

Escuela Superior Politécnica del Litoral

Facultad de Ingeniería Marítima y Ciencias Del Mar

Validación de datos de salinidad obtenidos de Copernicus Global Ocean

Physics Reanalysis para uso del IPIAP

INGE-2823

Proyecto Integrador

Previo a la obtención del Título de:

Ingeniero Oceanográfico

Presentado por:

Melissa Lizbeth Gómez Tapia

Guayaquil – Ecuador

Año: 2025

Dedicatoria

A mis padres Ana Tapia y Manuel Gómez,
por acompañarme en este largo camino,
por alentarme y motivarme cuando más lo
necesitaba. Gracias por creer en mí aun
cuando dudaba, por sostenerme en los
momentos difíciles y por celebrar conmigo
cada logro.

A mi hermana Paulina, quiero que este
trabajo sea para ti un reflejo de esfuerzo y
perseverancia, y de que todo lo que uno
se propone lo puede alcanzar.

Este logro no es solo mío, sino también de
ustedes que me acompañaron en este
recorrido. Les prometo que este será el
primero de muchos.

Con amor,

Melissa Gómez Tapia

Agradecimientos

Agradezco a Dios por siempre darme fuerzas para continuar, por guiarme en este largo camino y permitirme culminar esta etapa.

Mi más profundo agradecimiento al M.Sc. Telmo de la Cuadra, a la Química Patricia Macías, al Ing. Julio Prado y futura colega Paula Espinoza, sin su apoyo este proyecto no sería igual. Gracias por todas las veces que me corrigieron y me motivaron cuando más lo necesitaba, nunca olvidaré el infinito apoyo me brindaron. De igual manera agradecer al M.Sc. Jonathan Cedeño quien fue un gran tutor y estuvo dispuesto a resolver mis dudas a lo largo de todo este proyecto.

A Benjamín C. por ser una fuente de apoyo e inspiración, gracias por tomar mi mano a lo largo de este camino y motivarme cuando más lo necesité y por los valiosos momentos que hicieron que todo sea más sencillo. Este logro no sería igual sin ti.

Y a mis mejores amigos: Sofía B., Katherine O., Thomas N., y Jhossue C.,

quienes iniciaron este recorrido conmigo
y estuvieron en tantos momentos en los
que sentí rendirme, gracias por su apoyo,
por motivarme a continuar y por las risas
que jamás faltaron.

Declaración Expresa

Yo Melissa Lizbeth Gómez Tapia acuerdo y reconozco que:

La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por mí/nosotros durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me/nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi/nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique al autor que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 19 de septiembre del 2025.

Melissa Lizbeth Gómez Tapia

Evaluadores

Indira Yadira Nolivos Álvarez,
PhD.
Profesor de Materia Integradora

Jonathan Marcelo Cedeño
Oviedo,
MSc.
Tutor de Proyecto Integrador

Resumen

El proyecto tiene como objetivo validar datos de salinidad del reanálisis global de Copernicus en la costa ecuatoriana frente a Puerto López y Salinas, bajo la hipótesis de que las estimaciones modeladas reproducen las variaciones locales, pero con sesgos sistemáticos. Es necesario comprobar la representatividad local de los datos del modelo de reanálisis de salinidad con datos in-situ proporcionados por el Instituto Público de Investigación de Acuicultura y Pesca, ya que su validación fortalecerá el monitoreo oceanográfico mensual. Para ello, se recopilieron perfiles de salinidad con CTD in-situ (2022–2025) según protocolos CPPS y se obtuvieron datos del modelo de Copernicus en el mismo periodo. Posteriormente, las salinidades modeladas fueron interpoladas para compararlas con mediciones a profundidades estándar y en la ubicación deseada, mientras que las discrepancias fueron evaluadas con métricas de error y correlación. Los resultados indicaron que el reanálisis reproduce el ciclo estacional observado con alta correlación ($R \approx 0,83$). Sin embargo, presenta un sesgo negativo, subestimando la salinidad y mostrando mayores errores en la capa superficial. En aguas subsuperficiales (>50 m), el error es mínimo (RMSE = 0.08). En conclusión, Copernicus refleja la variabilidad salina costera, aunque sus estimaciones superficiales requieren ajustes locales.

