espacial, mientras que el clúster 2 presenta mayor cohesión, lo que coincide con su menor valor de silueta.

Las trayectorias de los componentes principales (**Figura 21c**) reflejan patrones dinámicos específicos para cada clúster. El clúster 2 mantiene trayectorias más compactas y organizadas, mientras que los clústeres 1 y 3 presentan una mayor dispersión y variabilidad en sus tendencias, lo que sugiere una dinámica más heterogénea a lo largo del tiempo.



**Figura 21.** **a.** índice de amplitud de Agrupamiento, **b.** representación de los centroides de cada clúster y **c.** clasificación de los componentes principales. El color verde corresponde al clúster 1, el rojo al clúster 2 y azul al clúster 3.

El análisis general de la superposición de las series temporales de las cinco primeras componentes principales (PC1-PC5, **Figura 22**) muestra patrones diferenciados que destacan tanto tendencias de largo plazo como fluctuaciones estacionales. Se identifican cuatro periodos principales: 1980-1983 (rojo), 1983-1996 (verde), 1996-2008 (azul-rojo) y 2008-2022 (rojo-azul). Es notable que el color rojo del clúster predomina tanto al inicio como al final de la serie, sugiriendo similitudes en las condiciones durante esos periodos. Por otro lado, el tercer periodo presenta una mayor variabilidad entre los clústeres, reflejando una dinámica más heterogénea en estos intervalos.

Al analizar la superposición de las series temporales de los desembarques (**Figura 23**), considerando los mismos cuatro periodos, se observa que en el primer periodo los desembarques son generalmente bajos, con la excepción de la sardina, que presenta capturas más significativas. Durante el segundo periodo, se registra un aumento generalizado en los desembarques de todas las especies. El tercer periodo se caracteriza por fluctuaciones marcadas en las capturas, evidenciando una mayor variabilidad en comparación con los demás. Finalmente, el cuarto periodo muestra una tendencia hacia la estabilidad, con desembarques relativamente constantes en todas las especies y al final del periodo se ve una disminución de las capturas.

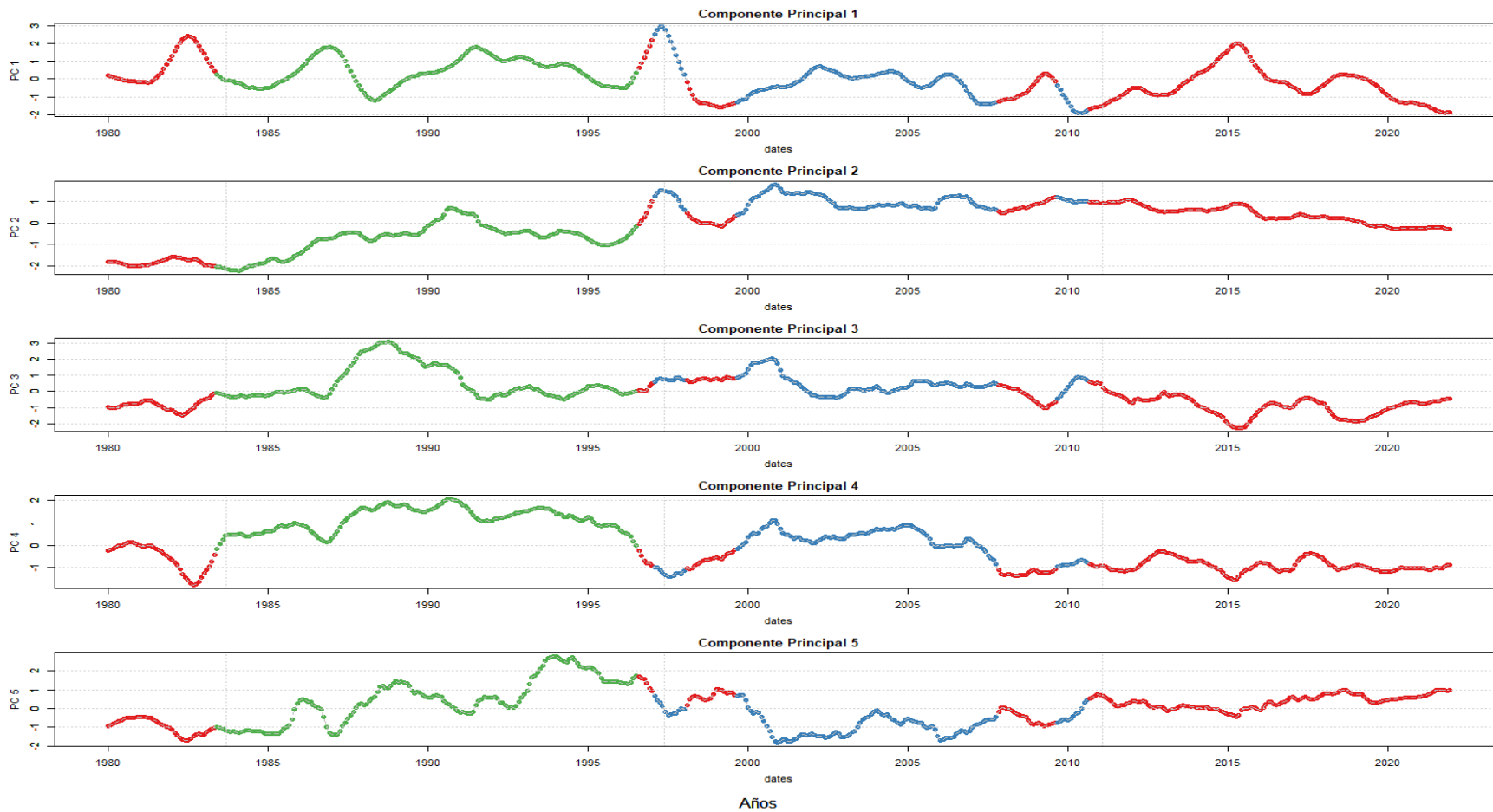
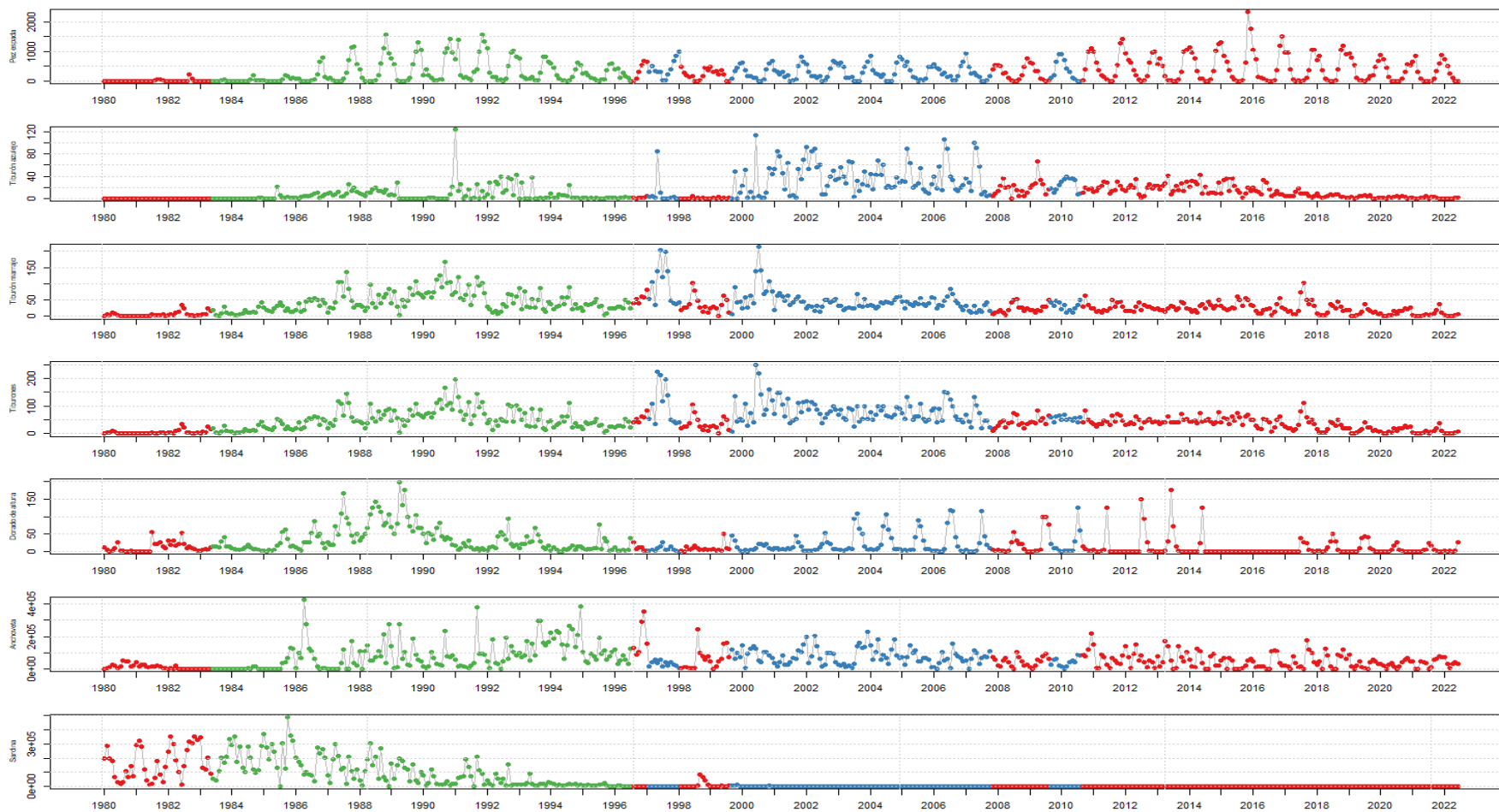


Figura 22. Serie de tiempo de la clasificación de los componentes principales.



**Figura 23.** Clasificación de la variable pesqueras (Desembarques de especies) **a.** pez espada, **b.** dorado de altura, **c.** tiburón azulejo, **d.** tiburón marrajo, **e.** tiburones, **f.** anchoveta y **g.** sardina

### **Análisis de Agrupamiento para todos los datos por especie de interés**

Se analizaron todas las variables ambientales y pesqueras, excluyendo de forma individual al pez espada, los tiburones y el dorado de altura. En esta ocasión, los tiburones fueron agrupados debido a que los resultados obtenidos por separado mostraban patrones muy similares, por lo que se optó por consolidar la información y evitar redundancias. Posteriormente, se evaluó el comportamiento de las variables excluyendo la variable de interés, con el fin de evitar la colinealidad en la modelación. Para ello, se exploraron las series de tiempo de los PCA y desembarques, visualizando la distribución de los clústeres a lo largo de la serie histórica.

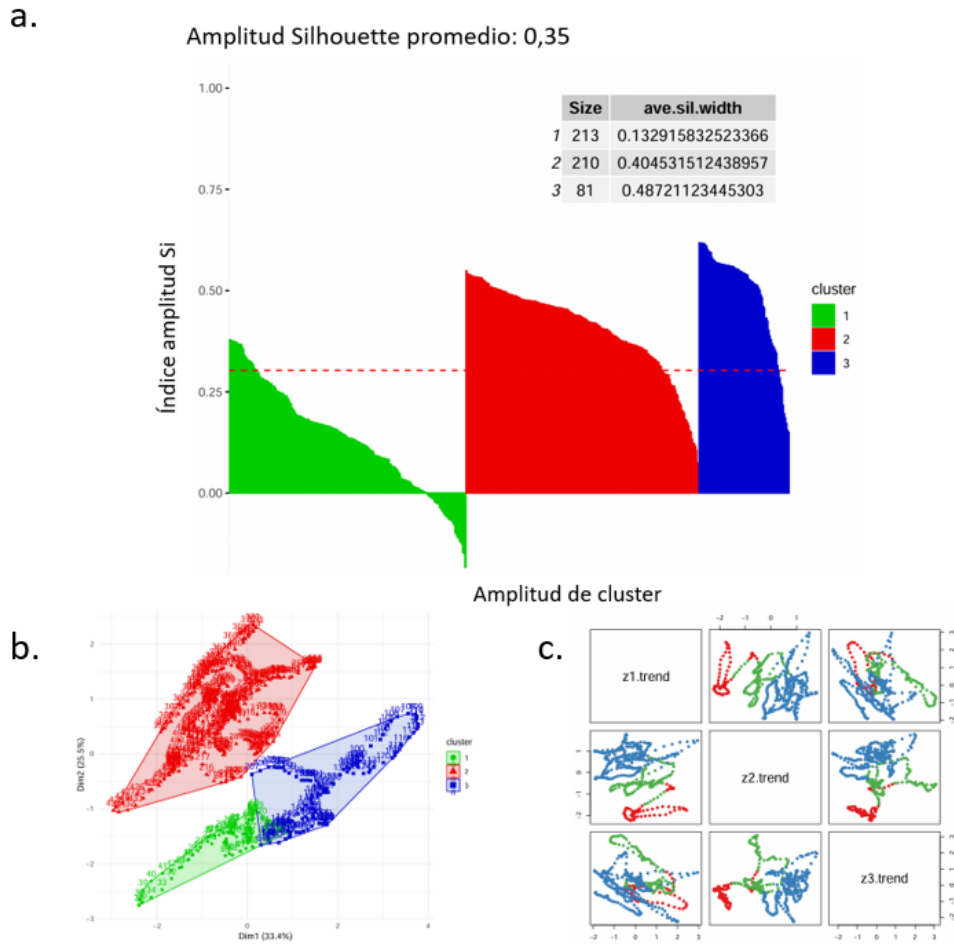
### **Análisis de Agrupamiento para pez espada**

Los resultados muestran que, según la frecuencia de los índices, el número óptimo de clústeres corresponde a 3, de acuerdo con la regla de la mayoría.

El análisis de clústeres, representado en la **Figura 24 a**, muestra un índice promedio de amplitud de silueta de 0,35, lo que indica una calidad de agrupamiento moderada. Se identificaron tres clústeres principales, con tamaños de 213, 210 y 81 elementos, correspondientes a los clústeres 1, 2 y 3, respectivamente. El clúster 3 presenta la mayor cohesión, con una amplitud de silueta promedio de 0,48, seguido del clúster 2 con 0,40, mientras que el clúster 1 presenta el menor valor (0,13), lo que refleja una mayor variabilidad interna y una separación menos definida respecto a los otros grupos.

En el gráfico tridimensional de representación de clústeres (**Figura 24 b**), se observa que los clústeres están distribuidos a lo largo de las dos primeras dimensiones principales (Dim1: 33,4% y Dim2: 25,2%). El clúster 2 muestra una distribución más compacta, indicando mayor cohesión dentro del grupo, mientras que el clúster 1 presenta una mayor dispersión, reflejando menor cohesión y una variabilidad más amplia. El clúster 3 muestra una distribución intermedia, destacándose por su clara separación de los otros grupos.

