



Universitat
Oberta
de Catalunya



UNIVERSITAT DE
BARCELONA

**PREDICCIÓN DE POTENCIALES
ZONAS DE PESCA DE LA ESPECIE
SCOMBER JAPONICUS
EN EL PACÍFICO ESTE, A TRAVÉS DE
REDES NEURONALES**

JESSICA VERA BERMUDEZ



Universitat
Oberta
de Catalunya



UNIVERSITAT DE
BARCELONA

MÁSTER UNIVERSITARIO EN BIOESTADÍSTICA Y BIOINFORMÁTICA

TUTORA: ROMINA REBRIJ
PRA: CARLES VENTURA ROYO

ENERO, 2023

CONTENIDOS



- INTRODUCCION
- OBJETIVOS

- PLAN DE TRABAJO
- MÉTODOS Y RESULTADOS

- CONCLUSIONES
- BIBLIOGRAFIA

INTRODUCCIÓN

1

PESCA Y SOBREPESCA

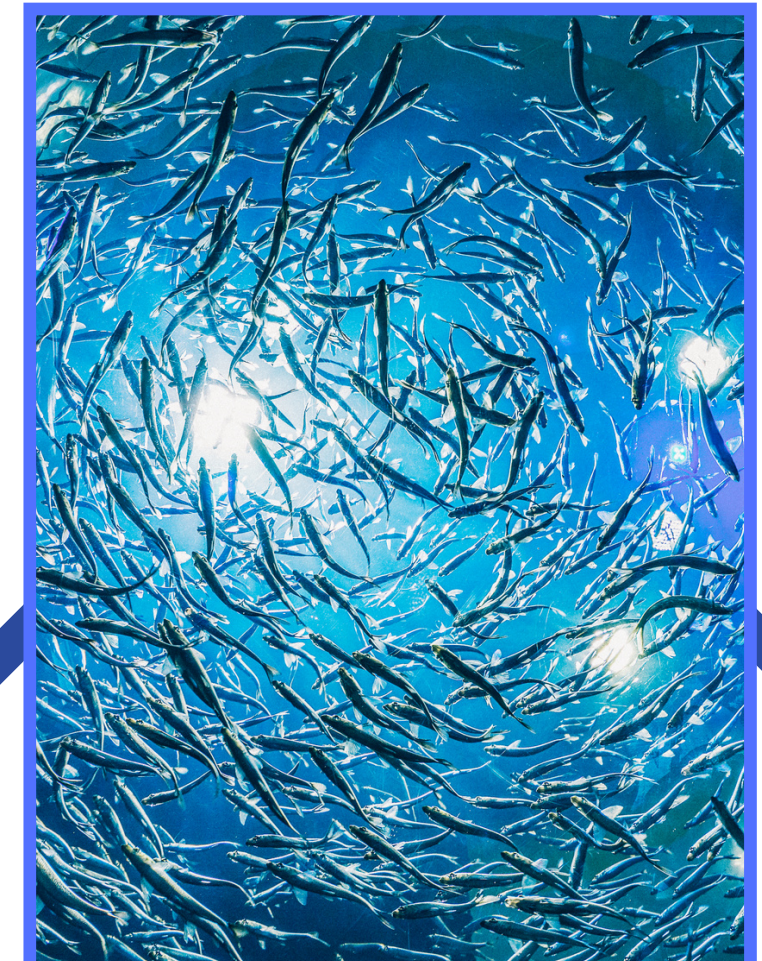
La pesca es un servicio ecosistémico de alta importancia. Sin embargo, la explotación de especies pesqueras conlleva a la sobrepesca, lo cual es actualmente una preocupación mundial.



2

SCOMBER JAPONICUS

Es una de las especies más capturadas mundialmente (1,360K Ton en 2020) y en Ecuador es la segunda más importante. Habita principalmente en el Océano Índico y Pacífico y se alimenta de especies zooplanctónicas.



INTRODUCCIÓN

3

INSOSTENIBILIDAD

Actualmente **la especie** está dentro de los **límites de pesca biológicamente sostenibles**, pero la zona del **Pacífico sureste**, donde se ubica la costa ecuatoriana, tiene el mayor porcentaje de **pesca insostenible** (66.7%).

4

ESTRATEGIAS

FAO* busca "conservar y utilizar sosteniblemente los océanos, los mares y los recursos marinos para el desarrollo sostenible" (ODS14) y de entre las propuestas está la **restauración de zonas pesqueras**

(*) Food & Agriculture Organization



INTRODUCCIÓN

5

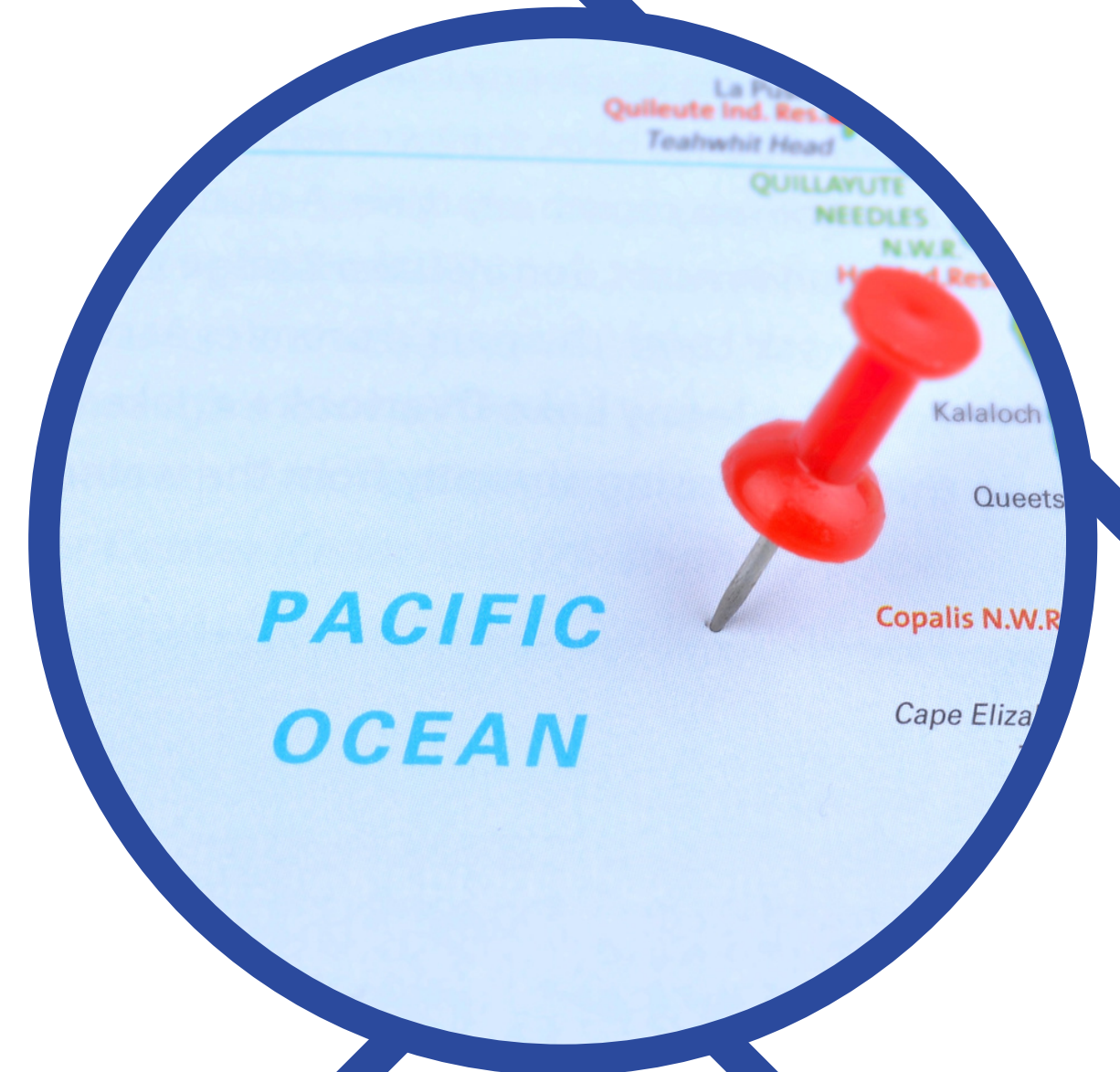
HERRAMIENTAS

*El conocer la distribución de la abundancia de la especie, permite el monitoreo y manejo de los recursos pesqueros, así como, identificar **áreas explotadas y no exploradas** potencialmente productivas.*

6

PROPUESTA

*La creación de mapas de la distribución de la **abundancia** predicha de *S. japonicus* en Ecuador, a través de redes neuronales, que permitan identificar tanto las zonas en necesidad de restauración, como aquellas potencialmente útiles en la pesca de la especie.*