Palabras clave: Reanálisis global, Subestimación, Salinidad, Sesgos sistemáticos, Métricas de error.

Abstract

The project aims to validate salinity data from the Copernicus global reanalysis along the Ecuadorian coast off Puerto López and Salinas, under the hypothesis that the modeled estimates reproduce local variations but with systematic biases. It is necessary to verify the local representativeness of the salinity reanalysis model data with in-situ measurements provided by the Public Research Institute of Aquaculture and Fisheries, since their validation will strengthen the monthly oceanographic monitoring. For this purpose, salinity profiles were collected with in-situ CTD (2022–2025) following CPPS protocols, and Copernicus model data were obtained for the same period. Subsequently, the modeled salinities were interpolated to compare them with measurements at standard depths and at the desired location, while discrepancies were evaluated using error and correlation metrics. The results indicated that the reanalysis reproduces the observed seasonal cycle with high correlation ($R \approx 0.83$). However, it presents a negative bias, underestimating salinity and showing greater errors in the surface layer. In subsurface waters (>50 m), the error is minimal (RMSE = 0.08). In conclusion, Copernicus reflects coastal salinity variability, although its surface estimates require local adjustments.

Keywords: Global reanalysis, Underestimation, Salinity, Systematic biases, Error metrics.

Contenido

1. Introducción	2
1.1 Planteamiento del Problema	2
1.2 Justificación.....	3
1.3 Objetivos	5
1.3.1 <i>Objetivo General</i>	5
1.3.2 <i>Objetivos Específicos</i>	5
1.4 Marco Teórico.....	5
1.4.1 <i>Área de Estudio</i>	5
1.4.2 <i>Salinidad como variable oceanográfica y su relación con la conductividad</i>	7
1.4.4 <i>Variabilidad estacional de la salinidad</i>	10
2. Metodología	13
2.1 Esquema de Análisis	13
2.2 Obtención de datos in-situ.....	14
2.4 Esquema de interpolación vertical.....	15
2.4.2 <i>Interpolación polinómica de segundo y tercer grado</i>	17
2.4.3 <i>Interpolación spline cúbico</i>	17
2.4.4 <i>Suavizado local LOESS (Locally Weighted Scatterplot Smoothing)</i>	17
2.5 Esquema de interpolación horizontal bidimensional	18
2.5.1 <i>Vecino más cercano</i>	19
2.5.2 <i>Interpolación por promedio de k vecinos más cercanos</i>	19
2.5.3 <i>Regresión espacial</i>	19
2.5.4 <i>Interpolación por Inverse Distance Weighting (IDW)</i>	20
2.5.5 <i>Kriging ordinario</i>	20
2.6 Validación de los resultados: Análisis exploratorio y pruebas no paramétricas	21
2.6.1 <i>Análisis exploratorio de precisión y exactitud</i>	21
2.6.2 <i>Error medio (Bias)</i>	22
2.6.3 <i>Error cuadrático medio (RMSD)</i>	22
2.6.5 <i>Error porcentual absoluto medio (MAPE)</i>	22
3. Resultados y Análisis	25
3.1 Ciclo anual de salinidad en Puerto López y Salinas (2022–2025)	25
3.2 Contraste: Época húmeda vs. época seca	27
3.3 Variación con la profundidad: Capa subsuperficial (50 – 100 m)	30
3.4 Indicadores de error y desempeño del modelo Copernicus por temporada	32
3.5 Variabilidad del error de salinidad en función de la profundidad	35
3.6 Estimación del Modelo	38
4. Conclusiones y recomendaciones.....	45