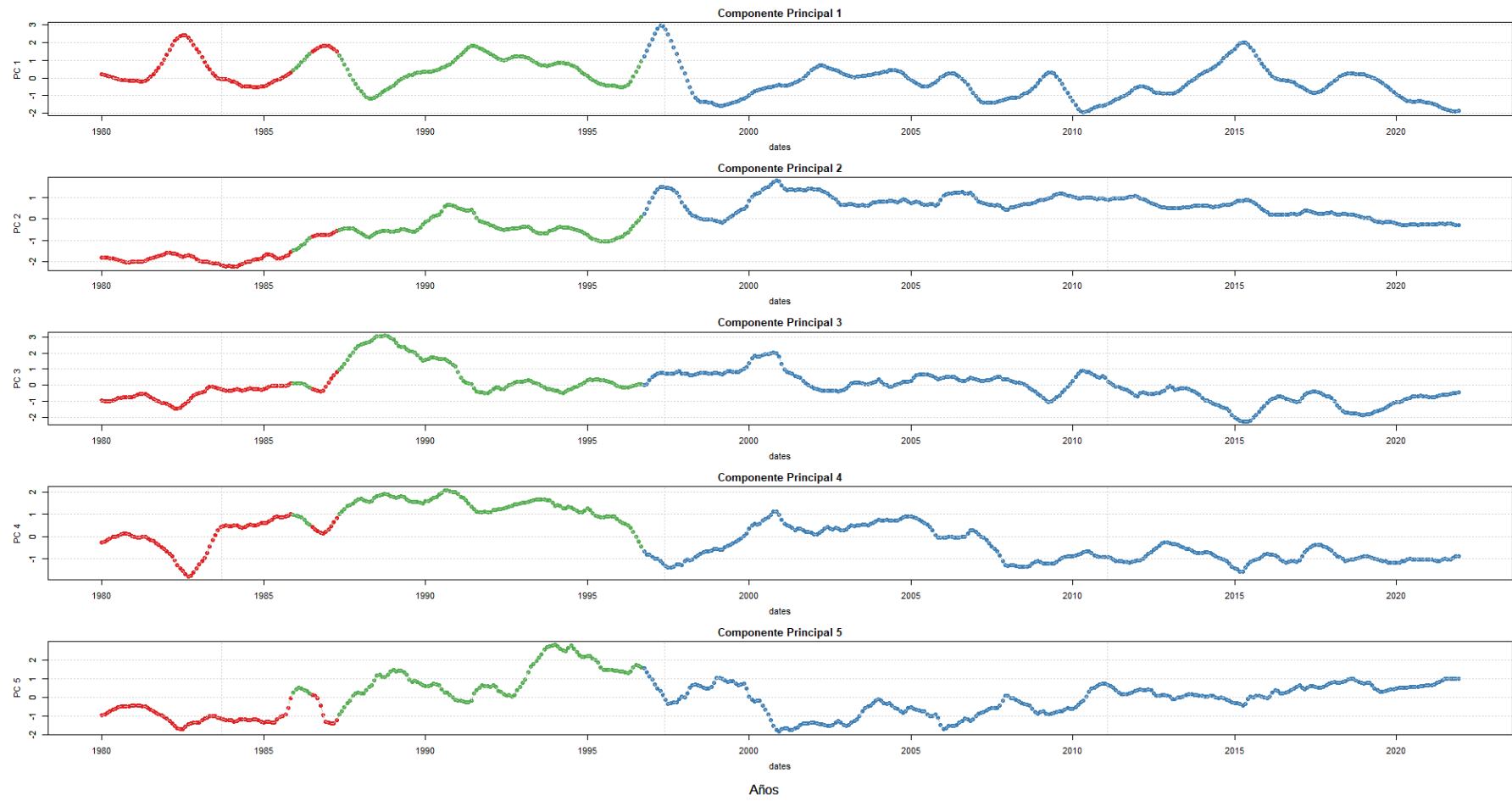
En cuanto a las trayectorias temporales de las componentes principales (**Figura 24 c**), los patrones dinámicos muestran diferencias claras entre los clústeres. El clúster 2 presenta trayectorias más organizadas y compactas, mientras que los clústeres 1 y 3 exhiben mayor dispersión y comportamientos heterogéneos. En particular, el clúster 3 refleja una menor variabilidad, mientras que el clúster 1 presenta mayores oscilaciones en sus trayectorias, sugiriendo dinámicas más complejas y menos estables a lo largo del tiempo.



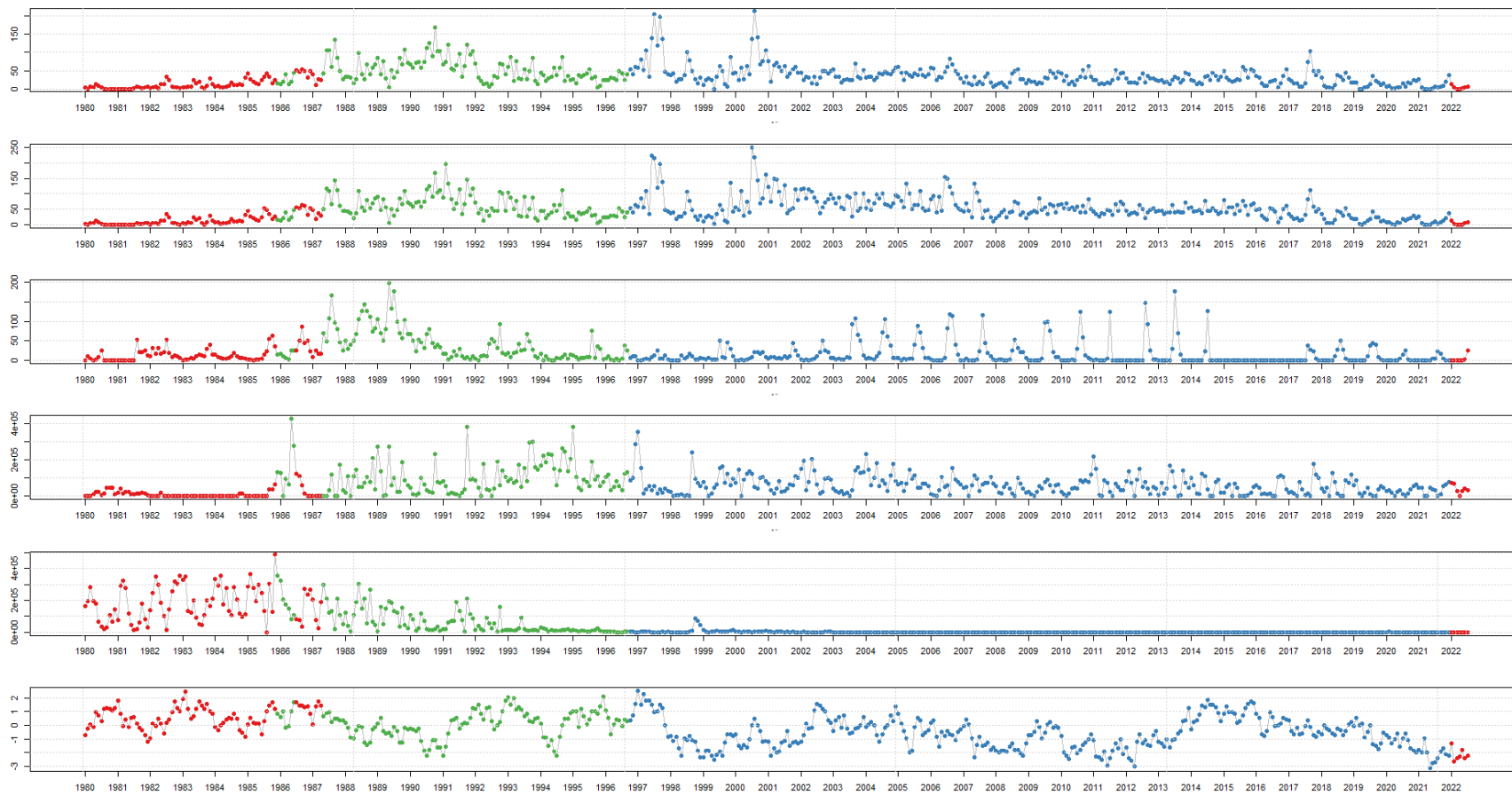
**Figura 24. a.** Índice de amplitud del Agrupamiento, **b.** representación de los centroides del clúster y **c.** clasificación de los componentes principales. El color verde corresponde al clúster 1, el rojo al clúster 2 y azul al clúster 3.

El análisis general de la superposición de las series temporales de las cinco primeras componentes principales (PC1-PC5, **Figura 25**) muestra patrones diferenciados que destacan tanto tendencias de largo plazo como fluctuaciones estacionales. Se identifican tres periodos principales: 1980-1986, 1986-1996 y 1996 - 2022.

Al analizar la superposición de las series temporales de los desembarques (**Figura 26**), las series muestran un crecimiento abrupto entre 1986 y 1990, seguido por una tendencia de alta variabilidad hasta finales de los años 90, con picos notorios. A partir del año 2000, los desembarques disminuyen y mantienen valores bajos con pequeños repuntes en los años recientes.



**Figura 25.** Serie de tiempo de la clasificación de los componentes principales sin considerar los desembarques de pez espada.



**Figura 26.** Clasificación de la variable pesqueras (desembarques de especies) **a.** dorado de altura, **b.** tiburón azulajejo, **c.** tiburón marrajo, **d.** tiburones, **e.** sardina y **f.** anchoveta.

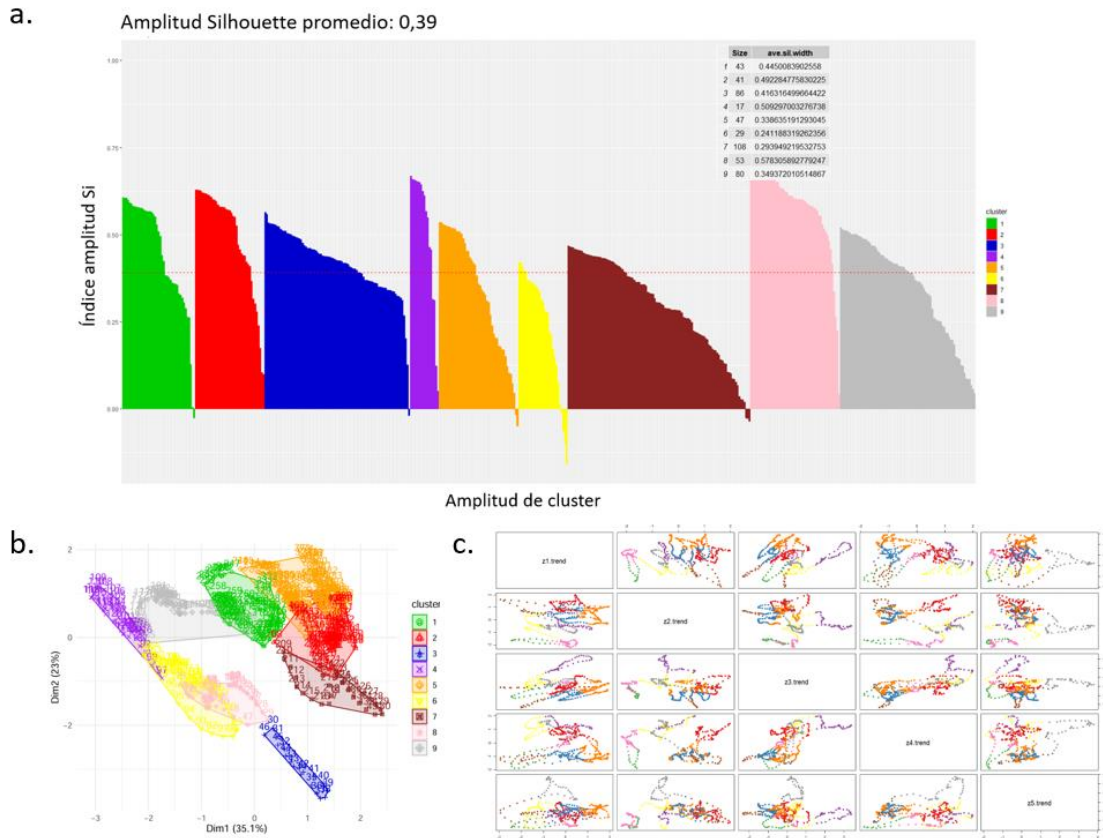
## **Análisis de Agrupamiento para tiburones**

Los resultados muestran que, según la frecuencia de los índices, el número óptimo de clústeres corresponde a 9, de acuerdo con la regla de la mayoría.

El análisis de clústeres (**Figura 27 a**) muestra un índice promedio de amplitud de silueta de 0,39, indicando una calidad de agrupamiento moderada. Se identificaron nueve clústeres con tamaños variados, donde el clúster 8 (rosado) presenta la mayor cohesión con una amplitud promedio de silueta de 0,57, seguido por el clúster 4 (morado) con 0,50. En contraste, los clústeres 6 (amarillo) y 7 (café) tienen las amplitudes de silueta más bajas, de 0,24 y 0,29 respectivamente, lo que refleja una menor cohesión interna y una mayor superposición con otros clústeres. Este comportamiento sugiere una heterogeneidad significativa en los patrones de variabilidad dentro de estos grupos.

En la representación tridimensional de los clústeres (**Figura 27 b**), los grupos se distribuyen a lo largo de las dos primeras dimensiones principales (Dim1: 35,1% y Dim2: 23,0%). Los clústeres 1 (verde), 2 (rojo) y 5 (café claro) están más compactos, indicando mayor cohesión interna y separación respecto a otros clústeres. Por el contrario, los clústeres 6 (morado) y 9 (gris) muestran una mayor dispersión, reflejando menor cohesión y una variabilidad más alta. Esto sugiere que los primeros grupos capturan patrones más definidos, mientras que los últimos representan dinámicas más complejas.