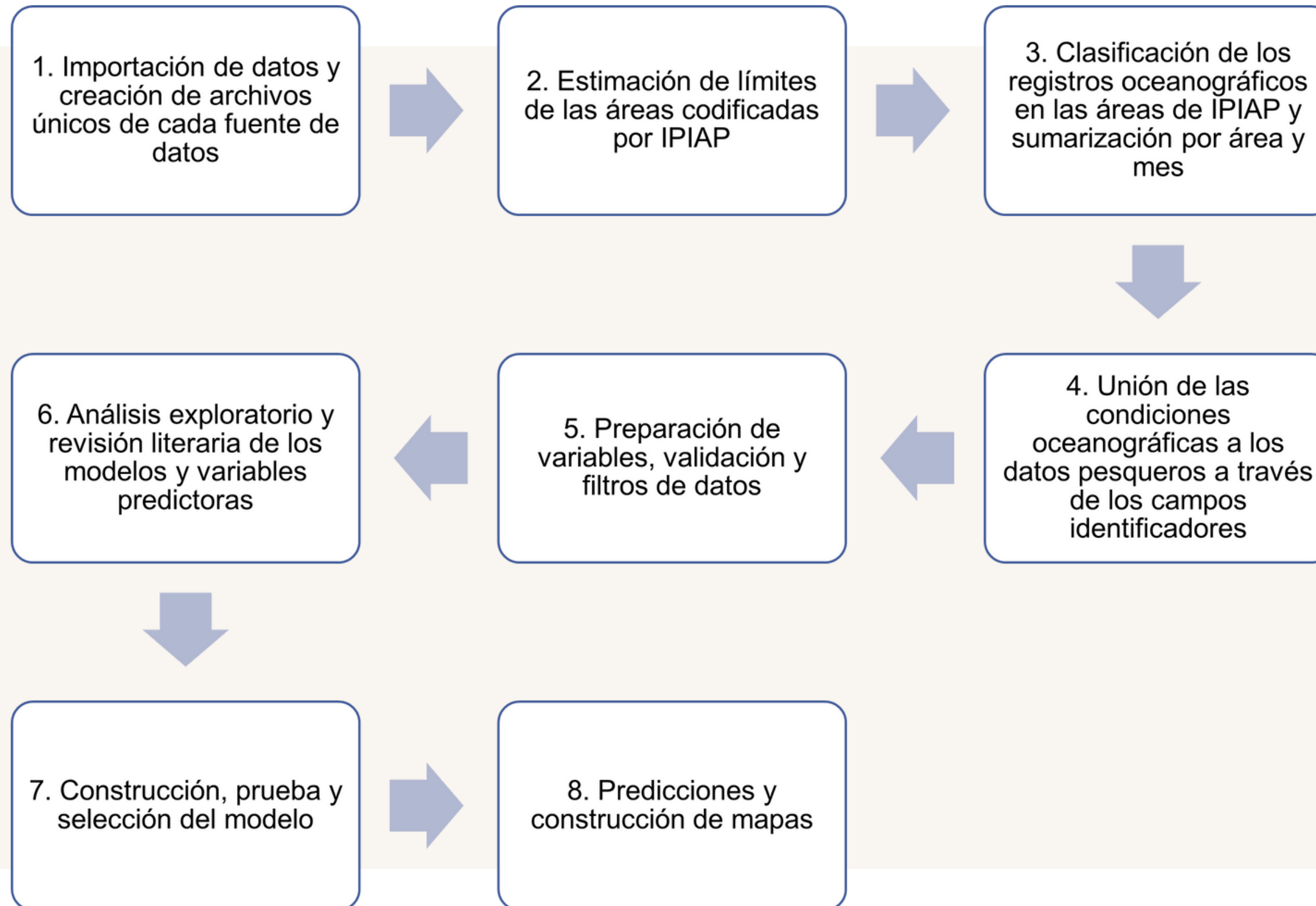


OBJETIVOS

- *Caracterizar la distribución de la especie *Scomber japonicus*, por área y mes.*
- *Predecir potenciales zonas de pesca que reúnen las características ambientales y espaciotemporales ideales para la especie *Scomber japonicus*, a través de redes neuronales.*