4.1 Conclusiones.....	45
4.2 Recomendaciones.....	47
Bibliografía.....	
Apéndice.....	

Abreviaturas

AES	Agua Ecuatorial Superficial
AESS	Agua Ecuatorial Subsuperficial
ASTS	Agua Subtropical Superficial
ATS	Agua Tropical Superficial
CMEMS	Servicio de Monitoreo Marino de Copernicus (Copernicus Marine Env. Monitoring Service)
COI	Comisión Oceanográfica Intergubernamental
CPPS	Comisión Permanente del Pacífico Sur
CTD	Conductivity-Temperature-Depth
ENSO	El Niño Oscilación Sur
EOS-80	Ecuación de Estado de 1980
IDW	Interpolación por Distancia Inversa
IPIAP	Instituto Público de Investigación de Acuicultura y Pesca
IPTS-68	Escala Internacional de Temperatura Práctica de 1968
ITS-90	Escala Internacional de Temperatura de 1990
LOESS	Locally Weighted Scatterplot Smoothing
PSS-78	Practical Salinity Scale 1978
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
ZCIT	Zona de Convergencia Intertropical

Índice de figuras

<i>Figura 1. Mapa del Área de Estudio en base a las estaciones Puerto López y Salinas.....</i>	<i>7</i>
<i>Figura 2. Esquema de análisis</i>	<i>13</i>
<i>Figura 3. Climatología Mensual de Salinidad 1997-2025</i>	<i>26</i>
<i>Figura 4. Promedios mensuales de salinidad superficial en Salinas (2023-2025).</i>	<i>28</i>
<i>Figura 5. Promedios mensuales de salinidad superficial en Puerto López (2023-2025).....</i>	<i>28</i>
<i>Figura 6. Promedios mensuales de Salinidad Subsuperficial en Salinas (2023-2025).....</i>	<i>29</i>
<i>Figura 7. Promedios mensuales de Salinidad Subsuperficial en Puerto López (2023-2025)</i>	<i>29</i>
<i>Figura 8. Indicadores de Error: In situ vs Copernicus.....</i>	<i>33</i>
<i>Figura 9. Errores por Época (Húmeda y Seca).....</i>	<i>33</i>
<i>Figura 10. Error MAPE por Temporada (Húmeda y Seca).....</i>	<i>34</i>
<i>Figura 11. Error MAPE del modelo por profundidad</i>	<i>36</i>
<i>Figura 12. Errores MAE Y RMSD por profundidad</i>	<i>37</i>
<i>Figura 13. Comparación de Subestimación y Sobreestimación del Modelo</i>	<i>39</i>
<i>Figura 14. Gráfico de Correlación: Datos In Situ vs Copernicus.....</i>	<i>40</i>
<i>Figura 15. Correlación por estacionalidad</i>	<i>41</i>
<i>Figura 16. Correlación por profundidad.....</i>	<i>42</i>

Índice de tablas

<i>Tabla 1 Métodos de Interpolación Geoespacial más utilizados con su porcentaje de error.</i>	<i>21</i>
<i>Tabla 2 Costos de la validación de datos de salinidad del modelo de reanálisis vs. in-situ.</i>	<i>43</i>

CAPÍTULO 1

1. Introducción

1.1 Planteamiento del Problema

El Instituto Público de Investigación de Acuicultura y Pesca (IPIAP), como organismo técnico-científico del Estado, tiene la responsabilidad de generar información precisa para el manejo sostenible de los ecosistemas marinos y costeros del Ecuador. Para cumplir esta función, requiere datos oceanográficos confiables y con resolución suficiente que permitan caracterizar adecuadamente las condiciones del medio marino. El IPIAP realiza monitoreos directos mediante campañas oceanográficas, estos presentan limitaciones en cuanto a su cobertura temporal y espacial, debido a factores logísticos, presupuestarios y operativos. Con el objetivo de complementar esta información, la institución ha incorporado productos de *Copernicus Marine Service*, una plataforma que provee datos oceanográficos de alcance global. Entre ellos, se encuentra el producto *Global Ocean Physics Reanalysis*, que proporciona estimaciones diarias de variables físicas como la salinidad del mar, con una resolución espacial de hasta $1/12^\circ$. Este tipo de producto representa una herramienta valiosa para análisis científicos, modelos oceanográficos y evaluaciones de recursos marinos. Sin embargo, los datos generados mediante modelos numéricos y técnicas de asimilación de datos pueden presentar errores, especialmente en regiones costeras donde la resolución de los modelos y la disponibilidad de observaciones locales influyen significativamente en la calidad de los resultados. Estudios previos han mostrado que los productos de reanálisis pueden presentar sesgos sistemáticos o subestimaciones/localizaciones erróneas por ejemplo de gradientes salinos, particularmente en zonas tropicales con fuerte influencia continental, como la costa ecuatoriana (Dossa et al., 2021).