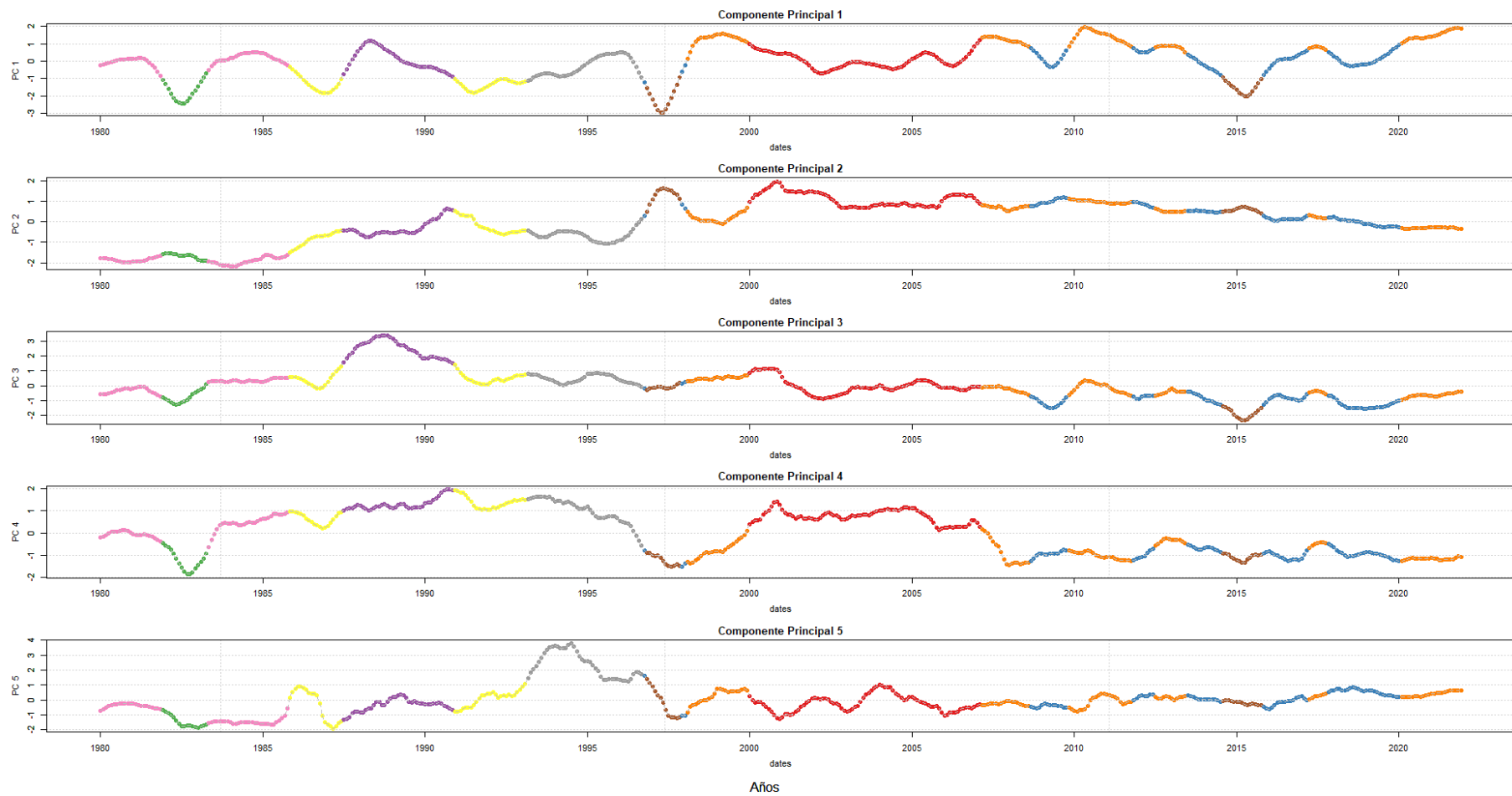
En las trayectorias temporales de las componentes principales (**Figura 27 c**), los patrones dinámicos destacan diferencias claras entre clústeres. Los clústeres 2 y 3 presentan trayectorias más organizadas y compactas, mientras que los clústeres 6 y 9 muestran mayor dispersión y oscilaciones marcadas. En particular, las trayectorias de los clústeres 6 (amarillo) y 8 (rosado) sugieren una transición gradual entre patrones dinámicos, al contrario que los clústeres 1 (verde) y 7 (café).



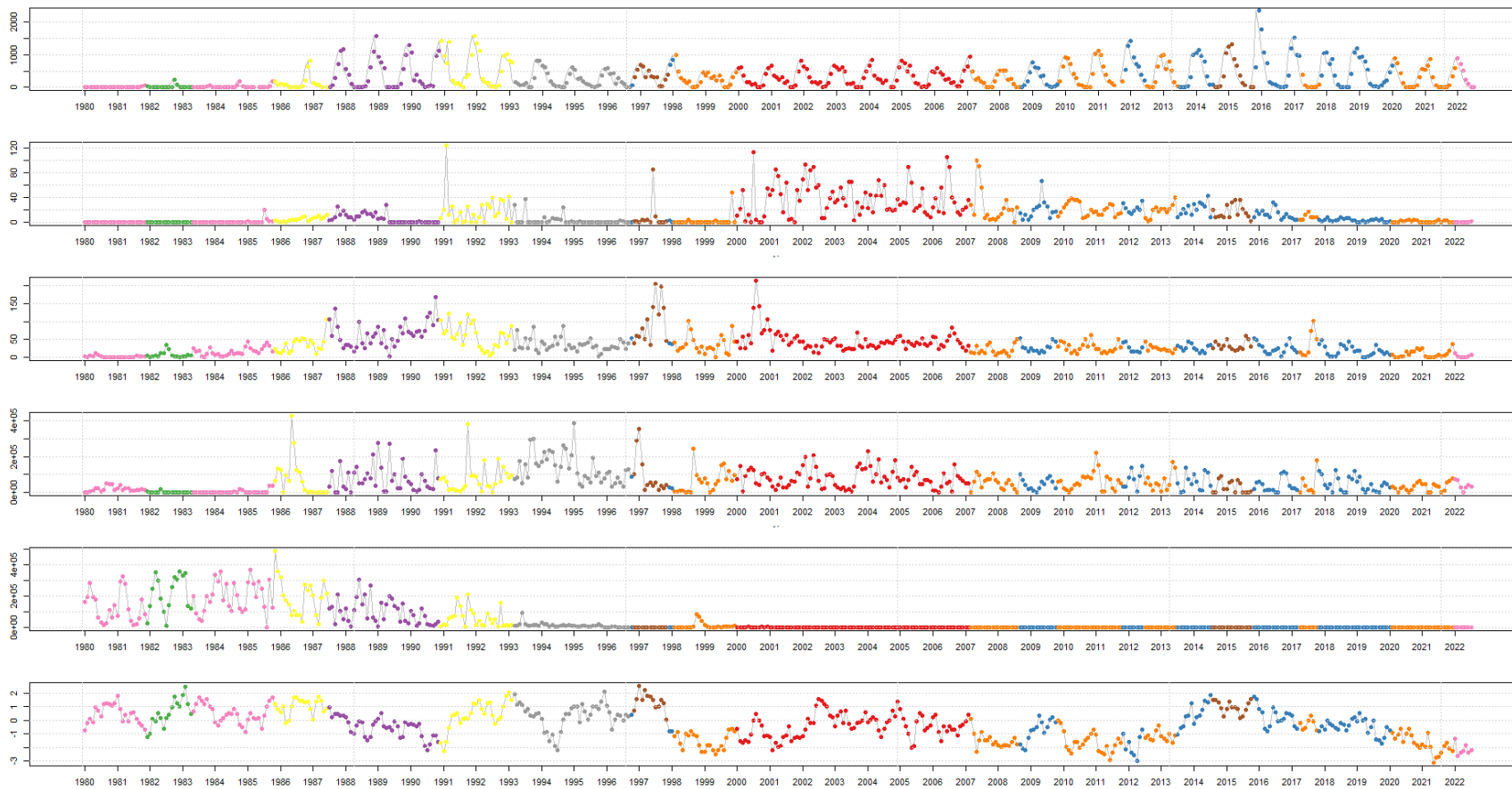
**Figura 27. a.** Índice de amplitud del Agrupamiento, **b.** representación de los centroides del clúster y **c.** clasificación de los componentes principales. El color verde corresponde al clúster 1, el rojo al clúster 2 y azul al clúster 3.

El análisis general de la superposición de las series temporales de las cinco primeras componentes principales (PC1-PC5, **Figura 28**) revela patrones temporales con variaciones notables a lo largo del tiempo. Antes del año 2000, se observa una mayor fluctuación y variabilidad en los clústeres, mientras que a partir de esta fecha se percibe una estabilización en las series, aunque persiste cierta alternancia en los clústeres dominantes.

Por otro lado, las series temporales de los desembarques pesqueros (**Figura 29**) reflejan una evolución clara en cuatro periodos principales. El primer periodo, de 1980 a 1985, está asociado a dos clústeres y predomina el de color rosado. El segundo periodo, de 1985 a 1997, se caracteriza por una mayor variabilidad en los clústeres, con alternancia notable del color amarillo. En el tercer periodo, de 1997 a 2008, se destaca un aumento en los desembarques, representado principalmente por el clúster rojo. Finalmente, el último periodo, de 2008 a 2022, muestra la presencia de cuatro clústeres, aunque con una alternancia predominante de los colores naranja y azul.



**Figura 28.** Serie de tiempo de la clasificación de los componentes principales sin consideras los desembarques de tiburones.



**Figura 29.** Clasificación de la variable pesqueras (desembarques de especies) **a.** pez espada, **b.** tiburón azulejo, **c.** tiburón marrajo, **d.** dorado de altura, **e.** sardina y **f.** anchoveta.

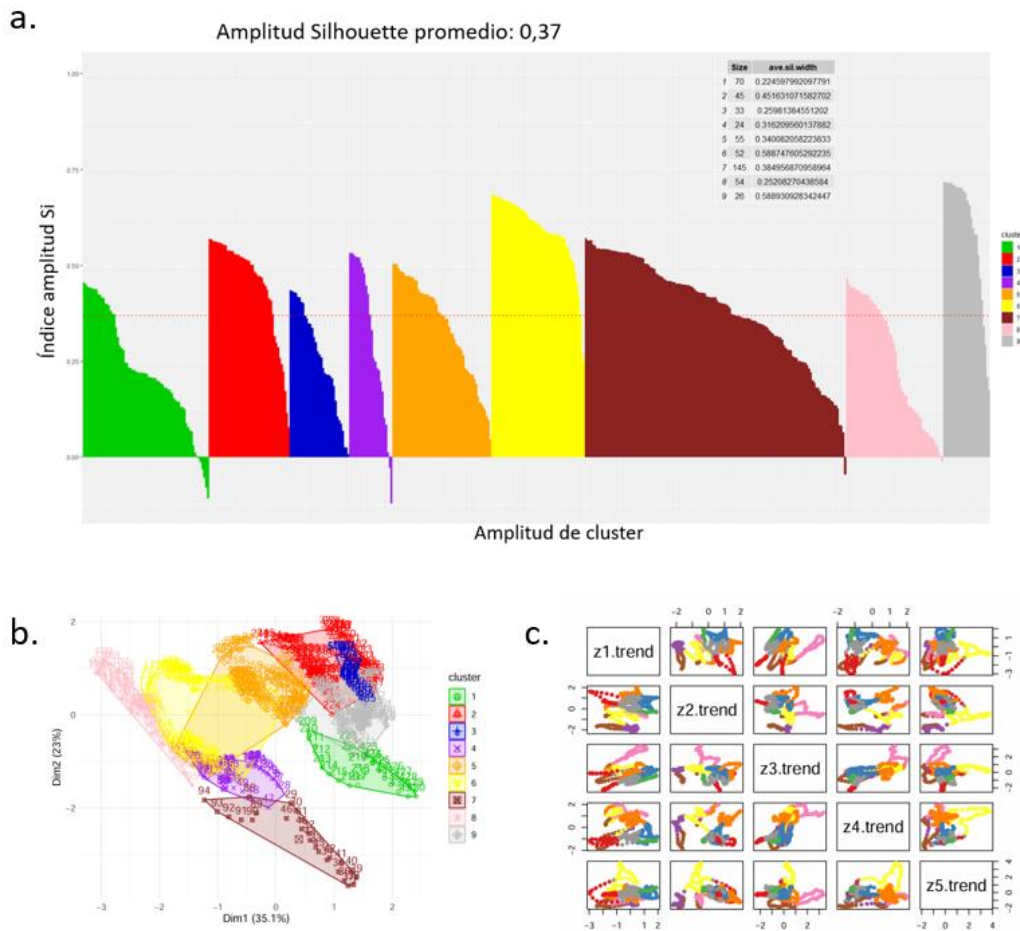
## **Análisis de Agrupamiento para dorado de altura**

Los resultados muestran que, según la frecuencia de los índices, el número óptimo de clústeres corresponde a 9, de acuerdo con la regla de la mayoría.

El análisis de clústeres (**Figura 30 a**) muestra un índice promedio de amplitud de silueta de 0,37, lo que indica una calidad de agrupamiento moderada. Se identificaron nueve clústeres, con tamaños que varían entre 24 y 145 elementos. El clúster 9 (gris) y 6 (amarillo) presenta la mayor cohesión, con una amplitud promedio de silueta de 0,58. Por el contrario, el clúster 1 (verde), 3 (azul) y 8 (rosado) tienen las menores amplitudes de silueta, de 0,22; 0,25 y 0,25 respectivamente, reflejando una menor cohesión interna y mayor solapamiento con otros clústeres. Estos valores sugieren que los patrones en algunos grupos están más definidos que en otros.

En la representación tridimensional de los clústeres (**Figura 30 b**), se observa que los grupos están distribuidos principalmente a lo largo de las dos primeras dimensiones principales (Dim1: 35,1% y Dim2: 23,0%). Los clústeres 2 (rojo) y 9 (gris) presentan una distribución más compacta, indicando una separación más clara entre grupos, mientras que los clústeres 5 (café claro) y 6 (amarillo) son más dispersos, reflejando mayor variabilidad interna. La disposición general de los clústeres resalta una estructura bien diferenciada en algunos casos y una interacción más compleja en otros.