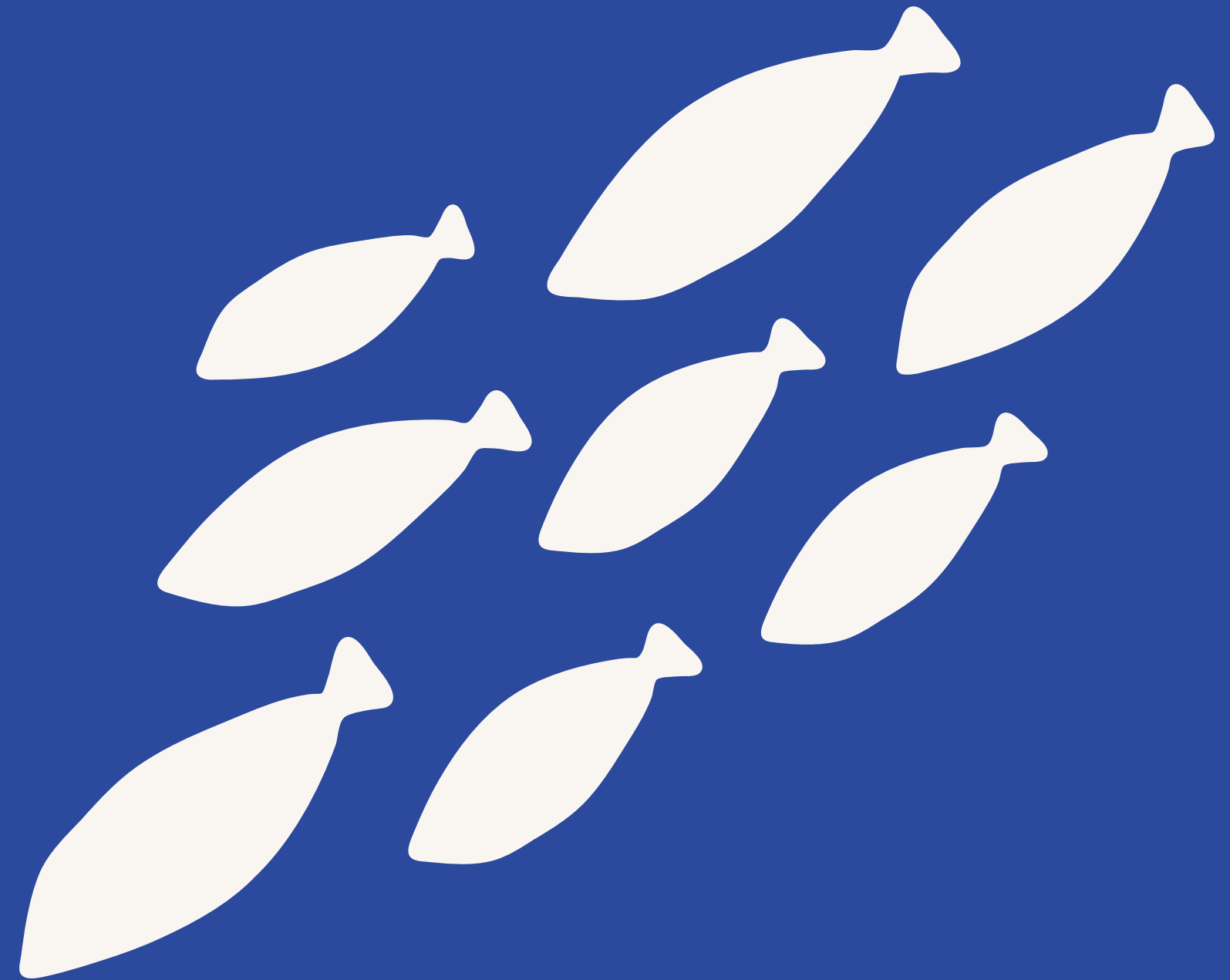


PLAN DE TRABAJO



MÉTODOS Y RESULTADOS

DETALLES DEL PLAN DE TRABAJO



1. Importación de datos y creación de archivos únicos

1

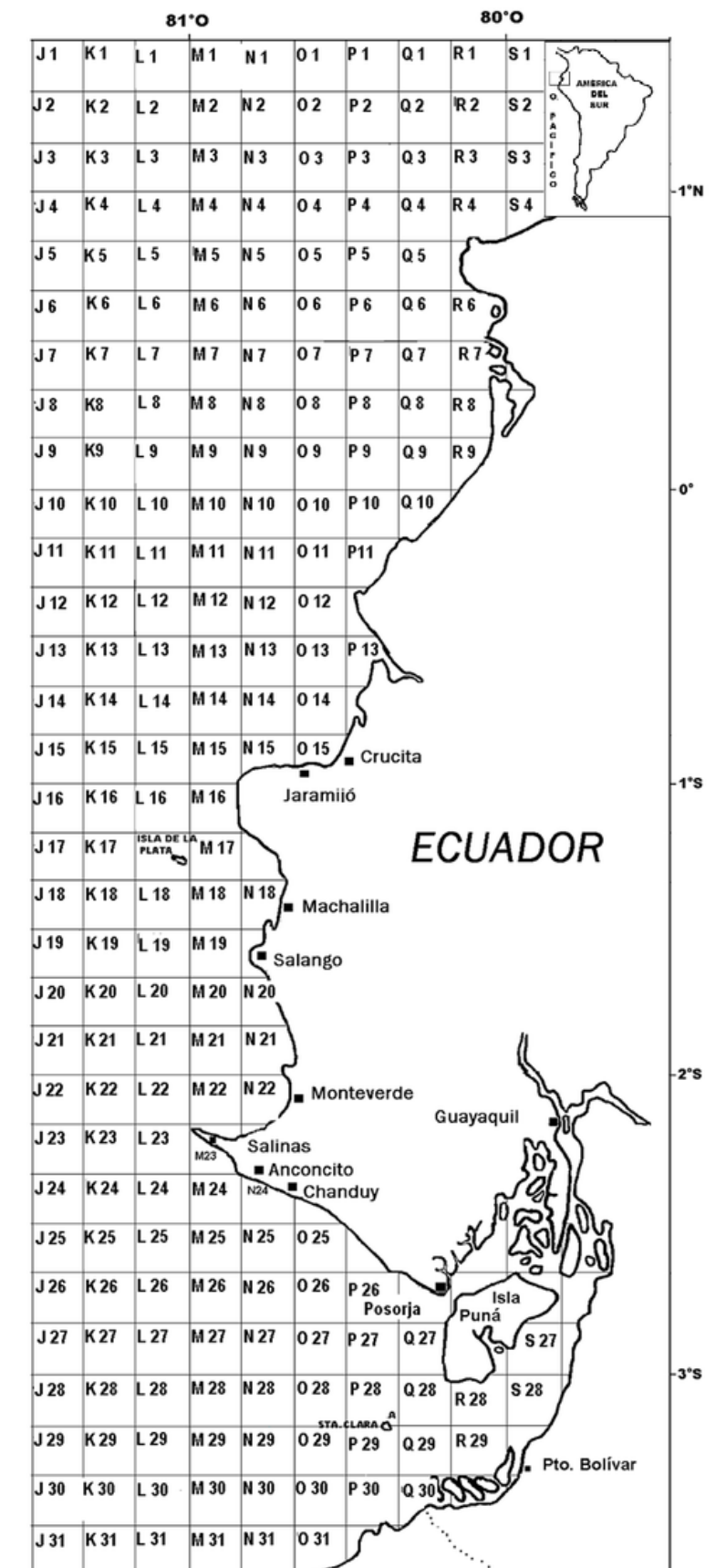
DATOS PESQUEROS

- 2,146 registros pesqueros diarios provistas IPIAP, que fueron tomados por los observadores a bordo, según planificación de IPIAP, entre 2017-01 y 2022-05
- 219 áreas de 5X11.5 millas (latitud x longitud), con la coordenada central identificada. Solo 66 contaban con registros pesqueros.
- Ubicación: latitudes entre 01°21' N y 03°35' S
- La unidad de medida de la captura fue toneladas

2

CONCENTRACIÓN DE CLOROFILA

- Tomada de los datos registrados por el sensor MODIS del satélite Aqua de la NASA.
- Medida en $mgxm^{-3}$
- Valores mensuales, con una resolución de 1 km



1. Importación de datos y creación de archivos únicos

3

TEMPERATURA SUPERFICIAL DEL AGUA

- *Tomada de los datos registrados por el sensor MODIS del satélite Aqua de la NASA*.*
- *Medida en grados Celsius*
- *Valores mensuales, con una precisión de 0.5° Celsius y una resolución de 1 km*

4

MAGNITUD DE LA VELOCIDAD DEL VIENTO

- *Tomada de los datos satelitales de NOAA*.*
- *Medida en m/s*
- *Valores mensuales, medidos a una altura de 10 metros sobre el agua y una precisión de 1x1 grados en latitud y longitud*

(*) NASA: Administración Aeronáutica Nacional del Espacio / NOAA: Administración Nacional Oceánica y Atmosférica

2. Estimación de límites de las áreas codificadas por IPIAP

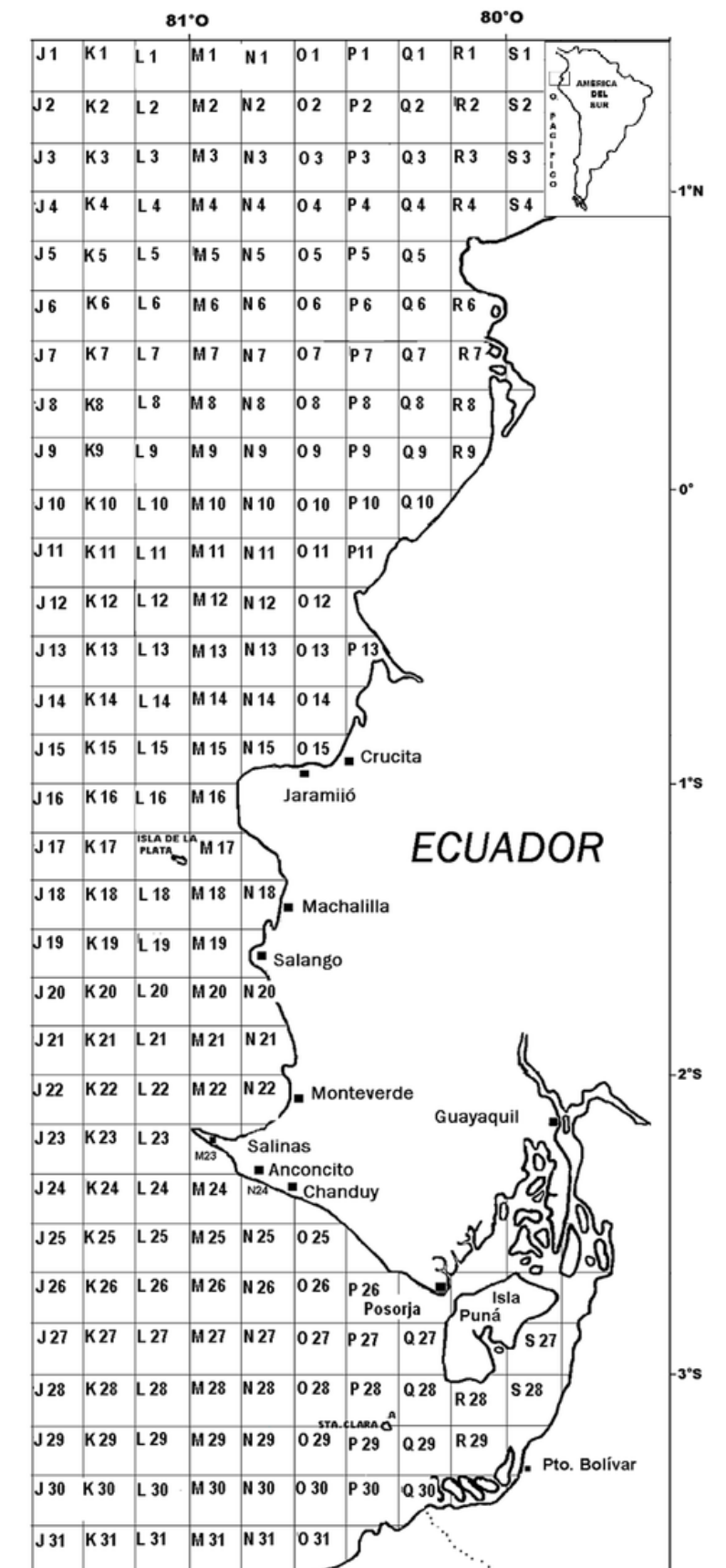
1

ESTIMACIÓN DE LÍMITES

- Con las **coordenadas centrales** (P_{xi} , P_{yi}), se **estimó los límites** de cada area de 5x11.5 millas, a través de la determinación de la longitud (Q_{xij}) y latitud (Q_{yij}) de los cuatro puntos **vértices** (j) de cada **polígono** (i).
Estimaciones basadas en la distancia inversa Haversina
- Se **asumió** que las distancias entre grados son similares para longitudes y latitudes (111 km) y que el radio de la tierra es 6,378.14 km, por la zona cercana al ecuador de la tierra

$$Q_{xij} = P_{xi} \pm \frac{\left[\frac{d_x}{2} \times \frac{180}{1000\pi r} \right]}{\cos\left(\frac{P_{yi}\pi}{180}\right)}$$

$$Q_{yij} = P_{yi} \pm \left[\frac{d_y}{2} \times \frac{180}{1000\pi r} \right]$$



3 y 4. Clasificación, sumalización y unión de datasets

1

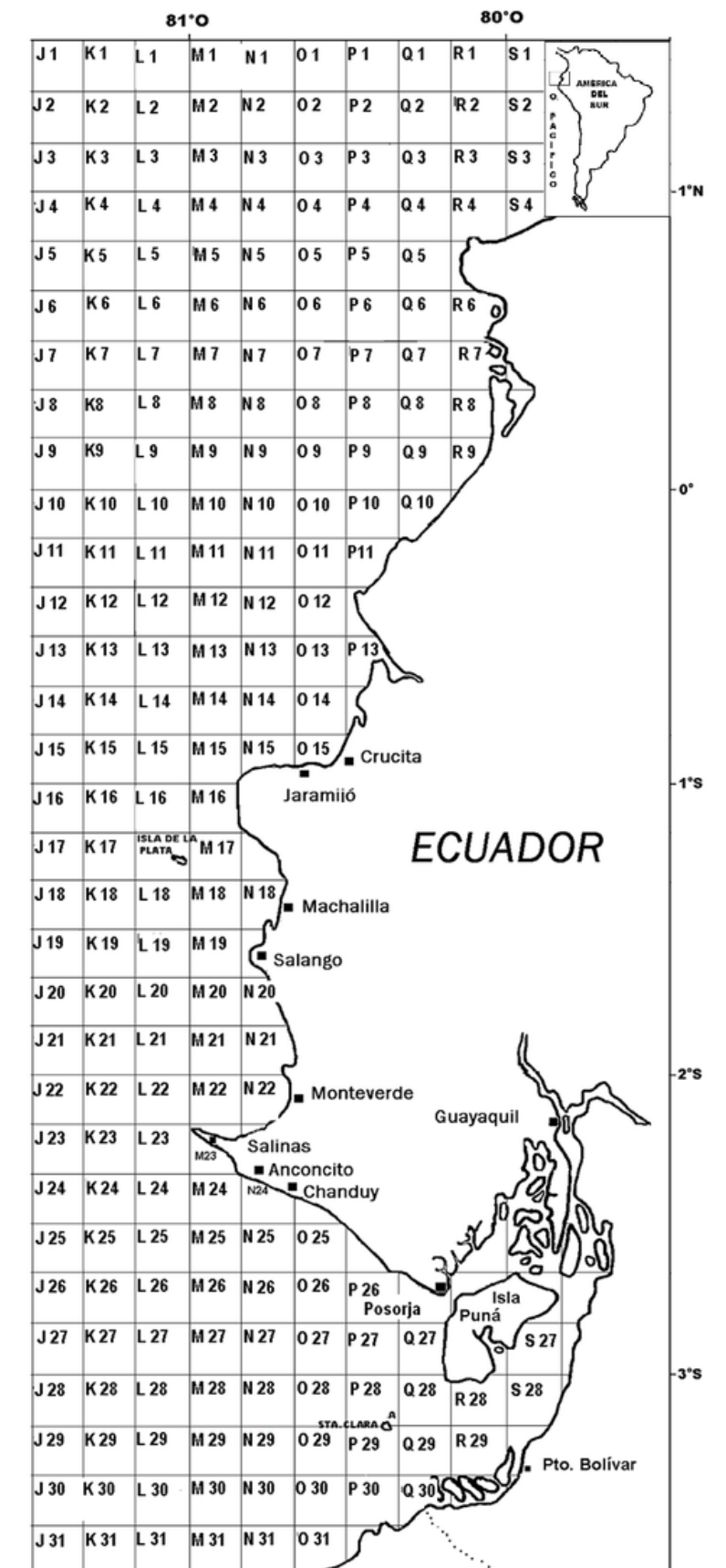
CORRECCIÓN DE LA ESTIMACIÓN

A continuación se procedió a realizar una **corrección** a las **latitudes límite**, de modo que, sean **coincidentes entre áreas contiguas** (áreas que comparten al menos un límite), repartiendo a ambas áreas la mitad de la distancia generada entre dichos límites.