A pesar del uso del producto dentro del IPIAP, no se ha realizado hasta el momento una evaluación formal de la incertidumbre del producto de salinidad de Copernicus en el contexto local, ni se ha implementado un procedimiento de validación sistemática. La ausencia de un proceso de verificación pone en duda la confiabilidad de la información que actualmente se utiliza como base para investigaciones científicas, análisis pesqueros y

aplicaciones de modelación numérica. Esta situación evidencia la necesidad de desarrollar un estudio que permita comparar los datos de salinidad proporcionados por Copernicus con mediciones in situ obtenidas en campañas oceanográficas, con el fin de identificar posibles errores y cuantificar su magnitud y para ajustes que mejoren su representatividad en aguas ecuatorianas.

La ejecución de un trabajo de estas características permitirá al IPIAP fortalecer su capacidad técnica asegurando mayor precisión en sus análisis. En este sentido, el problema central que se aborda en este proyecto integrador es la incertidumbre sobre la confiabilidad de los datos de salinidad del producto Copernicus en la región costera del Ecuador, y la falta de una metodología institucional validada conforme a las condiciones locales. Abordar esta necesidad contribuirá significativamente a mejorar el monitoreo marino nacional y optimizar el uso de herramientas oceanográficas dentro de la planificación científica y operativa del país.

1.2 Justificación

Generalmente en Ecuador se utilizan datos de salinidad derivados de modelos como el *Global Ocean Physics Reanalysis* de Copernicus, sin embargo, no se han encontrado estudios que hayan aplicado algún tipo de validación, considerando la complejidad de las características oceanográfica de la región (sistema de corrientes y eventos extremos). Esta incertidumbre implicaría que se desconoce el grado de exactitud de los datos arrojados por el modelo en la zona.

La salinidad es una de las principales variables físico-químicas que rigen la dinámica del océano, ya que influye en la densidad del agua, la circulación oceánica y la distribución de nutrientes y organismos marinos. En el contexto costero del Ecuador, caracterizado por la convergencia de masas de agua de distinta naturaleza y por la influencia de fenómenos climáticos como El Niño-Oscilación del Sur (ENOS), el conocimiento preciso de la salinidad permite a los oceanógrafos del Instituto Público de Investigación de Acuicultura y Pesca (IPIAP) comprender mejor los procesos que afectan los ecosistemas marinos, como la

dispersión larval, la productividad biológica y la variabilidad en las pesquerías (Riofrío & Chavarría, 2004; Perugachi-Salamea et al., 2014).

El uso de datos satelitales y de reanálisis oceánico, como los provistos por el producto Copernicus *Global Ocean Physics and Reanalysis*, ha permitido extender el monitoreo marino a escalas regionales y temporales más amplias. No obstante, este tipo de información, generada mediante modelos numéricos y técnicas de asimilación de datos, puede presentar errores sistemáticos o subestimaciones significativas en zonas costeras, donde la resolución del modelo y la escasez de datos in situ impactan en la calidad de la estimación (Dossa et al., 2021). Por ello, validar estos productos con datos reales recolectados en campo es una necesidad urgente para asegurar la confiabilidad de su uso en análisis oceanográficos, planificación pesquera y toma de decisiones ambientales.