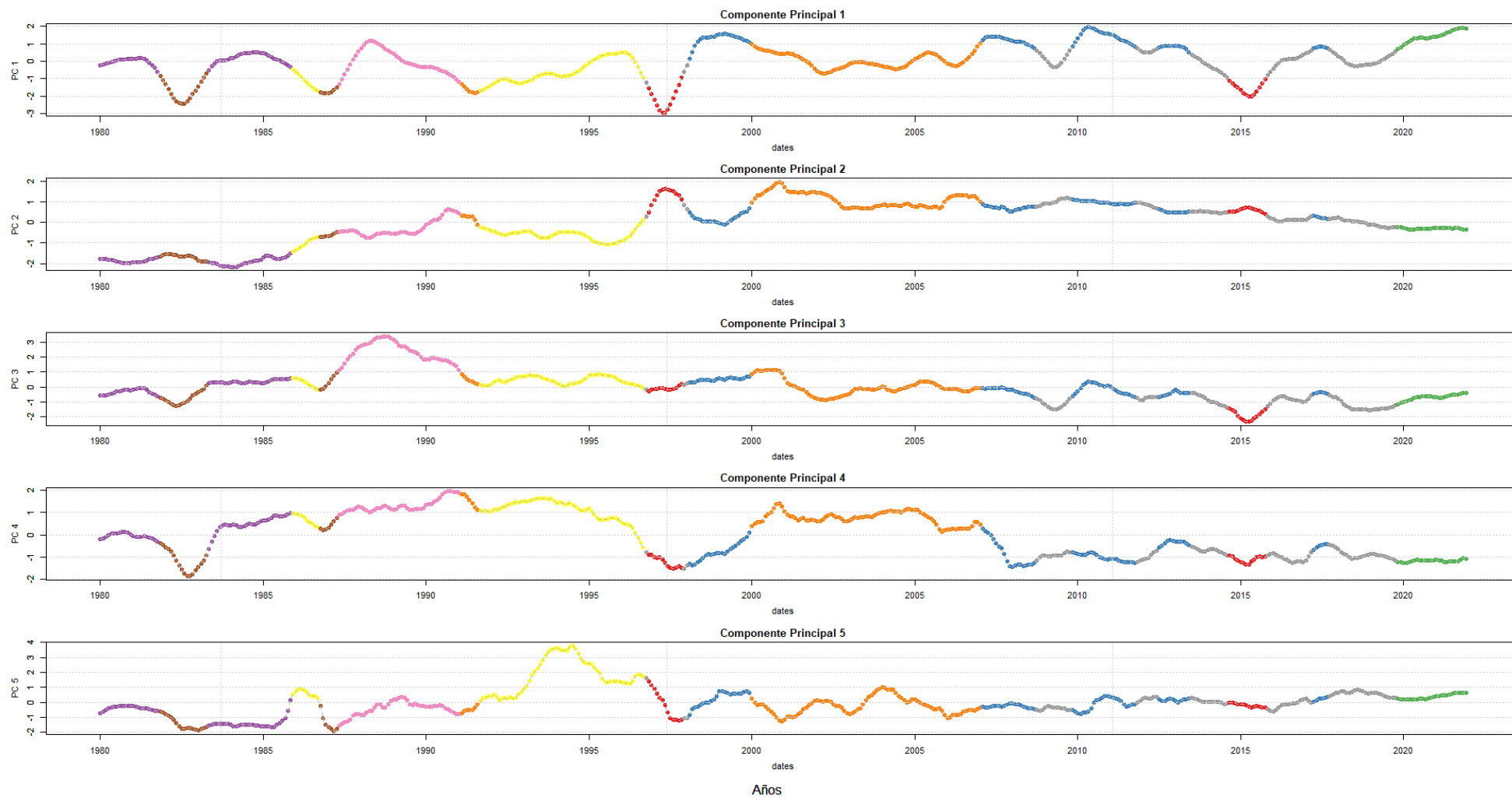
En las trayectorias de las componentes principales (**Figura 30 c**), se observan patrones dinámicos característicos para cada clúster. Los clústeres 7 (marrón) y 9 (gris) muestran trayectorias más organizadas y compactas. Las dinámicas de los clústeres 6 (amarillo) y 8 (rosado) reflejan transiciones entre diferentes patrones de variabilidad, mientras que los clústeres 5 (naranja) y 8 (morado) presentan trayectorias intermedias, con cierta definición, pero mayor dispersión en comparación con los clústeres más cohesivos.



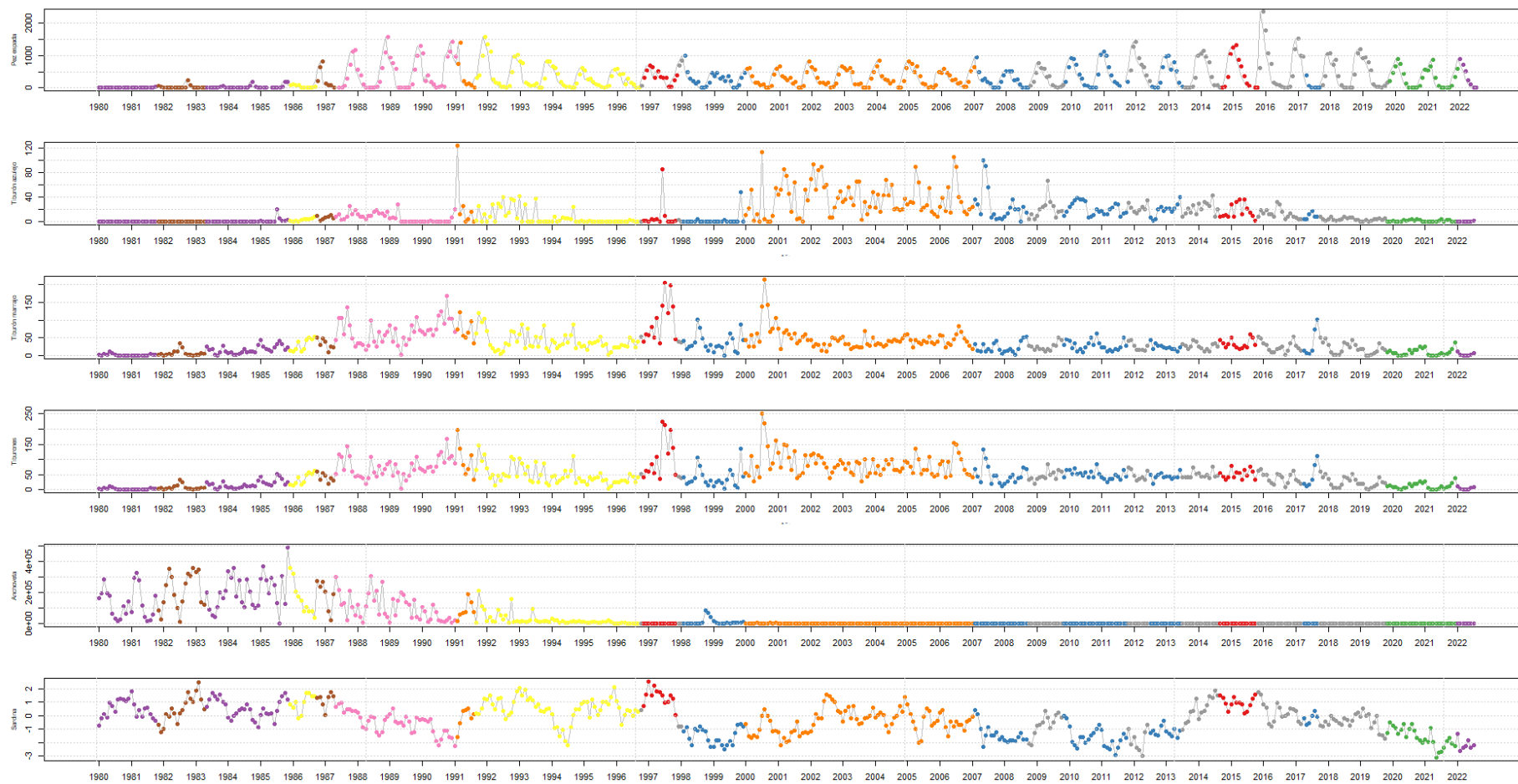
**Figura 30. a.** Índice de amplitud de Agrupamiento, **b.** representación de los centroides del clúster y **c.** clasificación de los componentes principales. El color verde corresponde al clúster 1, el rojo al clúster 2 y azul al clúster 3.

El análisis general de la superposición de las series temporales de las cinco primeras componentes principales (PC1-PC5, **Figura 31**) revela patrones temporales con variaciones significativas a lo largo del tiempo. Antes del año 2000, se observa una mayor fluctuación y variabilidad en los clústeres, mientras que a partir de esta fecha se aprecia una estabilización en las series. Sin embargo, persiste cierta alternancia en los clústeres dominantes, lo que sugiere dinámicas transitorias.

Por otro lado, las series temporales de los desembarques pesqueros (**Figura 32**) evidencian una evolución clara dividida en cuatro periodos principales. En el primer periodo (1980-1985), predominan dos clústeres, con el color morado como dominante. El segundo periodo (1985-1997) se caracteriza por una mayor variabilidad en los clústeres, con una alternancia notable del color amarillo y apariciones ocasionales de los colores café, naranja y plomo, lo que indica una transición entre diferentes dinámicas a largo plazo. En el tercer periodo (1997-2008), el clúster naranja se convierte en el predominante, aunque se observa alternancia con el clúster azul, indicando una fase de mayor actividad en los desembarques. Finalmente, el último periodo (2008-2022) muestra la presencia de cinco clústeres, con una alternancia predominante entre los colores plomo y azul, y hacia el final del periodo se destaca la presencia de los colores verde y morado, reflejando una posible nueva transición en las dinámicas de los desembarques.



**Figura 31.** Serie de tiempo de la clasificación de los componentes principales sin consideras los desembarques de dorado de altura.



**Figura 32.** Clasificación de la variable pesqueras (desembarques de especies) **a.** pez espada, **b.** tiburón azulejo, **c.** tiburón marrajo, **d.** tiburones, **e.** sardina y **f.** anchoveta.

#### 4.6 Implementación de modelo para desembarque de pez espada, tiburones y dorado de altura.

En este estudio se evaluó el desempeño de los modelos Random Forest y XGBoost para la predicción de los desembarques de pez espada (DESPA), tiburones (DTIBU) y dorado de altura (DDORA), considerando retardos temporales y variables ambientales. La **Tabla 7** muestra los valores de error cuadrático medio (MSE) y coeficiente de determinación ( $R^2$ ) para cada especie y modelo. En general, Random Forest presentó un mejor rendimiento en todas las especies, obteniendo valores de MSE menores y  $R^2$  superiores en comparación con XGBoost.

**Tabla 7.** Evaluación del rendimiento de los modelos para pez espada (DESPA), tiburones (DTIBU) y dorado de altura (DDORA) incluido como variables los retardos

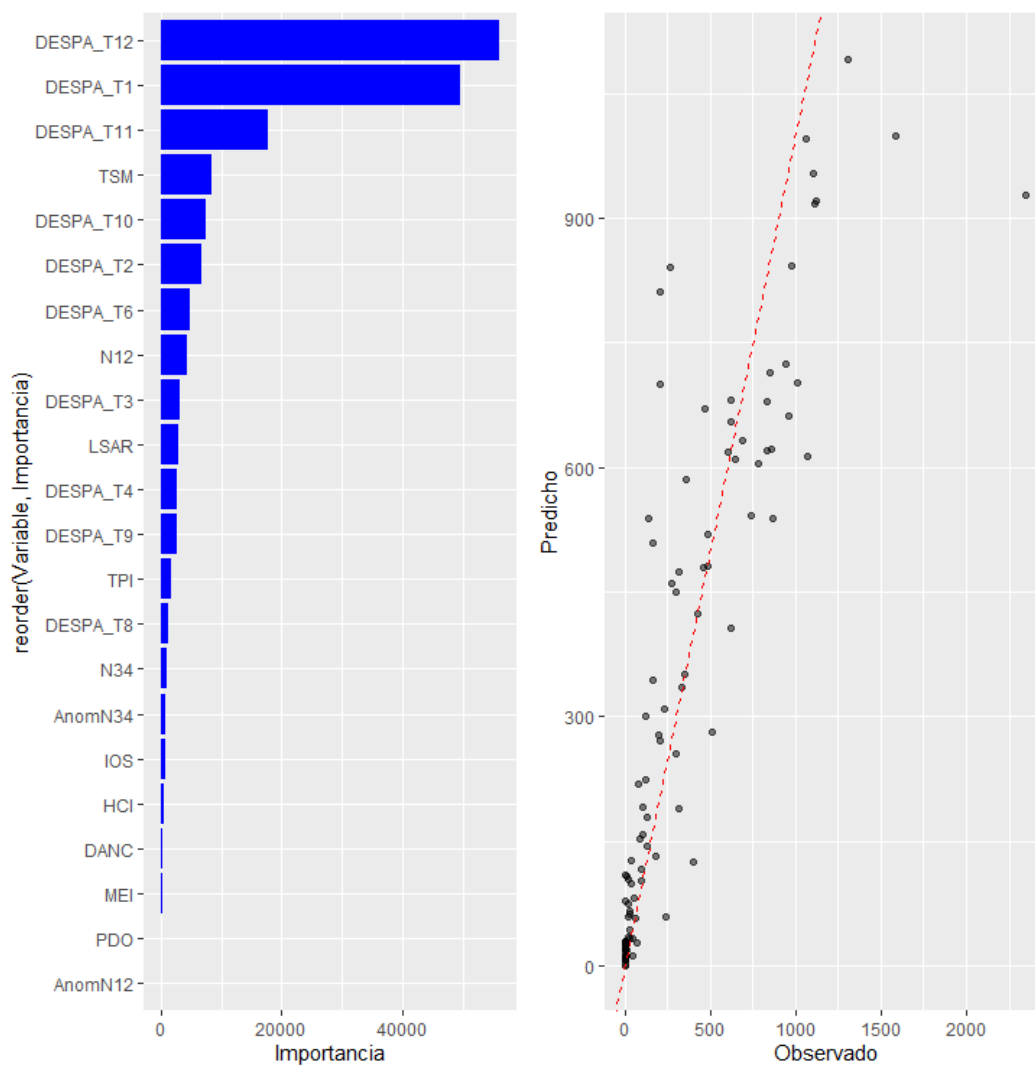
Variable	Modelo	MSE	R2
DESPA	Random Forest	51235	<b>0,74</b>
DESPA	XGBoost	54745	0,70
DTIBU	Random Forest	954	<b>0,35</b>
DTIBU	XGBoost	968	<b>0,35</b>
DDORA	Random Forest	469	<b>0,63</b>
DDORA	XGBoost	513	0,59

Para el pez espada (DESPA), el modelo Random Forest alcanzó un  $R^2$  de 0,74 y un MSE de 51.235, mientras que XGBoost presentó un desempeño ligeramente inferior, con  $R^2$  de 0,70 y MSE de 54.745. La **Figura 33** muestra la importancia de las variables en Random Forest, donde DESPA T-12, DESPA T-1, DESPA T-11 Y TSM fueron los predictores más influyentes en la variabilidad del desembarque. Asimismo, la gráfica de observados vs. predichos revela que Random Forest captura mejor la tendencia general del desembarque de DESPA en comparación con XGBoost (**Figura 34**), el cual presenta mayor dispersión en las predicciones.