2

UNIÓN DE CONJUNTOS DE DATOS

Con estos límites, se **agrupó las medidas oceanográficas**, de acuerdo con el **área y mes** en el que se enmarcaron; y, a su vez, se **promedió las respectivas mediciones**. La **unión** realizada fue **por izquierda** a favor del conjunto de datos pesqueros, (campos de unión área y mes de medición)



5. Preparación, validación y filtros de datos

1

CREACIÓN DE VARIABLE DE RESPUESTA

- Cálculo del índice de captura por unidad de esfuerzo (**CPUE**), como el promedio capturado por lance
- Construcción de un indicador de la bondad de captura, a través del índice de abundancia relativa (**IAR**), como la razón entre CPUE y el valor CPUE máximo del respectivo mes del registro.
- Este índice tiene un rango entre 0 y 1, donde 1 representa mayor abundancia de la especie.
- Como alternativa a un IAR con alto sesgo, se empleó la **transformación de raíz cuadrada**

$$CPUE_{ymi} = \frac{C_{ymi}}{f_{ymi}} \quad sqrtIAR_{ymi} = \sqrt{IAR} = \sqrt{\frac{CPUE_{ymi}}{\max(CPUE_m)}}$$

5. Preparación, validación y filtros de datos

2

VALIDACIÓN Y FILTROS DE DATOS

- Remoción de los datos registrados durante los **meses de veda** impuestos por IPIAP (marzo y septiembre)
- Identificación de **valores atípicos en CPUE**, a través del método *boxplot* ajustado para distribuciones asimétricas, empleando un bigote de longitud 3, ya que la distribución es naturalmente sesgada. Estos valores fueron descartados (36 casos)

Variable	Min	Max	Q1	Q2	Q3	Media	DE	Sesgo
CPUE	0.00	210	2.00	5.00	12.00	10.13	15.45	5.18

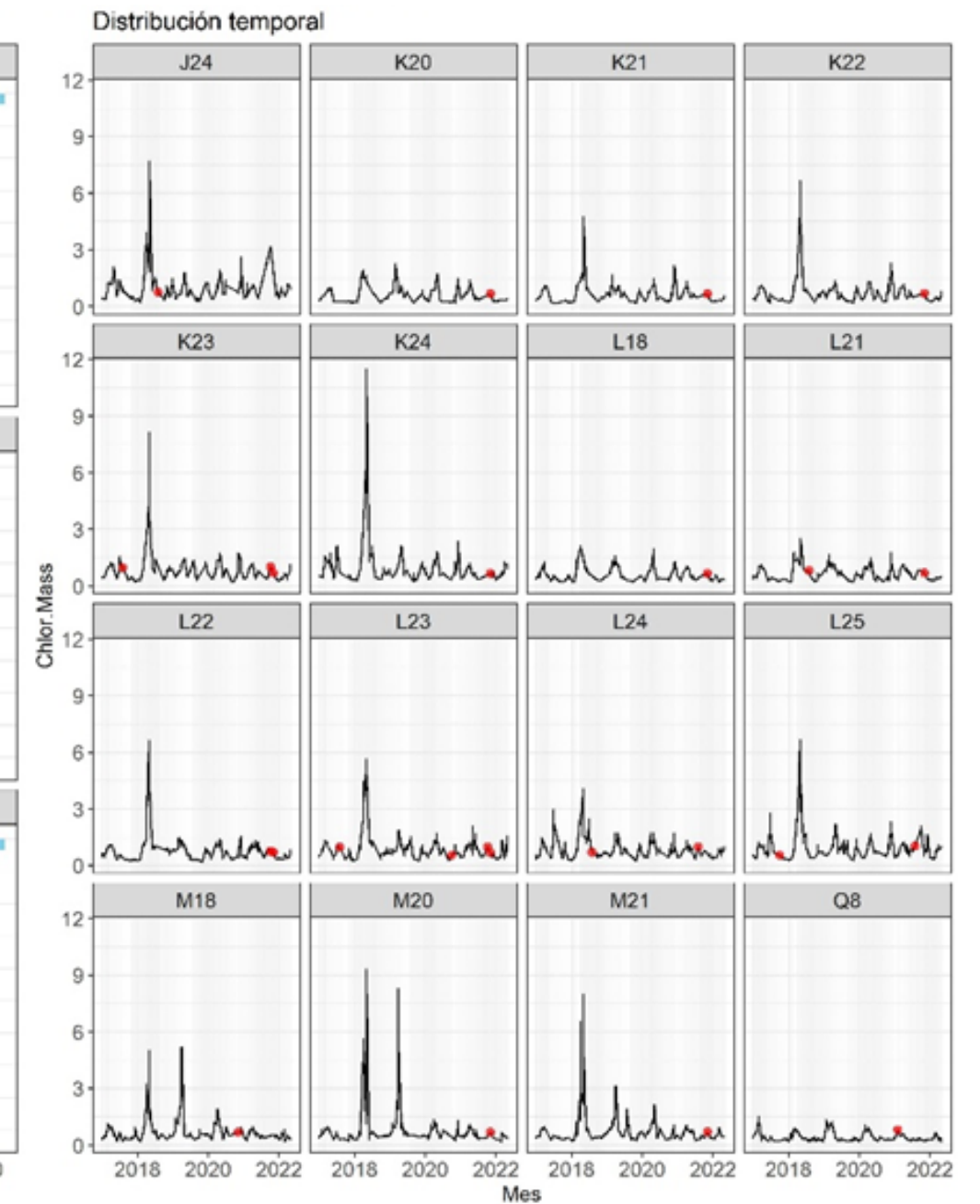
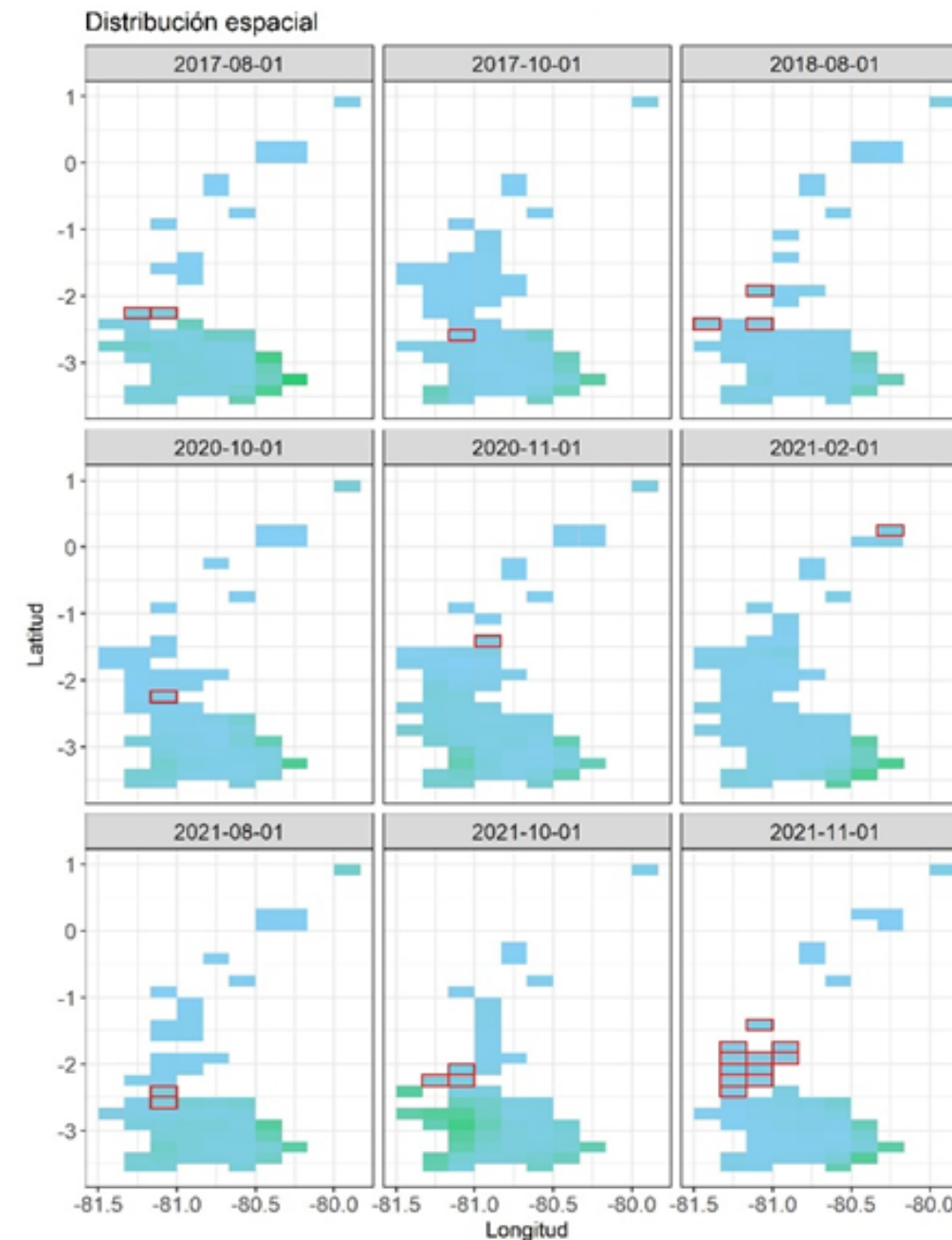
(*) Min representa el mínimo; max representa el máximo; Q1, Q2 y Q3 representan los cuartiles 1, 2 y 3. DE representa al valor de la desviación estándar. |