En este sentido, el IPIAP ejecuta campañas oceanográficas donde se obtienen mediciones con dispositivos CTD (conductividad, temperatura y presión, por sus siglas en inglés), complementadas con muestras tomadas mediante botellas Niskin para control de calidad. Estas actividades permiten contrastar los valores estimados por modelos globales con mediciones reales y contextualizadas. Implementar una validación sistemática entre ambos tipos de datos permitirá establecer el grado de precisión, exactitud y confiabilidad del producto Copernicus en la zona costera del país, aportando no solo a la mejora de los análisis científicos, sino también a la generación de insumos más precisos para la gestión pesquera y ambiental en Ecuador.

Este proyecto no solo responde a la necesidad técnica de evaluar la representatividad de los datos satelitales en condiciones locales, sino que también se alinea con estándares internacionales que promueven la validación cruzada de observaciones remotas mediante datos *in-situ* como parte fundamental de la vigilancia del océano (UNESCO-IOC, 2019). Además, fortalece el enfoque de soberanía científica y autosuficiencia técnica del IPIAP, al promover el desarrollo de capacidades internas para evaluar críticamente las herramientas de observación marina disponibles. Al validar localmente los datos de salinidad satelital mediante campañas oceanográficas en Ecuador,

este proyecto fortalece las capacidades de investigación marina nacional, mejora la calidad de la información científica utilizada en la gestión pesquera y contribuye a la reducción de la dependencia de modelos sin ser validados. Asimismo, promueve la aplicación de tecnología marina y el uso de metodologías estandarizadas para el monitoreo del océano, alineándose con los principios del ODS 14.a y posicionando al Ecuador en el camino hacia una gestión sostenible y basada en evidencia de sus recursos marinos (Naciones Unidas, 2015).

1.3 Objetivos

1.3.1 Objetivo General

Validar los datos de salinidad del producto “Copernicus Global Ocean Physics Reanalysis” en la región costera del Ecuador, con el fin de mejorar su aplicabilidad en estudios oceanográficos y pesqueros desarrollados por el IPIAP.

1.3.2 Objetivos Específicos

- Recopilar los registros de salinidad in situ obtenidos con dispositivos CTD durante cruceros y salidas de campo del IPIAP.
- Comparar estadísticamente los datos de salinidad del producto de Copernicus y los datos in situ obtenidos, cuantificando la magnitud de la incertidumbre en distintos niveles de profundidad y ubicaciones geográficas.
- Determinar el nivel de precisión, exactitud y confiabilidad del producto Copernicus en la estimación de salinidad para la zona costera del Ecuador, a partir del análisis de las diferencias observadas frente a los datos medidos en campo.
- Evaluar si existe una variación en el error de salinidad de Copernicus según la estacionalidad anual (época seca – época húmeda).

1.4 Marco Teórico

1.4.1 Área de Estudio

El área de estudio se localiza frente a la región costera suroeste del Ecuador, y está delimitada geográficamente entre las coordenadas **1.4° S a 2.2° S de latitud sur y 80.8° W a 81.2° W de longitud oeste**, según se observa en el **Mapa 1** de distribución de estaciones

oceanográficas. Debido a su ubicación presenta grandes variaciones estacionales en parámetros como el viento, corrientes, temperatura y salinidad (Alory et al., 2012), caracterizándose por ser una zona de transición entre los regímenes tropical y subtropical. En esta región, las principales masas de agua presentes son Agua Tropical Superficial (ATS) y Agua Subtropical Superficial (ASTS), las cuales se encuentran ubicadas al norte y sur del ecuador, separadas por una franja denominada Frente Ecuatorial (Sonnenholzner, 1993), y por debajo de estas dos masas se encuentra el Agua Ecuatorial Subsuperficial (AESS) (Cucalón, 1986).