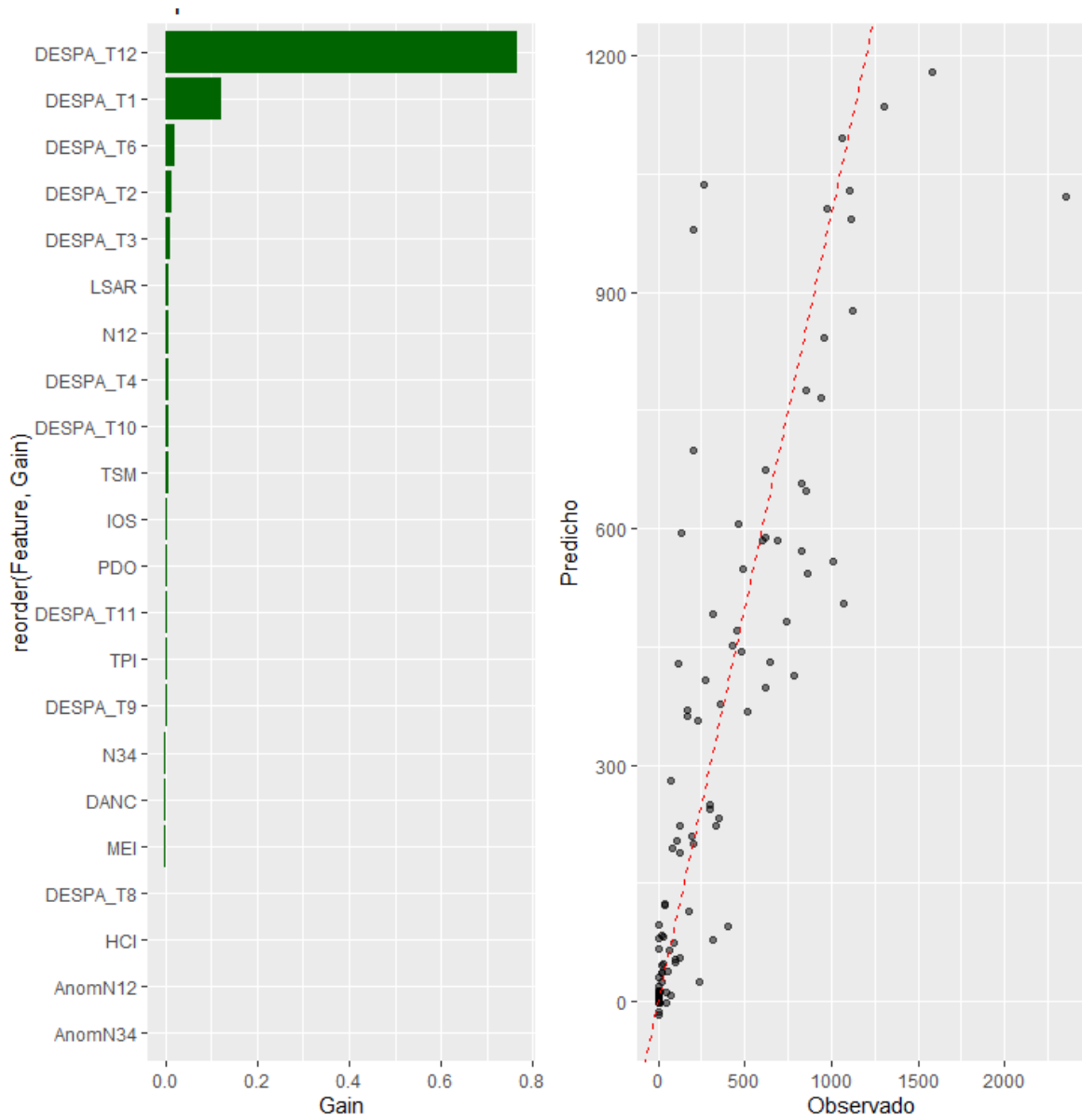
En el caso de los tiburones (DTIBU), ambos modelos obtuvieron un  $R^2$  de 0,35, indicando una menor capacidad predictiva en comparación con DESPA. No obstante, Random Forest mostró un MSE ligeramente menor (954) que XGBoost (968), lo que sugiere una mejor capacidad de generalización del primer modelo. La **Figura 35** destaca que DTIBU T-1, DTIBU T-2, DTIBU T-3 y TPI, fueron las variables más relevantes para la predicción, aunque su impacto en la variabilidad del desembarque es bajo. En la **Figura 36**, se observa que la dispersión de los valores predichos es elevada en ambos modelos, lo que sugiere que factores adicionales, podrían estar afectando los desembarques y no han sido capturados en los modelos.

Para el dorado de altura (DDORA), Random Forest nuevamente superó a XGBoost, con un  $R^2$  de 0,63 y MSE de 469, en comparación con el  $R^2$  de 0,59 y MSE de 513 obtenidos por XGBoost. La **Figura 37** indica que DDORA T-1, DDORA T-11, TSM y N34 fueron las variables con mayor influencia en el desembarque, lo que sugiere una fuerte relación con las condiciones oceanográficas. En la **Figura 38**, se observa que Random Forest predice con mayor precisión los valores observados en comparación con XGBoost, lo que refuerza su mejor desempeño para esta especie.

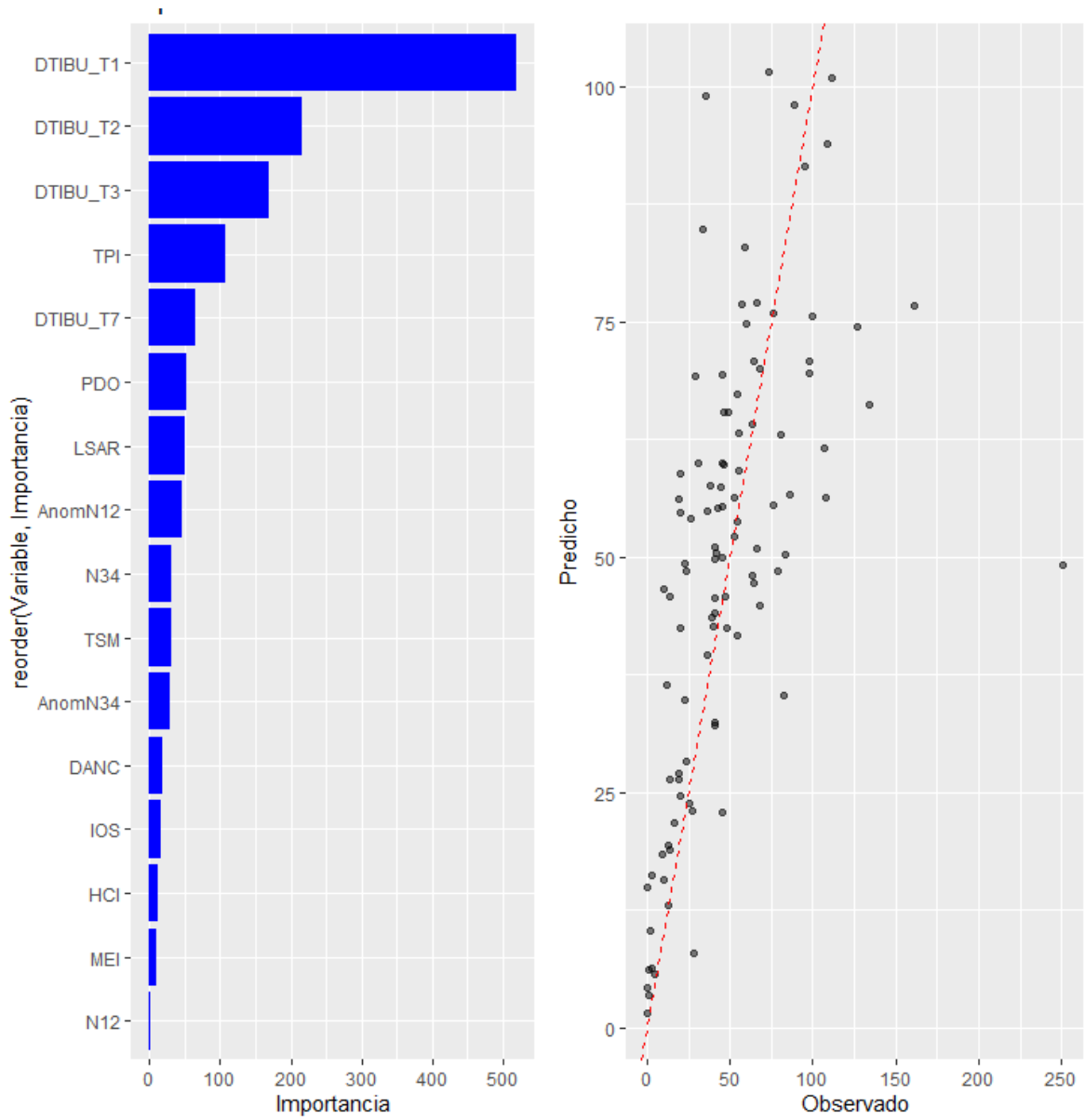
En general, los resultados indican que *Random Forest* es el modelo más adecuado para la predicción del desembarque de especies altamente migratorias, especialmente para el pez espada y el dorado de altura. No obstante, en el caso de los tiburones.



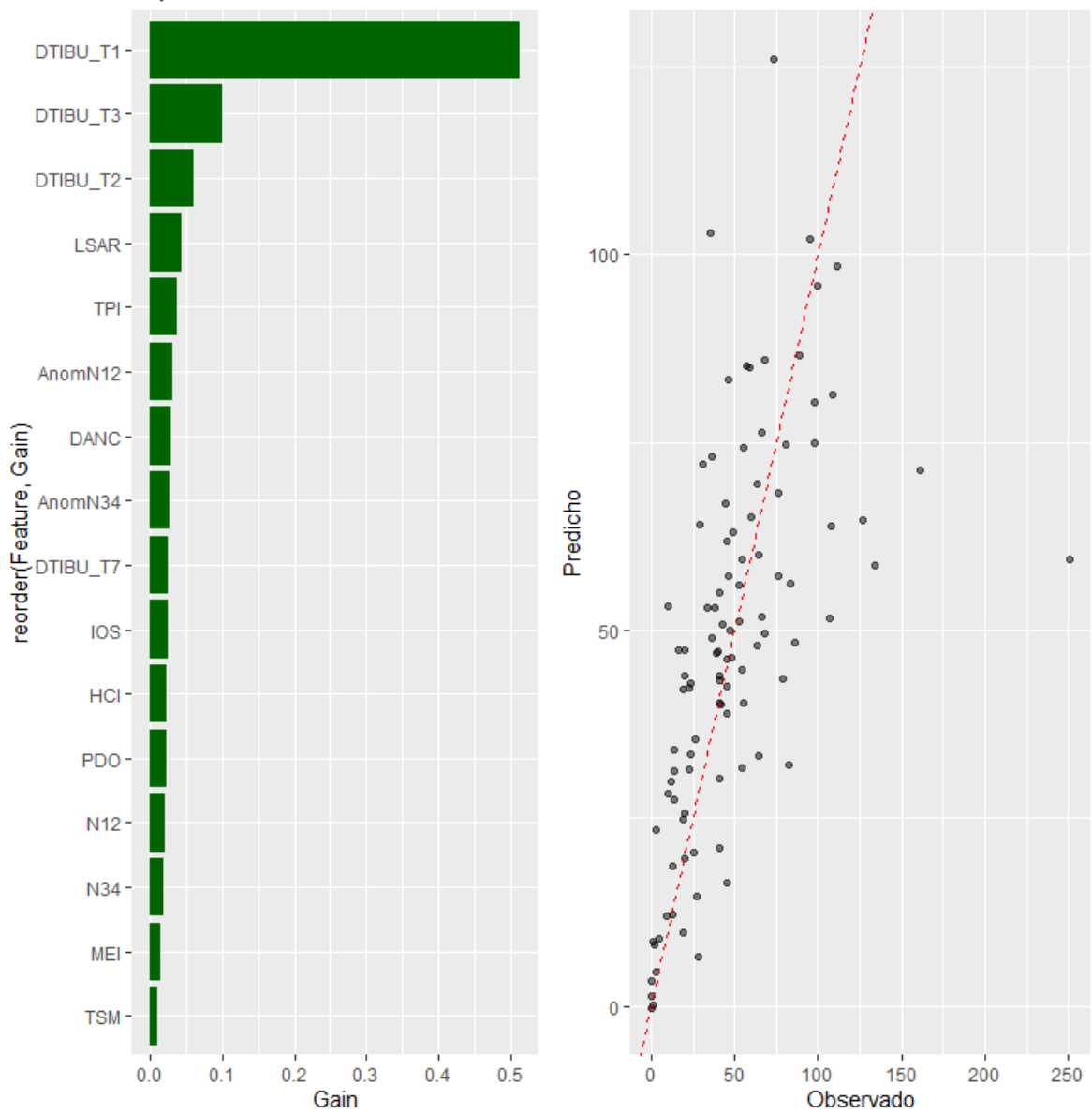
**Figura 33.** Importancia de Variable (izquierda) y grafico observados v/s predichos (derecha) del modelo *Random Forest* para la variable de interés DESPA.