5. Preparación, validación y filtros de datos

3

VALIDACIÓN Y FILTROS DE DATOS

- **Imputación** de 68 valores faltantes de la **concentración de clorofila**, a través del método predictivo de suavizado Ponderación de Distancia Inversa (IDW) con parámetro de potencia $\alpha=2$
- Así, tras realizar los filtros e imputaciones indicadas el número de registros del **conjunto de datos de trabajo se redujo a 1,836** observaciones.



6. Análisis exploratorio y selección de variables

1

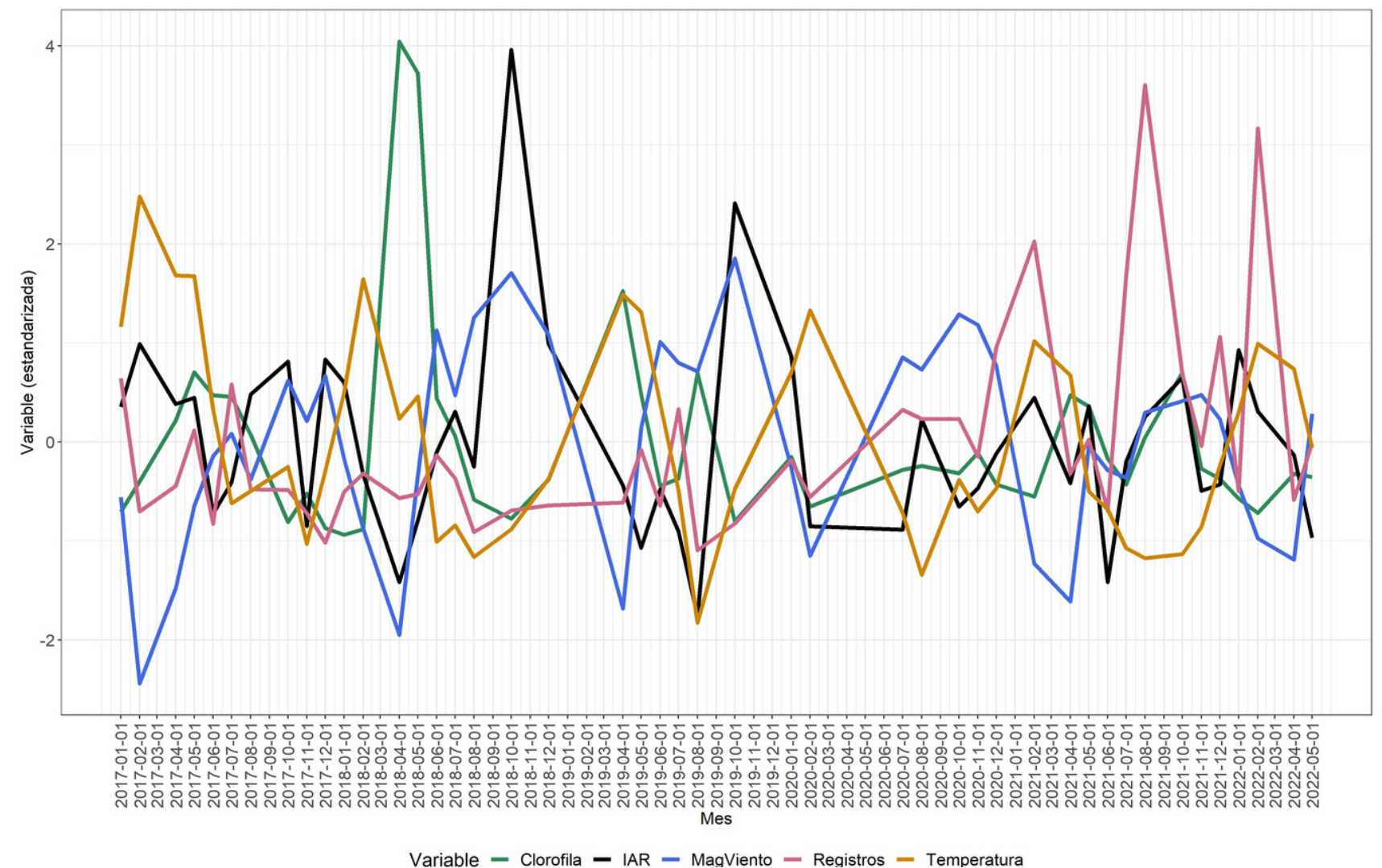
DIMENSIÓN TEMPORAL

Los años con mayor número de registros fueron 2021, 2020 y 2017. Mientras que, el año con menor número de registros fue el 2019. A nivel de meses, **febrero de 2022 y agosto de 2021** tuvieron mayor número de registros. A nivel de promedio estandarizado, los **incrementos de abundancia relativa** han ocurrido **cuando la temperatura ha disminuido**. Los meses con **mayor IAR** han sido **2018-10 y 2019-10**.

2

DIMENSIÓN ESPACIAL

Las áreas con mayor número de registros han sido las ubicadas en las longitudes **N y M** y latitudes **entre 28 y 27**, en general y para la mayoría de los meses. Sin embargo, las que presentaron **mayor IAR** fueron la **L30, L31 y P30**. En términos generales, el sur ha sido el más frecuente y ha ido en **incremento**.



6. Análisis exploratorio y selección de variables

3 VARIABLES PARA EL MODELO

- La concentración de **clorofila** varió entre 0.17 y 7.64 mg/m³, pero es **altamente sesgada**. La **temperatura** de la superficie del agua estuvo dentro del rango de 18°C y 29.63°C. La **magnitud del viento** varió entre 1.73 m/s y 7.15 m/s. Las dos últimas presentaron un comportamiento **aproximadamente simétrico**, respecto a sus medias.
- La CPUE varió entre 0 y 54 toneladas por lance, y el IAR indica que la **abundancia promedio es 0.18, independientemente del área o mes de captura**. La transformación de raíz cuadrada de IAR mantiene la dimensión, monotonía, y por tanto, interpretación.
- Todas las variables presentaron **relación lineal débil** con sqrtIAR.

Variable	Min	Max	Q1	Q2	Q3	Media	DE	Sesgo
Clorofila	0.17	7.64	0.50	0.70	0.97	0.80	0.51	5.11
Temperatura	18	29.13	23.24	24.28	26.12	24.66	1.69	0.25
Mag. Viento	1.73	7.15	4.27	4.63	4.96	4.59	0.63	-0.39
CPUE	0.00	54.00	2.00	5.00	10.00	8.56	9.41	1.92
IAR	0.00	1.00	0.04	0.11	0.25	0.18	0.19	1.86
sqrtIAR	0.00	1.00	0.21	0.33	0.50	0.37	0.21	0.78

(*) Min representa el mínimo; max representa el máximo; Q1, Q2 y Q3 representan los cuartiles 1, 2 y 3. DE representa al valor de la desviación estándar.

ρ_s / p-value	Clorofila	Temperatura	Mag. Viento	sqrtIAR
Clorofila	1.00	<0.001	<0.001	<0.001
Temperatura	-0.37	1.00	<0.001	<0.001
Mag. Viento	0.26	-0.56	1.00	<0.001
sqrtIAR	-0.09	0.08	-0.11	1.00

(*) ρ_s es el coeficiente de correlación de Spearman.

7. Construcción, prueba y selección del modelo

1 MUESTRAS EN EL MODELO

- *Entrenamiento: 1,088 ejemplos de 2017-1 a 2021-5 (59.3%)*
- *Validación: 373 ejemplos de 2021-6 a 2021-11 (20.3%)*
- *Prueba : 375 ejemplos de 2021-12 a 2022-5 (20.4%)*

2 CARACTERÍSTICAS DEL MODELO

- *Tipo: Perceptrón multicapas*
- *Número de capas: 4*
- *Nodos de entrada: 6*
- *Nodos en capa oculta 1: 32*
- *Nodos en capa oculta 2: 7*
- *Nodos de salida: 1*
- *Funciones de activación: ReLu, Lineal*
- *Total de parámetros: 463 (224, 231, 8)*

- *Épocas: 50*
- *Tamaño del batch: 32*
- *Optimizador: RMSProp*
- *Función de pérdida: MSE*
- *Métrica: MAE*