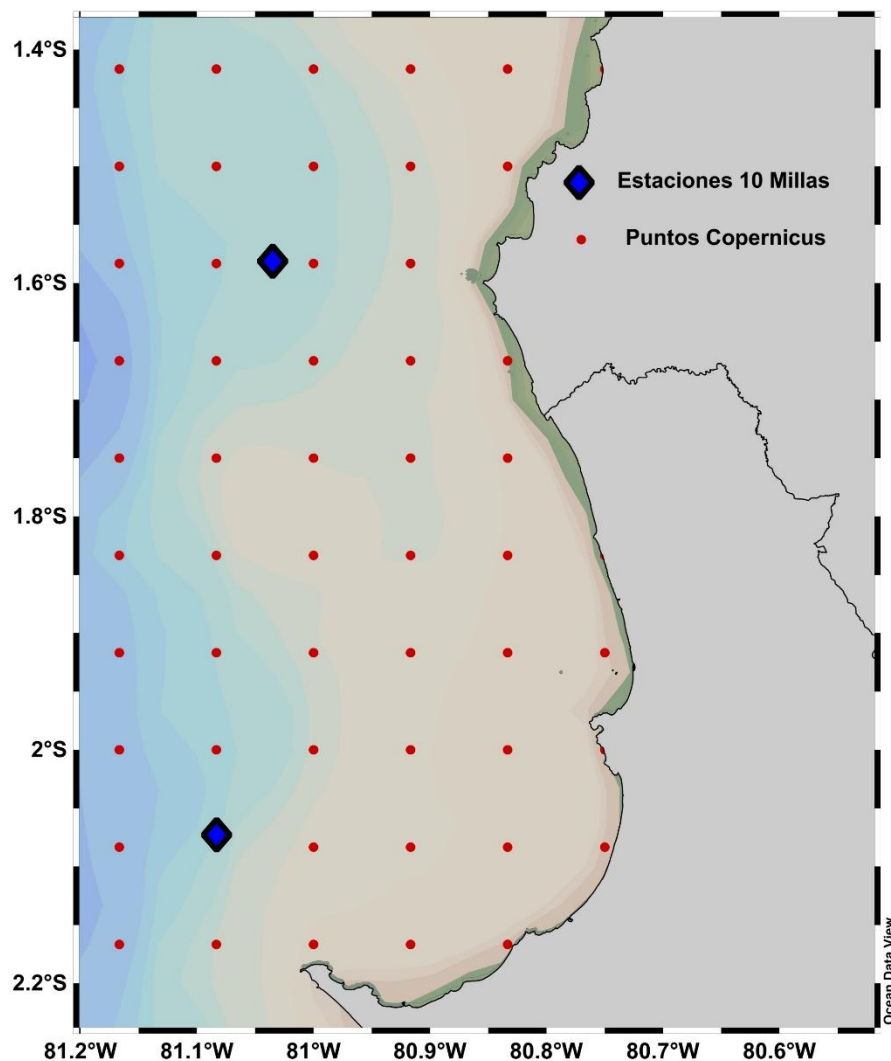
El Frente Ecuatorial es una de las características oceanográficas más relevantes de la costa del Ecuador. Durante la estación húmeda se encuentra casi ausente o muy desplazado al sur, debido al desarrollo de temperaturas relativamente altas para este período, siendo máximas en febrero-marzo. En la estación seca, el Frente se encuentra muy bien definido, siendo sostenido por un balance entre la fuerza de la corriente de Humboldt y el gradiente de presión hidrostática horizontal generado a través del frente (Cucalón, 1986).

Durante la estación húmeda (enero a abril), las temperaturas superficiales del mar alcanzan valores elevados (26–27 °C), asociados a uno de los máximos de la insolación en la zona ecuatorial (Mitchell y Wallace, 1992) y al debilitamiento estacional de los vientos Alisios, con máximos térmicos en febrero y marzo. En los meses de transición (abril-mayo), se observa una disminución progresiva de la temperatura superficial y un ligero hundimiento de la termoclina. En contraste, la estación seca (junio a diciembre) se caracteriza por temperaturas superficiales más bajas (24–25 °C), una capa de mezcla más profunda y condiciones de enfriamiento más intensas, con valores mínimos entre agosto y septiembre. En cuanto a la salinidad, durante la estación húmeda se registran valores relativamente bajos (32.5–33.8), con un marcado gradiente halino vertical, especialmente en enero y febrero. A partir de mayo-junio, la salinidad superficial aumenta (>34), coincidiendo con un debilitamiento del gradiente vertical. Posteriormente, en la estación seca la salinidad

disminuye progresivamente hasta diciembre (alcanzando 34–33), restableciéndose gradualmente el gradiente vertical (Riofrío & Chavarría, 2004).