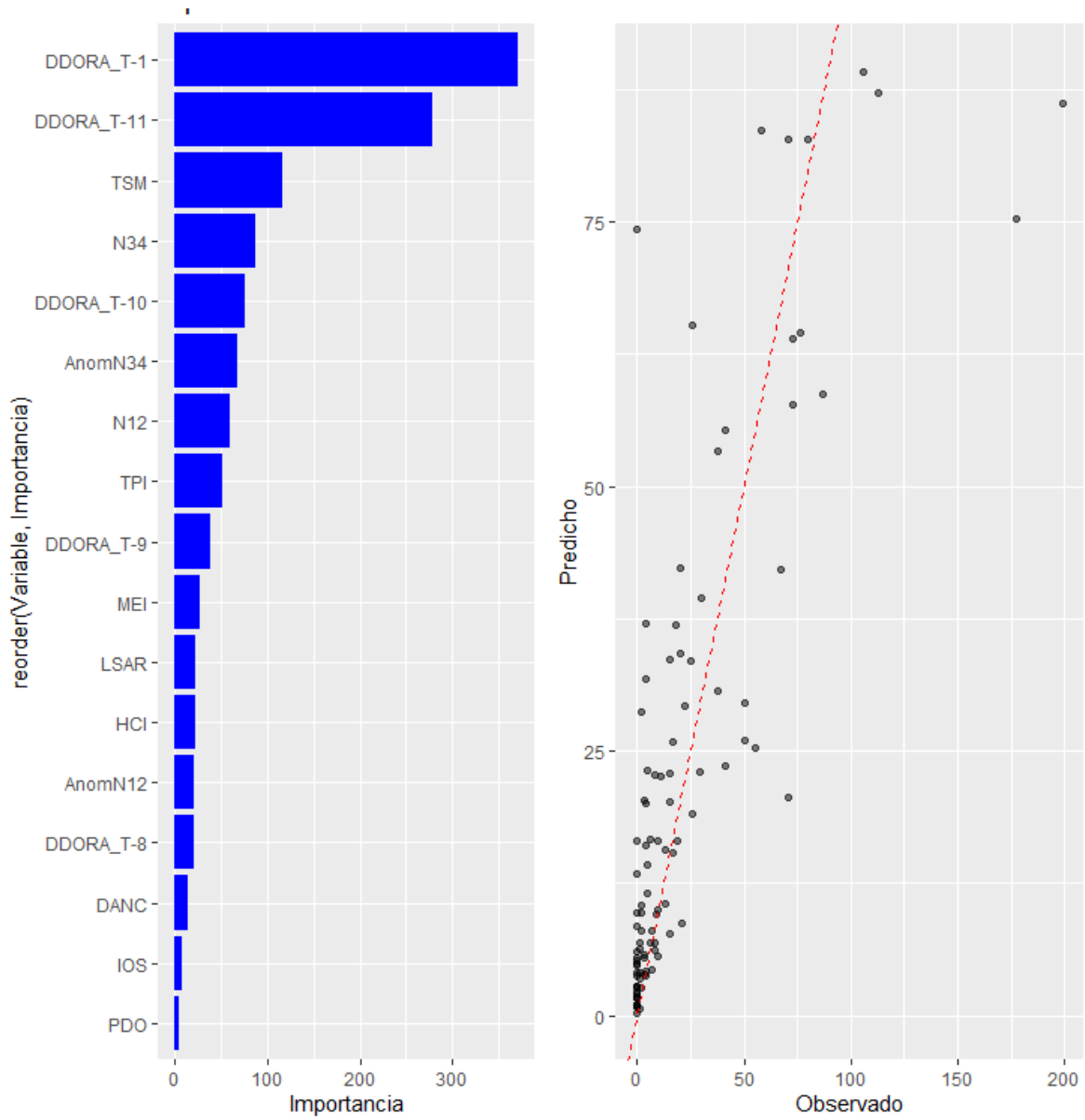
**Figura 34.** Importancia de Variable (izquierda) y grafico observados v/s predichos (derecha) del modelo XGBoost para la variable de interés DESPA



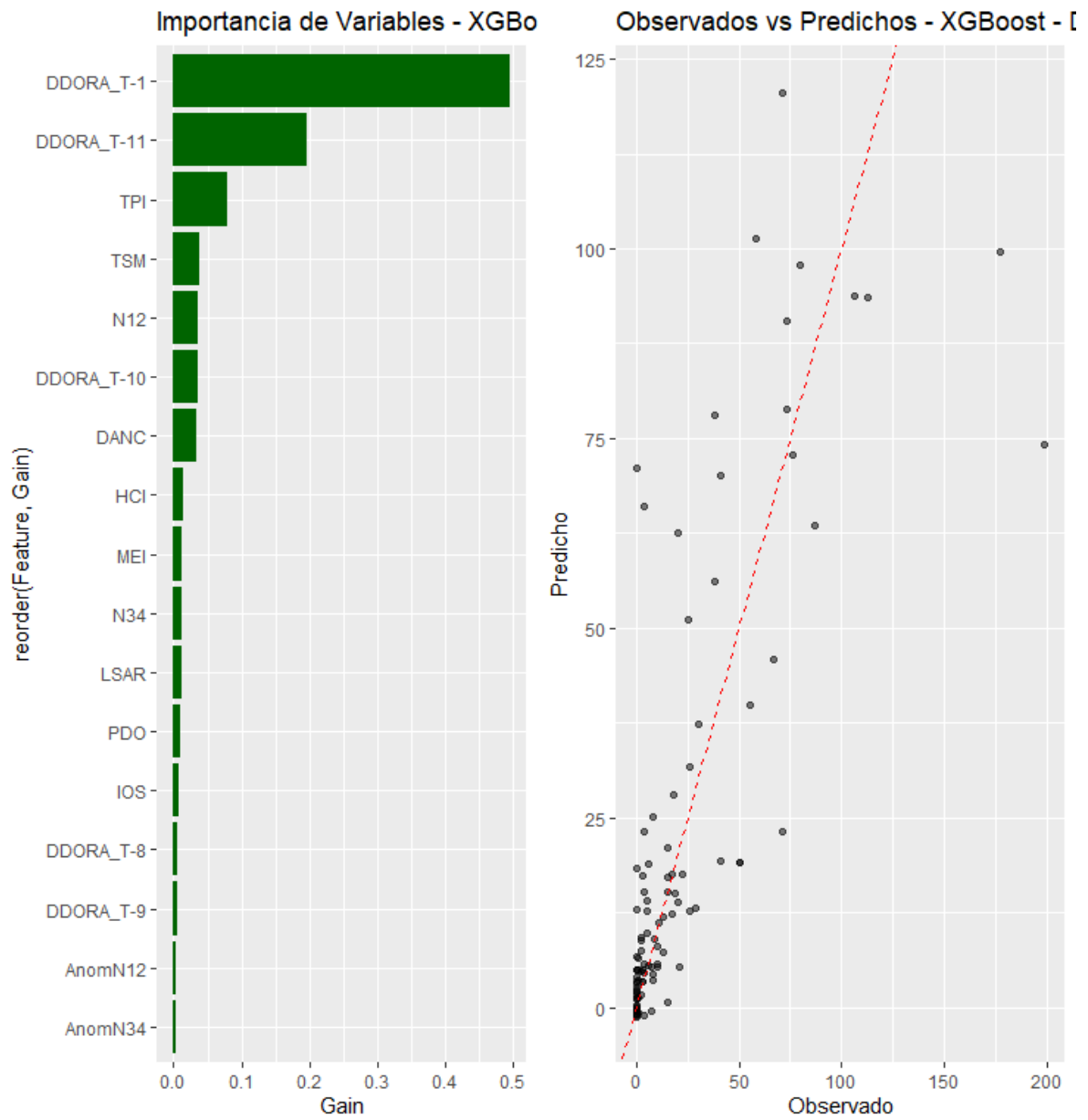
**Figura 35.** Importancia de Variable (izquierda) y grafico observados v/s predichos (derecha) del modelo *Random Forest* para la variable de interés DTIBU.



**Figura 36.** Importancia de Variable (izquierda) y grafico observados v/s predichos (derecha) del modelo *XGBoost* para la variable de interés DTIBU.



**Figura 37.** Importancia de Variable (izquierda) y grafico observados v/s predichos (derecha) del modelo *Random Forest* para la variable de interés DDORA.



**Figura 38.** Importancia de Variable (izquierda) y grafico observados v/s predichos (derecha) del modelo *XGBoost* para la variable de interés DDORA.

### 5 Discusión

---

El presente estudio analiza la influencia de las variables ambientales en las pesquerías altamente migratorias de la zona norte de Chile, aplicando un enfoque basado en el Descubrimiento del Conocimiento en los Datos (DCD) (Fayyad et al., 1996; Han et al., 2003). Para ello, se estructuró en seis etapas metodológicas que abarcaron desde la exploración de datos hasta la validación de modelos predictivos, combinando técnicas como Análisis de Componentes Principales (ACP), agrupamiento de patrones, modelado de series temporales y algoritmos de aprendizaje automático (*Random Forest* y *XGBoost*).

El uso de herramientas de aprendizaje automático ha permitido abordar la complejidad de la relación entre las condiciones ambientales y los desembarques pesqueros desde una perspectiva cuantitativa y sistemática. El ACP facilitó la reducción de dimensionalidad y la identificación de factores climáticos dominantes en la variabilidad de las capturas, mientras que los análisis de agrupamiento permitieron diferenciar períodos climáticos con impactos diferenciados en las especies. Asimismo, el análisis de series temporales permitió reconocer tendencias y ciclos en los desembarques, proporcionando un marco temporal para interpretar las fluctuaciones observadas. Finalmente, la implementación de modelos de aprendizaje automático, en particular *Random Forest* y *XGBoost*, permitió evaluar la capacidad predictiva de las variables ambientales sobre las capturas, destacando su utilidad para el conocimiento de la dinámica de la pesquería.

El calentamiento global ha generado cambios significativos en los hábitats marinos, obligando a muchas especies a desplazarse hacia aguas más profundas en busca de condiciones térmicas óptimas (Jordà et al., 2020). Este fenómeno afecta especialmente a especies pelágicas, como la anchoveta (*Engraulis ringens*), cuya migración está fuertemente influenciada por la temperatura superficial del mar (TSM) y la dinámica del Sistema de la Corriente de Humboldt (SCH) y la Oscilación Decadal del Pacífico (PDO) (Bakun & Weeks, 2008; Parada et al., 2013; Yáñez et al., 2017). A su vez, estos cambios en la distribución de especies presa impactan la disponibilidad de recursos para depredadores clave, como el pez espada, el tiburón azul, el marrajo y el dorado de altura, afectando las dinámicas pesqueras en la región (Silva et al., 2015).

Fenómenos como El Niño y La Niña generan alteraciones en la distribución y abundancia de especies marinas. Espíndola et al. (2011) documentaron una relación entre la posición de la isoterma de 18°C y los rendimientos de pesca del pez espada en el Pacífico suroriental, mientras que Nan Jay Su et al. (2020) destacaron que estos eventos modifican la productividad primaria y los patrones migratorios de las especies objetivo. Parada et al. (2013) señalaron que, además de influir en la abundancia, estos fenómenos alteran la estructura trófica del ecosistema, afectando la disponibilidad de alimento y la dinámica poblacional de especies pelágicas y demersales.

Estos cambios complican la gestión pesquera, ya que requieren ajustes en las estrategias de explotación y monitoreo debido a la variabilidad en la distribución de las especies y el aumento de costos operativos para la flota. Además, índices climáticos de largo plazo, como PDO y el Índice Tripolar del Pacífico (TPI), generan transformaciones estructurales en el ecosistema, impactando la disponibilidad y distribución de especies altamente migratorias.

Las variables climáticas con mayor correlación con la PC1, como el PDO, IOS, MEI, N12, N34, las anomalías 1+2 y 3.4, y el TPI, están estrechamente relacionadas con la temperatura superficial del mar (TSM). Estos índices reflejan patrones oceánicos y atmosféricos de gran escala que influyen directamente en las condiciones térmicas del océano. Por ejemplo, el MEI integra múltiples indicadores climáticos, incluyendo la TSM en diversas regiones del Pacífico, lo que le permite capturar la variabilidad climática en distintas zonas. Este comportamiento es consistente con estudios previos, como los de Yáñez et al. (2010) y Plaza et al. (2017), quienes identificaron una relación entre las fluctuaciones en estos índices y cambios significativos en la distribución y abundancia de especies marinas. De manera similar, Silva et al. (2015) modelaron los efectos de la TSM sobre el pez espada y la sardina, evidenciando cómo el aumento de temperatura afecta su disponibilidad y distribución. La interacción entre estas variables climáticas sugiere que los índices mencionados capturan adecuadamente la influencia del ambiente en las zonas evaluadas, lo cual es crucial para comprender las dinámicas pesqueras y anticipar cambios en la disponibilidad de los recursos.

El pez espada y los tiburones presentan correlaciones significativas con índices climáticos como MEI y N34, que actúan como predictores de los desembarques. Estos índices podrían ser aplicados para anticipar fluctuaciones en las capturas, permitiendo una planificación más efectiva de las temporadas de pesca y estrategias de conservación. Además, las flotas pesqueras podrían optimizar sus rutas y esfuerzos de captura dependiendo de las condiciones ambientales previstas, minimizando así costos operativos y reduciendo el impacto sobre las poblaciones vulnerables. Estas estrategias también podrían incluir la implementación de áreas marinas protegidas temporales durante eventos climáticos extremos para favorecer la recuperación de las especies. Este hallazgo es consistente con estudios previos que resaltan cómo las anomalías térmicas y las fluctuaciones en la circulación oceánica afectan las dinámicas pesqueras (Espíndola et al., 2011). Asimismo, las capturas de tiburón azul y marrajo muestran un declive significativo desde los años 2000, posiblemente asociado a la sobrepesca, la presión ambiental y cambios en la oferta y demanda pesquera (Convención sobre Especies Migratorias, 2007). En contraste, el dorado de altura presenta una marcada estacionalidad con picos de captura asociados a períodos de mayor TSM, subrayando la influencia de las condiciones climáticas locales, lo que refuerza la necesidad de estrategias de manejo diferenciadas (Adams et al., 2016).

Los análisis de patrones recurrentes en las dinámicas pesqueras y climáticas también muestran que, durante eventos intensos de El Niño, como los ocurridos en 1997-1998, las fluctuaciones en los desembarques son más pronunciadas, especialmente en especies como el tiburón azul y el pez espada. Estas fluctuaciones se correlacionan con picos destacados en PC2, que captura los efectos combinados de las anomalías térmicas y estas especies. Durante eventos de La Niña, en cambio, se observa una mayor estabilidad en las capturas, especialmente en el dorado de altura, lo que sugiere una mayor resiliencia de esta especie a las condiciones frías (Barria et al., 2020). Este comportamiento también está vinculado a la variabilidad en índices de surgencia y TSM, que influyen directamente en la biomasa de presas clave como la anchoveta y la sardina, fundamentales para la red trófica (Yáñez et al., 2017; Plaza et al., 2017).

Al analizar las series de tiempo, se observa que la PC1, asociada a variables ambientales, presenta picos positivos en los años 1982, 1987, 1991, 1997, 2009 y 2015, los cuales coinciden con períodos cálidos del fenómeno El Niño. En contraste, los años 1984, 1988, 1996, 1999, 2006, 2010, 2017 y 2022 están asociados a fases frías del fenómeno La Niña. Estos eventos ocurren dentro de un sistema de alternancia entre condiciones cálidas y frías, reflejando la variabilidad de

la serie temporal. La clasificación térmica del fenómeno de El Niño (Null, 2024) muestra eventos débiles (0.5 a 1 °C), moderados (1 a 1.5 °C), fuertes (1.5 a 2 °C) y muy fuertes ( $\geq 2$  °C). Estos patrones permiten comprender cómo la variabilidad climática afecta las capturas pesqueras y sugiere que los modelos predictivos pueden mejorar la planificación pesquera y la sostenibilidad en la región.