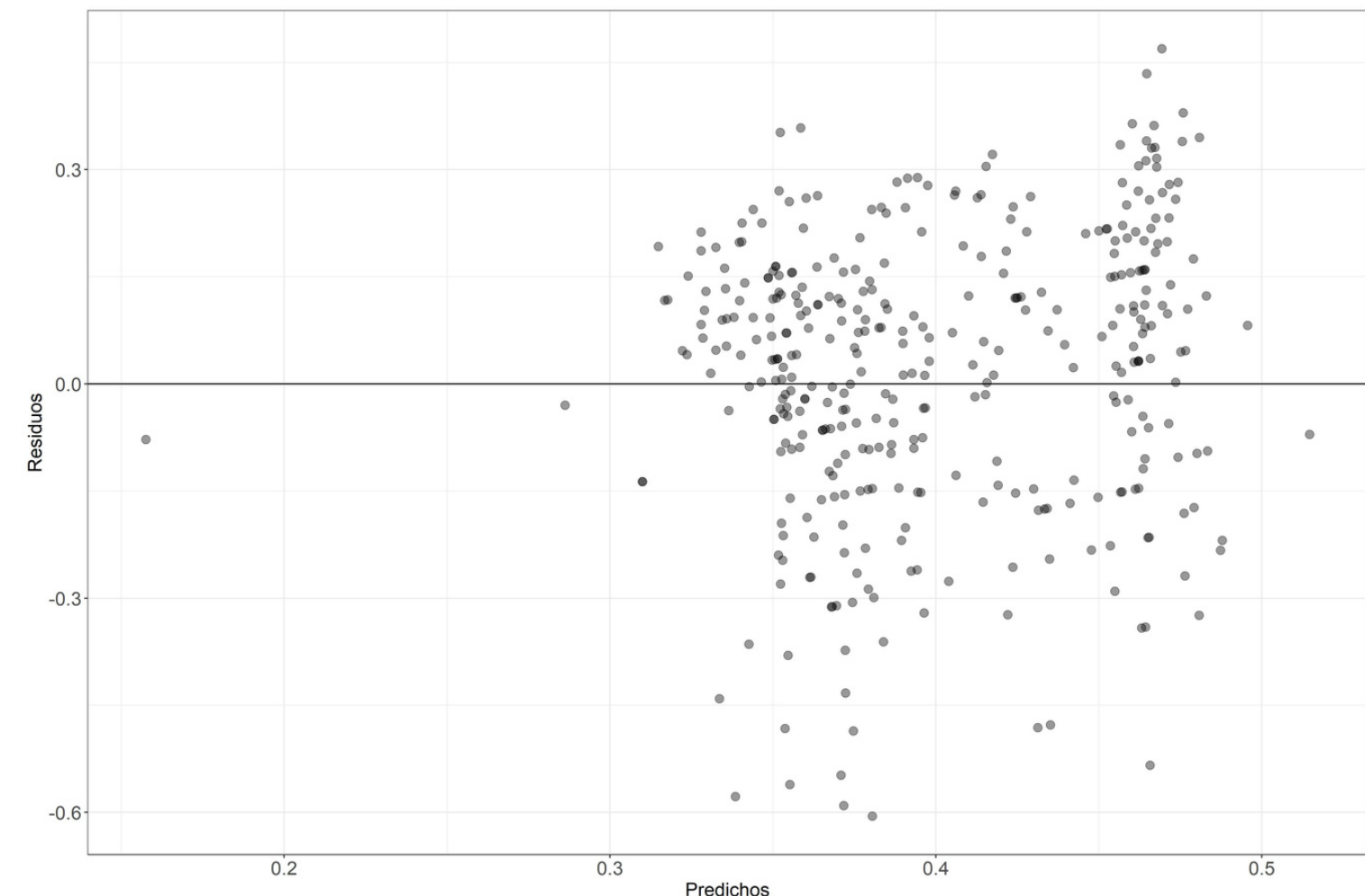
(*) Todo el análisis se realizó en R versión 4.1.2, haciendo uso de la interfaz RStudio. Los modelos se realizaron a través de Tensorflow y Keras para R.

7. Construcción, prueba y selección del modelo

3 DESEMPEÑO DEL MODELO

- En promedio, el valor de **RMSE** en la muestra de prueba fue **0.2** y el **MAE de IAR** predicho fue en promedio de **0.02**.
- El **porcentaje no explicado** por el modelo es en promedio 51%.
- Se puede observar que el modelo tiende a la **subestimación** en ciertos puntos.

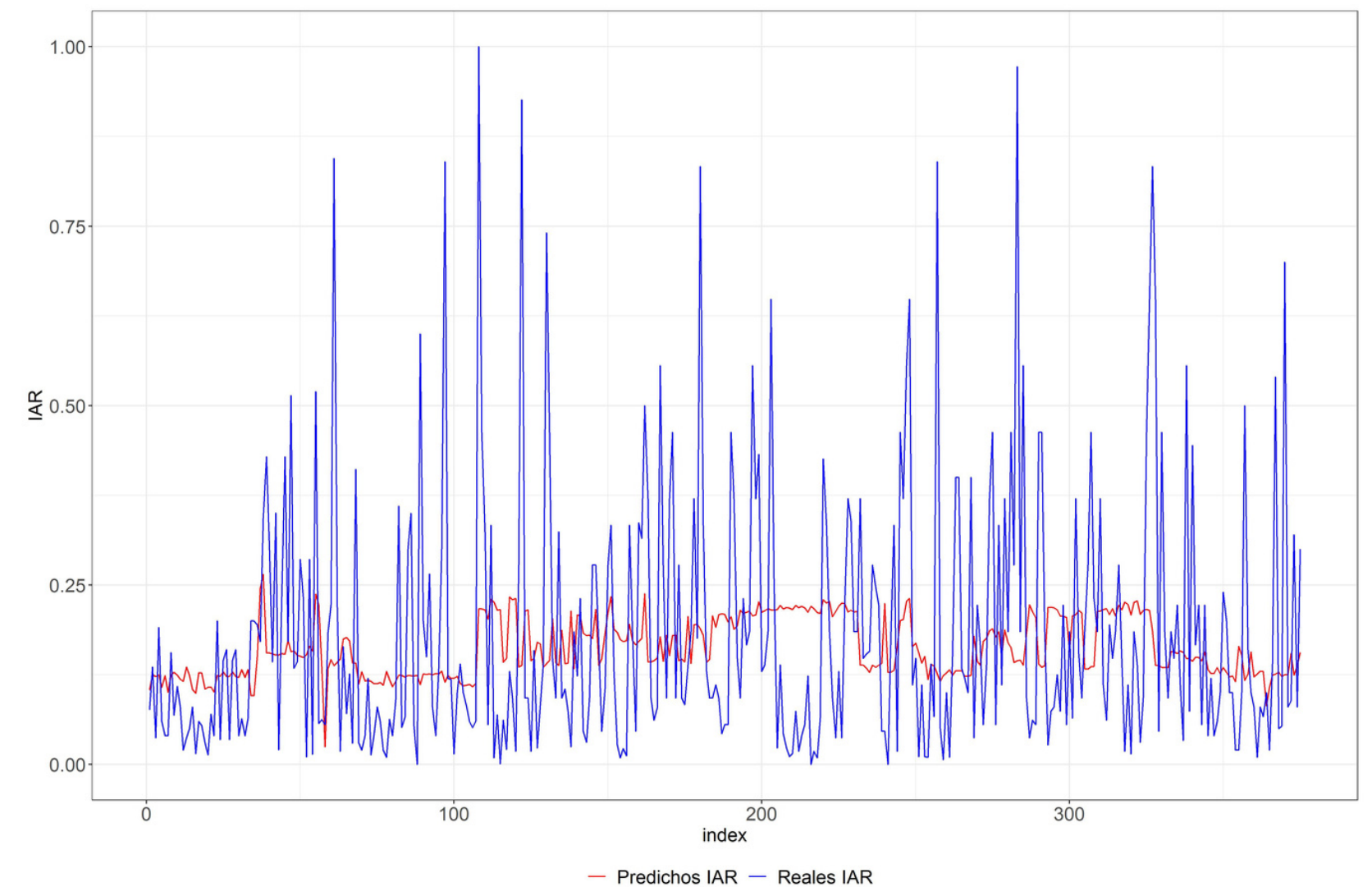
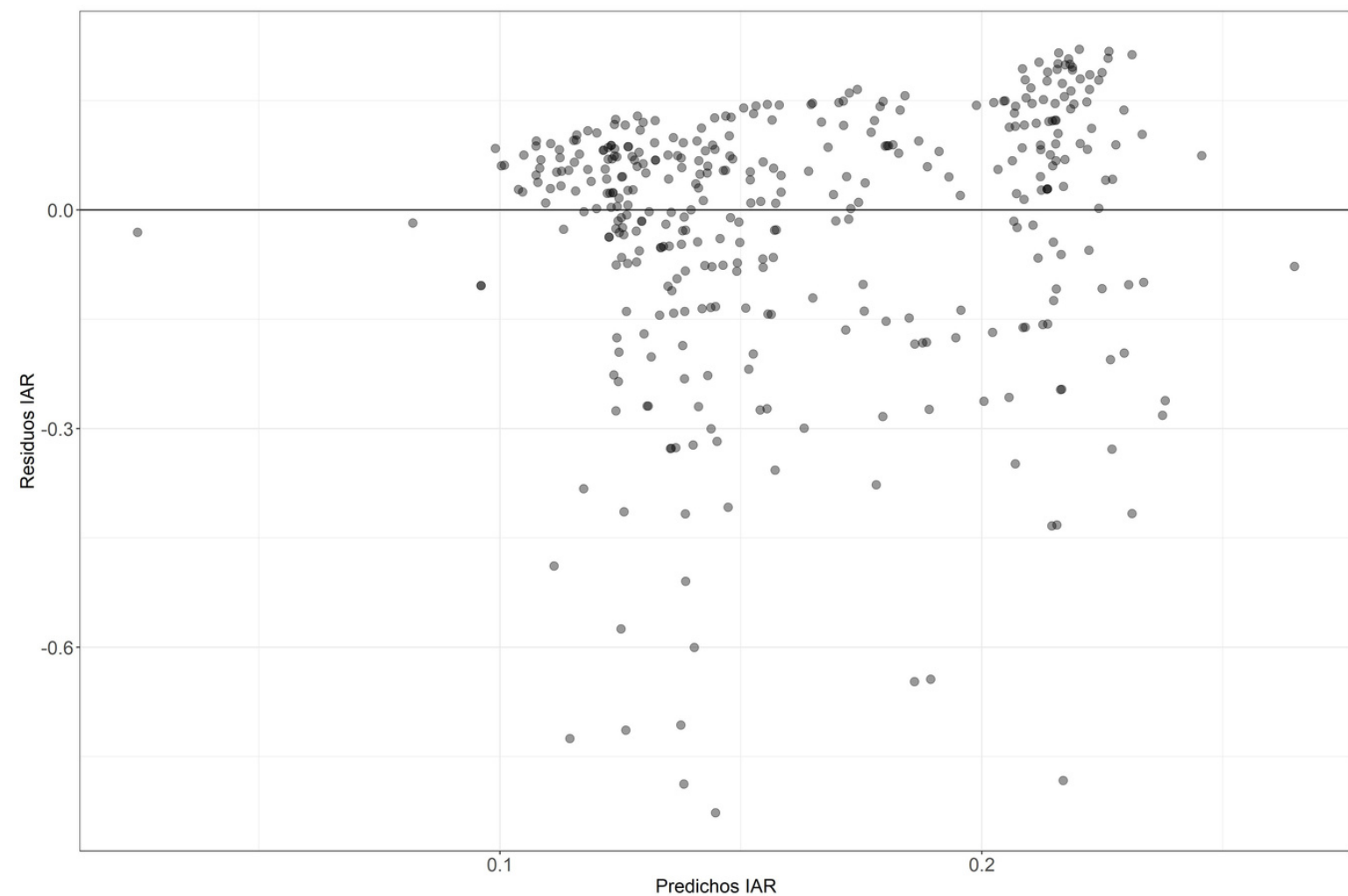
Indicador	Muestra	Iteración										Promedio
		1	2	3	4	5	6	7	8	9	10	
MSE	Train	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	Val	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	Test	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
MAE (sqrtIAR)	Train	0.17	0.17	0.16	0.16	0.17	0.17	0.16	0.17	0.17	0.17	0.17
	Val	0.16	0.16	0.17	0.16	0.16	0.16	0.16	0.17	0.16	0.16	0.16
	Test	0.15	0.15	0.17	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16
RMSE	Train	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
	Val	0.21	0.21	0.21	0.21	0.21	0.20	0.21	0.21	0.20	0.20	0.21
	Test	0.19	0.19	0.20	0.19	0.19	0.19	0.19	0.20	0.20	0.20	0.20
MAE (IAR)	Train	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	Val	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	Test	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.02
RMSE _{test} / \bar{y}_{test}		0.50	0.51	0.53	0.51	0.51	0.51	0.51	0.52	0.51	0.52	0.51