El análisis multivariado realizado confirma que el 73% de la variabilidad total en los datos es explicada por las cinco primeras componentes principales (PCA). La PC1, dominada por variables climáticas como PDO, IOS y MEI, subraya la influencia de los ciclos climáticos de gran escala, mientras que la PC2 refleja la interacción entre variables pesqueras (DESPA, DTIBU, DMARR) y climáticas (TPI). Estos resultados concuerdan con la literatura, que enfatiza cómo los cambios en el SCH afectan la distribución y abundancia de especies pelágicas menores, impactando a los depredadores que dependen de ellas (Bakun & Weeks, 2008; Gatica et al., 2009). En cuanto a la PC2, esta está principalmente asociada a variables pesqueras de interés como DESPA (pez espada), DAZUL (tiburón azul), DMARR (marrajo), DTIBU (tiburones en general) y DSAR (sardina), junto con variables climáticas como el Índice de Corriente de Humboldt (HCI) y el Índice Tripolar del Pacífico (TPI). Las series de tiempo muestran fluctuaciones notables en los desembarques de estas especies, con periodos de mayores capturas vinculados a condiciones ambientales favorables. Por ejemplo, el pez espada exhibe una tendencia creciente en el segundo periodo (1986-1997), mientras que el marrajo y el tiburón azul muestran una mayor variabilidad en los clústeres asociados a periodos cálidos. Por otro lado, la PC3 se relaciona con variables como DDORA (dorado de altura), DMARR (marrajo) y DTIBU (tiburones en general), además de la temperatura superficial del mar (TSM) y el Índice Climático de la región Niño 1+2 (N12). Las series de tiempo indican que los desembarques de dorado de altura tienden a ser más altos durante periodos neutros y cálidos, destacando su mayor resiliencia a las condiciones de El Niño. En contraste, el tiburón marrajo presenta picos de desembarques asociados a eventos específicos como 1987, 1997, 200 Y 2017 lo que sugiere una fuerte dependencia de las condiciones climáticas regionales. La utilización de PCA ha sido ampliamente utilizado en estudios pesqueros para caracterizar la estructura de las capturas y detectar patrones subyacentes en la dinámica de los recursos (Plaza et al., 2017; Hernández-Santoro & Suárez-Ahumada, 2018; Silva et al., 2015). En este sentido, la relación observada entre las principales componentes y las variables ambientales sugiere que fenómenos climáticos como la Oscilación Decadal del Pacífico (PDO), el Índice de Oscilación del Sur (IOS) y los índices asociados a El Niño y La Niña juegan un papel clave en la estructuración de la serie temporal de desembarques.

A diferencia de estudios previos, que han analizado periodos más limitados—Hernández-Santoro & Suárez-Ahumada (2018) con datos de 2004 a 2012 y Silva et al. (2015) de 2001 a 2012—este estudio cubre un lapso de 1980 a 2022, proporcionando una visión más amplia de la variabilidad interanual y de largo plazo. Mientras que estos trabajos identificaron fluctuaciones a corto y mediano plazo en la pesca de especies como la anchoveta y la sardina, la extensión de esta serie temporal permite evaluar con mayor precisión la recurrencia de eventos extremos y la influencia acumulativa del cambio climático en los desembarques. Además de la ventaja temporal, este estudio analiza un conjunto más amplio de variables ambientales y pesqueras, lo que permite una caracterización más detallada de la interacción entre el clima y las pesquerías. Plaza et al. (2017) se enfocaron en la relación entre desembarques de anchoveta y sardina con algunas variables climáticas clave, mientras que Hernández-Santoro & Suárez-Ahumada (2018) utilizaron GLM y clustering para evaluar la composición de especies en una ventana temporal más corta. Por su parte, Silva et al. (2015) aplicaron modelos GLM y GAM para evaluar los efectos del cambio

climático en la distribución y abundancia relativa del pez espada y la sardina, encontrando que las variaciones en la TSM explicaban hasta un 60.6% de la variabilidad en la CPUE de sardina y un 58.7% en la del pez espada. Estos resultados refuerzan la importancia de incluir múltiples enfoques estadísticos y series temporales extensas para comprender la influencia climática sobre las pesquerías.

Por otro lado, el análisis de Agrupamiento en series temporales identificó tres grupos principales de comportamiento en los desembarques cuando se excluyó el pez espada, y hasta nueve clústeres cuando se excluyeron los tiburones o el dorado de altura, lo que indica una mayor heterogeneidad en la estructura de los datos. La segmentación de las series temporales en distintos períodos reveló que la dinámica pesquera se ha modificado en función de la variabilidad climática, con cambios estructurales significativos alrededor de 1986-1997, 1997-2008 y 2008-2022. Estos cambios coinciden con el fenómeno de El Niño y fluctuaciones de largo plazo en los índices oceánicos, lo que sugiere que la estructura de los desembarques de las especies está estrechamente vinculada a estas oscilaciones (Chavez et al., 2003; Lluch-Cota et al., 2001; Schwing et al., 2010). Además, la menor cohesión de ciertos clústeres refleja una mayor variabilidad interna, lo que podría estar asociado a la influencia de factores adicionales no modelados, como la presión pesquera o cambios en la gestión de los recursos.

En relación con los modelos y su desempeño, los resultados mostraron que Random Forest, seguido de KNN y XGBoots, obtuvo los mejores desempeños, mientras que los modelos de Redes Neuronales y Redes Neuronales Profundas (MLP) presentaron un rendimiento menor. Esto concuerda con estudios previos que sugieren que los modelos de aprendizaje profundo no son la opción más eficiente para datos tabulares, ya que los enfoques basados en árboles de decisión, como Random Forest y XGBoost, suelen ser más adecuados en este tipo de problemas (Borisov et al. 2022; Grinsztajn, Oyallon y Varoquaux 2022; Shwartz-Ziv y Armon 2022). En este contexto, la estructura de los datos tabulares favorece el uso de modelos de ensamblado basados en árboles, los cuales pueden capturar la heterogeneidad y las relaciones no lineales de manera más eficiente, ofreciendo además mayor interpretabilidad y estabilidad en los resultados.

El análisis de series temporales y autocorrelación de los desembarques de pez espada, tiburones y dorado de altura evidenció una fuerte dependencia temporal en los registros históricos, reflejada en la significancia de los retardos detectados mediante la función de autocorrelación parcial (PACF). La presencia de autocorrelación significativa sugiere que los valores pasados de los desembarques influyen en los futuros, lo que refuerza la idea de que los patrones de captura no son aleatorios, sino que responden a dinámicas cíclicas y estacionales asociadas tanto a la disponibilidad del recurso como a factores ambientales.

El Test de Ljung-Box confirmó la existencia de correlación temporal en todas las especies analizadas, lo que valida la inclusión de retardos en los modelos predictivos. Sin embargo, la periodicidad observada en la PACF varía según la especie, sugiriendo que los ciclos de desembarque no son homogéneos. En particular, para desembarque de pez espada, se detectaron retardos significativos hasta 12 meses, mientras que para los tiburones y el dorado de altura, los retardos relevantes se concentran en los primeros meses y en períodos más amplios, reflejando diferentes patrones de respuesta temporal. Destaca especialmente el hecho de que la variable más relevante en la predicción de los desembarques de pez espada haya sido DESPA\_t-12, lo que sugiere una alta tendencia a seguir los valores de captura de hace un año. Este comportamiento puede estar vinculado a la estacionalidad de la especie, influenciada por su

migración anual, la planificación del esfuerzo pesquero u oferta/demanda que determinan la estacionalidad en las capturas.

A partir de estos hallazgos, se implementaron modelos de aprendizaje automático (Random Forest y XGBoost) con el objetivo de capturar relaciones no lineales y evaluar la influencia de las variables ambientales en la predicción de los desembarques. Ambos modelos lograron identificar la importancia de factores climáticos como la temperatura superficial del mar (TSM), el Índice Niño 1+2 (N12) y el Índice Tripolar del Pacífico (TPI) en la variabilidad de las capturas. En particular, la TSM mostró un impacto diferenciado en las especies analizadas, siendo especialmente relevante para el pez espada.

Además, los resultados confirman que los modelos capturan mejor las tendencias generales de los desembarques que sus fluctuaciones a corto plazo, lo que refuerza la necesidad de incorporar modelos dinámicos que integren efectos temporales y espaciales de manera más explícita. En este sentido, la integración de técnicas de aprendizaje automático con datos en tiempo real podría mejorar la planificación pesquera, optimizando la distribución del esfuerzo pesquero y facilitando estrategias adaptativas ante escenarios de cambio climático.

Los modelos predictivos basados en Agrupamiento y análisis de series temporales han demostrado ser herramientas eficaces para identificar patrones recurrentes en las dinámicas pesqueras y climáticas de estas especies. Métodos como el análisis multivariado y los algoritmos de agrupamiento han permitido prever fluctuaciones en los desembarques y sincronizar eventos climáticos con variaciones en la disponibilidad de los recursos pesqueros (Witten et al., 2011; Queiroz et al., 2010).

Los modelos predictivos aplicados en este estudio han demostrado ser herramientas eficaces para identificar patrones recurrentes en las dinámicas pesqueras y climáticas. En particular, el análisis multivariado y los algoritmos de agrupamiento han permitido anticipar fluctuaciones en los desembarques y relacionarlas con variaciones en las condiciones ambientales (Witten et al., 2011; Queiroz et al., 2010). Estos enfoques, combinados con modelos de envolvente bioclimática como los utilizados por Silva et al. (2015) para predecir la distribución del pez espada y la sardina común en Chile bajo escenarios de cambio climático, refuerzan la importancia de integrar factores ambientales en la planificación pesquera. Además, la implementación de plataformas de monitoreo en tiempo real permitiría optimizar la gestión pesquera mediante ajustes dinámicos en las cuotas de captura y la delimitación de zonas de pesca, mejorando la sostenibilidad del recurso y la eficiencia del esfuerzo pesquero en un contexto de variabilidad climática creciente.

Adicionalmente, los métodos de minería de datos han posibilitado la detección de patrones no evidentes, como la relación entre fluctuaciones en la temperatura superficial del mar y los cambios en los volúmenes de desembarque durante eventos climáticos extremos (Fayyad et al., 1996; Plaza et al., 2017). Sin embargo, este estudio representa una primera aproximación, y los resultados podrían fortalecerse significativamente si se incorporara información sobre la Captura por Unidad de Esfuerzo (CPUE) en los modelos predictivos. La inclusión de esta variable permitiría cuantificar de manera más precisa la disponibilidad de los recursos y su relación con las condiciones ambientales, proporcionando una base más sólida para la toma de decisiones y aumentando la robustez y aplicabilidad de los resultados en la gestión pesquera.

No obstante, la implementación efectiva de estos modelos requiere la participación de equipos multidisciplinarios, que integren expertos en estadística, dinámica poblacional de las especies y

oceanografía, asegurando una interpretación adecuada de los resultados y su aplicación en la gestión pesquera. A pesar de estos avances, la ausencia de un comité de manejo, un comité científico-técnico y la falta de cuotas de extracción específicas para estas especies representan una barrera para la adopción de estrategias de regulación basadas en evidencia científica.

### 6 Conclusiones

---

Este estudio se estructuró en seis etapas metodológicas que abarcaron desde la exploración de datos hasta la aplicación y validación de modelos predictivos. La combinación de técnicas como el Análisis de Componentes Principales (ACP), el agrupamiento de patrones, el modelado de series temporales y algoritmos de aprendizaje automático (*Random Forest* y *XGBoost*) permitió evaluar la influencia de las variables ambientales en el desembarque de especies altamente migratorias en la zona norte de Chile.

Los hallazgos obtenidos refuerzan la relevancia de los factores climáticos en la dinámica pesquera, destacando el impacto de fenómenos como El Niño y La Niña. Los ciclos de gran escala, representados por índices como PDO, MEI y TPI, influyen significativamente en la abundancia de especies clave como el pez espada, el tiburón azul, el marrajo y el dorado de altura. Eventos intensos de El Niño, como los de 1997-1998 y 2015-2016, generaron fluctuaciones marcadas en los desembarques, especialmente en los tiburones y el pez espada. En contraste, La Niña se asoció con una mayor estabilidad en las capturas, particularmente del dorado de altura, lo que sugiere diferencias en la resiliencia de las especies ante condiciones climáticas adversas.

El análisis de componentes principales evidenció que la PC1, dominada por variables climáticas como MEI, N3.4 y Anom 1+2, explica gran parte de la variabilidad observada en los desembarques. La PC2 refleja la interacción entre factores pesqueros y ambientales, mientras que la PC3 resalta el papel de la temperatura superficial del mar y los índices climáticos regionales en la distribución de especies como el dorado de altura y el tiburón marrajo.

El estudio confirmó la presencia de patrones temporales recurrentes en los desembarques de especies altamente migratorias en Chile, con autocorrelaciones significativas que respaldan el uso de modelos predictivos basados en retardos temporales. En particular, el pez espada mostró una fuerte dependencia de sus desembarques previos a 12 meses, evidenciando un patrón estacional estable en su dinámica pesquera.

El uso de modelos de aprendizaje automático, como *Random Forest* y *XGBoost*, permitió identificar relaciones no lineales entre los desembarques y las variables ambientales, resaltando el impacto de la temperatura superficial del mar (TSM), el Índice Niño 1+2 (N12) y el Índice Tripolar del Pacífico (TPI). En este contexto, el TPI mostró una influencia diferenciada sobre el dorado, lo que coincide con su preferencia por aguas cálidas, mientras que la persistencia de DESPA<sub>t-12</sub> refuerza la existencia de una dinámica cíclica en la pesquería del pez espada.

Los modelos predictivos aplicados demostraron ser herramientas efectivas para identificar patrones en las capturas. No obstante, la variabilidad a corto plazo sigue representando un desafío, lo que sugiere la necesidad de incorporar indicadores adicionales, como la captura por unidad de esfuerzo (CPUE), para mejorar la capacidad predictiva de los modelos. Entre las variables ambientales más influyentes para las especies de interés se identificaron la TSM, N1+2 y TPI, lo que concuerda con la influencia del sistema de surgencia del Humboldt en la distribución de estas especies.

Asimismo, la combinación de técnicas de agrupamiento y análisis de series temporales permitió caracterizar períodos cálidos, fríos y neutros, proporcionando información clave para la gestión pesquera. En general, los eventos cálidos asociados a El Niño coincidieron con incrementos en los desembarques de ciertas especies, mientras que las condiciones frías vinculadas a La Niña favorecieron la estabilidad en las capturas de otras.

Estos resultados destacan la importancia de integrar índices climáticos de gran escala en la planificación pesquera, facilitando la adaptación de la actividad a las condiciones ambientales predominantes. La incorporación de modelos predictivos avanzados en la gestión pesquera permitiría optimizar la distribución del esfuerzo y minimizar el impacto sobre los recursos, promoviendo estrategias de conservación que eviten la explotación en periodos de mayor vulnerabilidad.

En un contexto de creciente incertidumbre ambiental, la aplicación de enfoques adaptativos basados en modelos estadísticos y aprendizaje automático representa una herramienta clave para mitigar los efectos del cambio climático y contribuir a una mejor comprensión de la dinámica de estas especies.