7. Construcción, prueba y selección del modelo

3 DESEMPEÑO DEL MODELO

Al transformar los residuos a la dimensión original, se observa que varios casos en la muestra de prueba fueron **predichos con una abundancia de más de 0.6 unidades por debajo del valor real**.

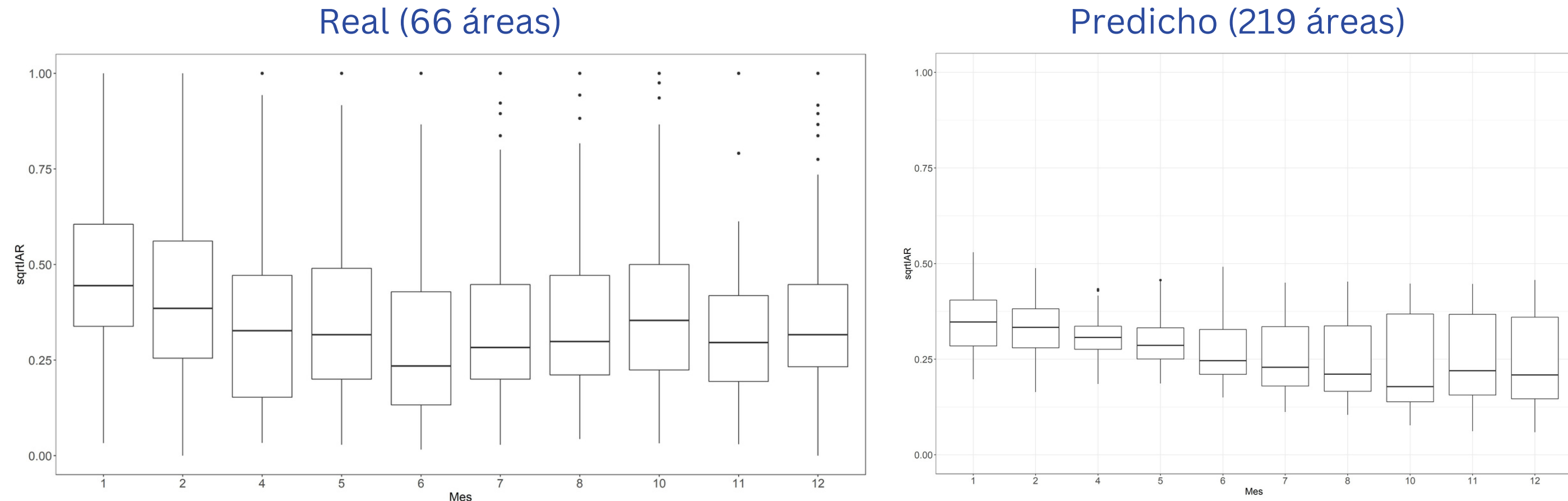


8. Predicciones y construcción de mapas

1

PREDICCIONES

Al usar la información oceanográfica mensual del total de áreas codificadas por IPIAP (219), se comparó las distribuciones mensuales reales frente a las predichas de la variable sqrtIAR y se observó que **el modelo aprendió sobre la tendencia de los meses, pero los subestima**. Esto afecta principalmente a las predicciones de Enero y Febrero

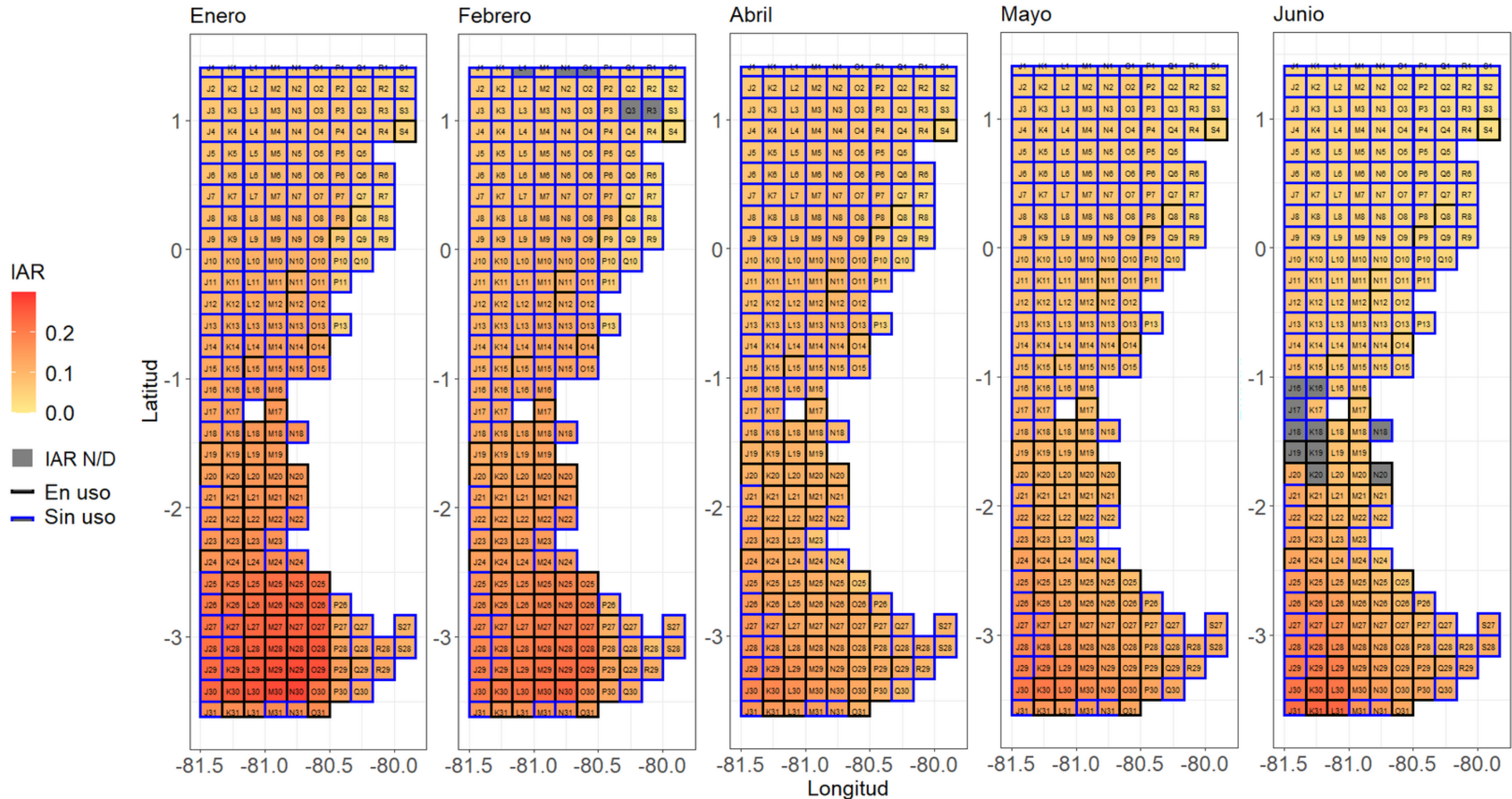


8. Predicciones y construcción de mapas

2 MAPAS

Varias áreas predichas con mayor abundancia han sido visitadas. No obstante, algunas con poca frecuencia.

Las cuadrículas de la J25 a la J31 y demás latitudes en la línea 31, no han sido exploradas y se predicen como abundantes.

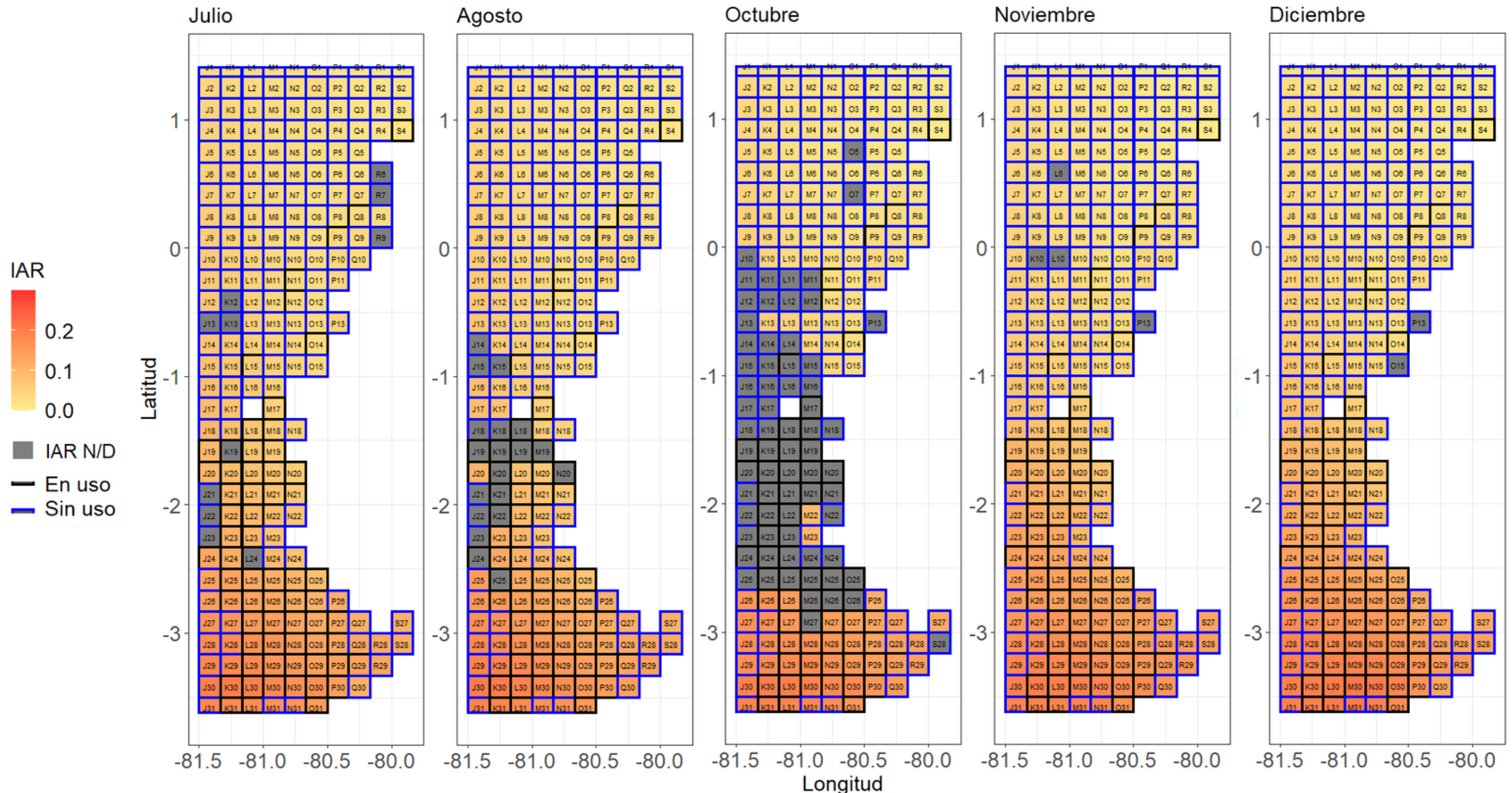


8. Predicciones y construcción de mapas

2 MAPAS

En los meses de julio a agosto existen **áreas no predichas**, debido a los valores **faltantes** en la concentración de clorofila.

Se observa una **menor abundancia** en el **segundo semestre** del año, tal y como ocurre en la realidad



CONCLUSIONES

- La pesca de *S. japonicus* ha **registrado** solo un **30% del total de áreas** de la costa ecuatoriana, de las cuales, las de **mayor abundancia** relativa desde el 2017 han sido las codificadas como **L30 y L31**, mientras que las **más concurridas** por los barcos pesqueros fueron las áreas **N y M entre 28 y 27**, independientemente del mes.
- **Febrero y enero** han sido los **meses más concurridos** y los de **mayor abundancia** relativa.
- Las zonas predichas como las más **abundantes** están **al sur** de Ecuador (Castro Hernández et al. 2000).
- El **modelo explica un 49%** de la VT y su principal debilidad es la estimación de las altas abundancias. Pese a ello, **aprendió sobre la estacionalidad y vecindad** espacial.
- **No hubo relación lineal** entre las variables de entrada y la de respuesta. Pero, se observó una **relación inversa entre temperatura e IAR**, a lo largo de los meses (Canales y Jurado 2021).
- Las **zonas potencialmente de pesca** están **entre la J25 y la J31**, así como las de la **fila 31**. Además de las áreas visitadas con poca frecuencia dentro del sur de la costa ecuatoriana.

- Al realizar las pruebas de modelos se concluyó que un **modelo de menor complejidad** obtuvo mejores resultados (Armas et al. 2022; Wang et al. 2015). Así como, un **menor tamaño de batch** y el uso de la función **ReLU**, mejoraron el **rendimiento y eficiencia** del modelo (Masters y Luschi 2018).
- Este estudio ha permitido reflexionar sobre la importancia de **toma de mediciones** oceanográficas **a bordo**, y de la implementación de un **muestreo** de barcos **adecuado** que asegure la homogeneidad o heterogeneidad controlada de los barcos.
- Los principales **desafíos** se encontraron en la obtención de datos pesqueros, en la unión de los datasets y la búsqueda de un modelo
- Los **mapas** construidos impactan **positivamente al ODS14**, siempre y cuando sean empleados como herramientas para IPIAP.
- Las principales **limitaciones** encontradas están en la **calidad y tamaño del conjunto de datos**. Además de la falta de exploración de modelos dedicados a explorar relaciones espaciotemporales subyacentes.
- No obstante, este estudio representa una **primera aproximación** hacia la **exploración de la distribución de la abundancia** de los recursos pesqueros en **Ecuador**, abriendo oportunidades de mejora en el muestreo, procesamiento y modelamiento de los datos.

BIBLIOGRAFÍA

- Armas, Elier, Hugo Arancibia, y Sergio Neira. 2022. "Identification and Forecast of Potential Fishing Grounds for Anchovy (*Engraulis ringens*) in Northern Chile Using Neural Networks Modeling". *Fishes* 7 (4).
- Beysolow II, Taweh. 2017. "Introduction to Deep Learning". En *Introduction to Deep Learning Using R: A Step-by-Step Guide to Learning and Implementing Deep Learning Models Using R*, editado por Taweh Beysolow II, 1-9. Berkeley, CA: Apress.
- Canales, Cristian M, y Viviana Jurado. 2021. "Evaluación del stock de recursos pelágicos pequeños del Ecuador. Año 2021". *Technical*. Guayaquil, Ecuador: Instituto Público de Investigación de Acuicultura y Pesca.
- Castro Hernández, José J., y Ana T. Santana Ortega. 2000. *Synopsis of Biological Data on the Chub Mackerel (Scomber Japonicus Houttuyn; 1782)*. Rome, Italy: FAO Fisheries Synopsis.
- Chen, Xinjun, Gang Li, Bo Feng, y Siquan Tian. 2009. "Habitat suitability index of Chub mackerel (*Scomber japonicus*) from July to September in the East China Sea". *Journal of Oceanography* 65 (febrero): 93-102.
- Chollet, François, Tomasz Kalinowski, y Joseph J Allaire. 2022. *Deep Learning with R*. 2nd ed. New York, NY: Manning.
- FAO. 2022. *The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation. The State of World Fisheries and Aquaculture (SOFIA) 2022*. Rome, Italy: FAO.
- Huang, Guang-Bin. 2003. "Learning capability and storage capacity of two-hidden-layer feedforward networks". *IEEE Transactions on Neural Networks* 14 (2): 274-81.
- Hubert, Wayne A, y Mary C Fabrizio. 2007. "Relative Abundance and Catch per Unit Effort". En *Analysis and Interpretation of Freshwater Fisheries Data*, 279-325. Bethesda, Maryland: American Fisheries Society.
- Plant, Richard E. 2018. *Spatial Data Analysis in Ecology and Agriculture Using R*. 2a ed. Boca Raton: CRC Press.
- Wang, Jintao, Wei Yu, Xinjun Chen, Lin Lei, y Yong Chen. 2015. "Detection of potential fishing zones for neon flying squid based on remote-sensing data in the Northwest Pacific Ocean using an artificial neural network". *International Journal of Remote Sensing* 36 (13): 3317-30.
- Wilcox, Rand R. 2013. "Some Outlier Detection Methods". En *Introduction to Robust Estimation and Hypothesis Testing*, 3rd ed., 96-100. St. Louis: Elsevier Science & Technology.