## 7 Bibliografía

---

Bakun, A., & Weeks, S. J. (2008). The marine ecosystem off Peru: What are the secrets of its fishery productivity and what might its future hold? *Progress in Oceanography*, 79(2–4), 290–299. <https://doi.org/10.1016/j.pocean.2008.10.027>

Barría, P., González, A., Devia, D., Mora, S., Miranda, H., Barraza, A., Cerna, F., Cid, L., & Ortega, J. (2019). *Seguimiento pesquería recursos altamente migratorios. Aspectos biológicos pesqueros, 2018. Informe Final*. IFOP - Subsecretaría de Pesca.

Barría, P., González, A., Devia, D., Mora, S., Miranda, H., Barraza, A., Cerna, F., Cid, L., & Ortega, J. (2020). Programa de seguimiento de las principales pesquerías nacionales, año 2019: Recursos altamente migratorios. Aspectos biológicos-pesqueros. Informe final. IFOP - Subsecretaría de Pesca.

Bedriñana-Romano, L., Zarate, P. M., Huckle-Gaete, R., Viddi, F. A., Buchan, S. J., Cari, I., ... & Zerbini, A. N. (2022). Abundance and distribution patterns of cetaceans and their overlap with vessel traffic in the Humboldt Current Ecosystem, Chile. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-14465-7>

Bedriñana-Romano, L., Huckle-Gaete, R., Viddi, F. A., Johnson, D., Zerbini, A. N., Morales, J., ... & Palacios, D. M. (2021). Defining priority areas for blue whale conservation and investigating overlap with vessel traffic in Chilean Patagonia, using a fast-fitting movement model. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-82220-5>

Borisov, Vadim, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk, y Gjergji Kasneci. 2022. «Deep neural networks and tabular data: A survey». *IEEE transactions on neural networks and learning systems*.

Bustamante, C., Acuña, E., Tapia-Jopia, C. & Vargas-Caro, C. (2023). Actualización del Plan de Acción Nacional para la conservación y manejo de tiburones de Chile. Informe Final Proyecto del Fondo de Investigación Pesquera y de Acuicultura (FIPA N°2021-24). Universidad de Antofagasta.

Cady, F. (2017). *The data science handbook*. John Wiley & Sons.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 - Step-by-Step Data Mining Guide. Retrieved from <https://the-modeling-agency.com/crisp-dm.pdf>

Chavez, F. P., Ryan, J., Lluch-Cota, S. E., & Ñiquen C., M. (2003). From anchovies to sardines and back: Multidecadal change in the Pacific Ocean. *Science*, 299(5604), 217-221. <https://doi.org/10.1126/science.1075880>

Closas, A. H., Arriola, E. A., Kuc, C. I., Amarilla, M. R., & Jovanovich, E. C. (2013). *Análisis multivariante, conceptos y aplicaciones en Psicología Educativa y Psicometría*. Enfoques, 25(1), 65-92. Recuperado de <https://www.redalyc.org/articulo.oa?id=25930006005>

Cios, K. J. (2007). *Data Mining: A Knowledge Discovery Approach*. Springer. <https://doi.org/10.1007/978-0-387-36795-8>

CMS. (2007). Documento preliminar sobre el estado de conservación de tiburones migratorios y posibles opciones de cooperación internacional bajo la Convención sobre las Especies Migratorias. Reunión para la Identificación y Elaboración de una Opción para la Cooperación Internacional sobre Tiburones Migratorios bajo la Convención sobre las Especies Migratorias, Mahé, Seychelles, 11-13 de diciembre de 2007. [https://www.cms.int/sites/default/files/document/Doc\\_04\\_Background\\_Paper\\_S\\_0.pdf](https://www.cms.int/sites/default/files/document/Doc_04_Background_Paper_S_0.pdf)

Cope, J., & Punt, A. (2009). Drawing the lines: Resolving fishery management units with simple fisheries data. *Canadian Journal of Fisheries and Aquatic Sciences*, 66(8), 1256–1273.

Daneri, G., Dellarossa, V., Quiñones, R., Jacob, B., Montero, P., & Ulloa, O. (2000). Primary production and community respiration in the Humboldt Current System off Chile and associated oceanic areas. *Marine Ecology Progress Series*, 197, 41–49.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications*. New York: Wiley & Sons.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3), 32–57. <https://doi.org/10.1080/01969727308546046>

Dutton, J. (2021, September 1). *Pacific Decadal Oscillation*. World Climate Service. Retrieved from <https://www.worldclimateservice.com/2021/09/01/pacific-decadal-oscillation/>

Enomoto, K., Ishikawa, S., Hori, M., Sitha, H., Song, S., Thuok, N., & Kurokura, H. (2011). Data mining and stock assessment of fisheries resources in Tonle Sap Lake, Cambodia. *Fisheries Science*, 77(5), 713–722.

Espino, M. (2013). *El jurel Trachurus murphyi y las variables ambientales de macroescala*. En Csirke, J., Guevara-Carrasco, R., & Espino, M. (Eds.), *Ecología, pesquería y conservación del jurel (Trachurus murphyi) en el Perú*. Revista Peruana de Biología, 20(1), 9-20.

Espíndola, F., Yáñez, E., & Barbieri, M. Á. (2011). El Niño Southern Oscillation and spatial-temporal variability of the nominal performances of swordfish (*Xiphias gladius*) in the southeastern Pacific. *Revista de Biología Marina y Oceanografía*, 46(2), 231–242. <https://doi.org/10.4067/s0718-19572011000200012>

Espíndola, F., Quiroz, J. C., Wiff, R., & Yáñez, E. (2016). Incorporating sea surface temperature into the stock-recruitment relationship: An application for Chilean jack mackerel (*Trachurus murphyi*). *Revista de Biología Marina y Oceanografía*, 51(1), 137-145. DOI: 10.4067/S0718-19572016000100013.

FAO. (2020). *El estado mundial de la pesca y la acuicultura 2020. La sostenibilidad en acción*. Roma. <https://doi.org/10.4060/ca9229es>

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 3(17), 37–54.

Gatica, C. (2009). Asociación entre la Corriente de Deriva de los Vientos del Oeste y la abundancia relativa del pez espada (*Xiphias gladius*) frente a la costa de Chile. *Latin American Journal of Aquatic Research*, 37(1), 97–105. <https://doi.org/10.3856/vol37-issue1-fulltext-8>

Grinsztajn, Léo, Edouard Oyallon, y Gaël Varoquaux. 2022. «Why do tree-based models still outperform deep learning on typical tabular data?» *Advances in neural information processing systems* 35: 507-20.

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5ta ed.). Boston, MA: Houghton Mifflin.

Hernández-Santoro, C., & Suárez-Ahumada, B. (2018). Análisis de la composición de especies en la pesquería de cerco orientada a la captura de anchoveta (*Engraulis ringens*) en la Zona Norte de Chile. *Revista de Biología Marina y Oceanografía*, 53(1), 51-60. <https://doi.org/10.4067/S0718-19572018000100051>.

Hormazábal, S., Shaffer, G., & Pizarro, O. (2002). Tropical Pacific control of intraseasonal oscillations off Chile by way of oceanic and atmospheric pathways. *Geophysical Research Letters*, 29(6). <https://doi.org/10.1029/2001GL013481>

Hormazábal, S., Shaffer, G., Letelier, J., & Ulloa, O. (2001). Local and remote forcing of sea surface temperature in the coastal upwelling system off Chile. *Journal of Geophysical Research Atmospheres*, 106(C8), 19657–19671. <https://doi.org/10.1029/2001JC900008>

Huettmann, F., Arthukin, Y., Gilg, O., & Humphries, G. (2011). Predictions of 27 Arctic pelagic seabird distributions using public environmental variables, assessed with colony data: A first digital IPY and GBIF open access synthesis platform. *Marine Biodiversity*, 41, 141–179.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>

Jordà, G., Marbà, N., Bennett, S., Santana-Garcón, J., Agustí, S., & Duarte, C. M. (2020). Ocean warming compresses the three-dimensional habitat of marine life. *Nature Ecology & Evolution*, 4(1), 109–114. <https://doi.org/10.1038/s41559-019-1058-0>

Kaiser, H. F. (1960). "The Application of Electronic Computers to Factor Analysis." *Educational and Psychological Measurement*, 20(1), 141–151.

Liao, S., Chu, P., & Hsiao, P. (2012). Data mining techniques and applications: A decade review from 2000 to 2011. *Expert Systems with Applications*, 12(39), 11303–11311.

Lluch-Cota, S. E., Wooster, W. S., & Hare, S. R. (2001). Sea surface temperature variability in coastal areas of the northeastern Pacific related to the El Niño-Southern Oscillation and the

Pacific Decadal Oscillation. *Geophysical Research Letters*, 28(10), 2029-2032.  
<https://doi.org/10.1029/2000GL012429>

Montecino, V., & Lange, C. B. (2009). The Humboldt Current System: Ecosystem components and processes, fisheries, and sediment studies. *Progress in Oceanography*, 83(1-4), 65-79.  
<https://doi.org/10.1016/j.pocean.2009.07.041>

Nan Jay Su, Zhou, S., Lyne, V., Du, Y., & Shi, W. (2020). A data-mining approach to determine the spatio-temporal relationship between environmental factors and fish distribution. *Ecological Modelling*, 174(4), 421–431. <https://doi.org/10.1016/j.ecolmodel.2020.104407>

Naranjo, L., Plaza, F., Yáñez, E., Barbieri, M. Á., & Sánchez, F. (2015). *Forecasting of jack mackerel landings (Trachurus murphyi) in central-southern Chile through neural networks*. *Fisheries Oceanography*, 24(3), 219-228. <https://doi.org/10.1111/fog.12105>

Null, J. (2024). El Niño and La Niña years and intensities. Golden Gate Weather Services. Recuperado el [diciembre 2024], de <https://ggweather.com/enso/oni.htm>

Nurdin, Taufiq, Fajriana, & Zia Ulhaq, M. (2023). Mapping System Model and Clustering of Fishery Products using K-Means Algorithm with Web GIS Approach. *International Journal of Intelligent Systems and Applications in Engineering*, 11(3). DOI:10.12691/jbms-9-4-4.

Pearson, K. (1901) Sobre líneas y planos de ajuste más cercano a sistemas de puntos en el espacio. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2, 559-572. <https://doi.org/10.1080/14786440109462720>

Plaza, F., Araya, H., & Yáñez, E. (2023). *Environmental effect on the variability of anchovy (Engraulis ringens) in northern Chile: Autoregressive conditional heteroskedastic approach with exogenous variable and missing values*. *Fisheries Research*, 260, 106607. <https://doi.org/10.1016/j.fishres.2022.106607>

Plaza, F., Salas, R., & Yáñez, E. (2017). Identifying ecosystem patterns from time series of anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) landings in northern Chile. *Journal of Statistical Computation and Simulation*, 88(10), 1863–1881. <https://doi.org/10.1080/00949655.2017.1410150>

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)

Schwing, F. B., Mendelssohn, R., Bograd, S. J., Overland, J. E., Wang, M., & Ito, S. (2010). Climate change, teleconnection patterns, and regional processes forcing marine populations in the Pacific. *Journal of Marine Systems*, 79(3-4), 245-257. <https://doi.org/10.1016/j.jmarsys.2008.11.027>

Silber, G. (1990). Presencia y distribución de la vaquita (*Phocoena sinus*) en el norte del Golfo de California. *Fishery Bulletin*, 88, 339–346.

Shwartz-Ziv, Ravid, y Amitai Armon. 2022. «Tabular data: Deep learning is not all you need». *Information Fusion* 81: 84-90.

Silva, C., Yáñez, E., Barbieri, M. Á., Bernal, C., & Aranís, A. (2015). *Forecasts of swordfish (Xiphias gladius) and common sardine (Strangomera bentincki) off Chile under the A2 IPCC climate change scenario*. *Progress in Oceanography*, 134, 343-355. <https://doi.org/10.1016/j.pocean.2015.03.004>

Su, F., Zhou, S., Lyne, V., Du, Y., & Shi, W. (2004). A data-mining approach to determine the spatio-temporal relationship between environmental factors and fish distribution. *Ecological Modelling*, 174(4), 421-431.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the Gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423. <https://doi.org/10.1111/1467-9868.00293>

Witten, I., Frank, E., & Hall, A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

Yang, Y., Cai, X., & Herricks, E. (2008). Identification of hydrologic indicators related to fish diversity and abundance: A data mining approach for fish community analysis. *Water Resources Research*, 44(4). <https://doi.org/10.1029/2006WR005763>

Yáñez, E., Plaza, F., & Barbieri, M. A. (2010). Aplicación de redes neuronales artificiales en pesquerías. *Journal of Applied Statistics*, 37(2), 241-254.

Yáñez, E., Barbieri, M. Á., Plaza, F., & Silva, C. (2012). *Cambio climático y pesquerías en Chile*. En *Pesquerías y Acuicultura en Chile: Desafíos y Oportunidades* (pp. 141-150). Pontificia Universidad Católica de Valparaíso.

Yáñez, E., Gutiérrez, A., & Barbieri, M. A. (2016). Efectos del cambio climático en el SCH bajo escenarios proyectados. *Marine Environmental Research*, 120, 45-54.

Yáñez, E., Plaza, F., Silva, C., Sánchez, F., Barbieri, M. Á., & Aranís, A. (2016). *Pelagic resources landings in central-southern Chile under the A2 climate change scenarios*. *Ocean Dynamics*, 66(10), 1333-1351. <https://doi.org/10.1007/s10236-016-0984-5>

Yáñez, E., Plaza, F., & Silva, C. (2017). Modelos predictivos en pesquerías frente a la variabilidad climática. *Fish and Fisheries*, 18(4), 563-579.

Yáñez, E., Lagos, N. A., Norambuena, R., Silva, C., Letelier, J., Muck, K. P., San Martín, G., Benítez, S., Broitman, B. R., Contreras, H., Duarte, C., Gelcich, S., Labra, F. A., Lardies, M. A., Manríquez, P. H., Quijón, P. A., Ramajo, L., González, E., Molina, R., Gómez, A., Soto, L., Montecino, A., Barbieri, M. Á., Plaza, F., Sánchez, F., Aranís, A., Bernal, C., & Böhm, G. (2018). *Impacts of climate change on marine fisheries and aquaculture in Chile*. En B. F. Phillips & M. Pérez-Ramírez (Eds.), *Climate Change Impacts on Fisheries and Aquaculture: A Global Analysis, Volume I* (pp. 239-244). John Wiley & Sons Ltd.

Zárate, P., Clavijo, C. I., Devia, L., Bedriñana, D., Hucke, L., Viddi, R., ... & Vásquez, C. (2021). *Pesquería Recursos Altamente Migratorios, Enfoque Ecosistémico, año 2020. Informe Final*. Instituto de Fomento Pesquero.