Figura 1

Mapa del Área de Estudio en base a las estaciones Puerto López y Salinas [Autor, 2025]



1.4.2 Salinidad como variable oceanográfica y su relación con la conductividad

En oceanografía, la salinidad se define clásicamente como la masa de sales disueltas en una masa dada de agua de mar. Sin embargo, medir directamente la salinidad absoluta (g de sal por kg de agua) ha resultado poco práctica, por lo que se recurrió a métodos indirectos como la clorinidad (titulación de haluros) y, más recientemente, a la conductividad eléctrica de la muestra. La conductividad es sensible a la concentración total

de iones en solución, por lo que proporciona una medida rápida y precisa de la salinidad *in-situ*. Para aprovechar esta ventaja, en 1978 se adoptó internacionalmente la Escala Práctica de Salinidad 1978 (PSS-78, por sus siglas en inglés), que redefinió la salinidad en términos de la conductividad eléctrica de la muestra (Lewis, 1980).

El término *agua de mar normalizada con KCl* se emplea para referirse a cualquier muestra de agua marina cuya conductividad eléctrica específica es equivalente a la de una solución patrón compuesta por cloruro de potasio (KCl) disuelto en agua pura, con una fracción másica precisa de 32,4356 g/kg. Esta referencia se establece bajo condiciones estandarizadas de temperatura (14,996 °C, conforme a la escala ITS-90) y presión atmosférica normal (101 325 Pa). La solución estándar utiliza KCl con composiciones isotópicas naturales de potasio y cloro, tal como lo define la Unión Internacional de Química Pura y Aplicada (IUPAC) (Wieser, 2006).

La salinidad práctica (S), establecida por la PSS-78, se define operacionalmente a partir de un cociente de conductividades. Dicho cociente, denominado K_{15} , corresponde a la razón entre la conductividad de una muestra de agua de mar a 15 °C y 1 atm, y la de la solución patrón de KCl preparada en las condiciones mencionadas (Lewis, 1980). Esta solución de referencia se definió de modo que una muestra de agua de mar estándar con salinidad 35 produzca exactamente $K_{15} = 1,0000$. Así, en la práctica, a una conductividad relativa unitaria respecto al KCl estándar se le asigna por definición el valor $S = 35$.

Es importante notar que la salinidad práctica no tiene unidades físicas; se expresa como un número puro derivado de relaciones de conductividad. Aunque históricamente se emplearon símbolos como “‰” o la abreviatura PSU (Practical Salinity Unit), en sentido estricto S es adimensional según la PSS-78 (Millero, Feistel & Wright, 2008).

Relación con la densidad, temperatura y presión

La densidad del agua de mar depende de su salinidad, temperatura y presión de forma no lineal. La formulación clásica adoptada por UNESCO (1983) conocida como EOS-80 (Equation of State 1980), expresaba ρ (densidad *in-situ*) como función de la salinidad

práctica (escala PSS-78, variable adimensional), de la temperatura *in-situ* y de la presión, mediante un ajuste polinómico empírico (McDougall T. , et al., 2009). Por ejemplo, dicha ecuación calcula la densidad como:

$$\rho(S, T, P) \approx \rho_w(T, p) + a_1 S + a_2 S^{\frac{3}{2}} + a_3 S^2 + \dots$$

Es importante destacar que en 2010 se estableció la TEOS-10¹ como el nuevo estándar, introduciendo la salinidad absoluta S_A [g/Kg] como la variable de salinidad termodinámicamente consistente, en reemplazo de la salinidad práctica S_p . Para propósitos de este estudio, sin embargo, la salinidad utilizada es S_p que continúa siendo la salinidad referencial para oceanografía operacional, para su almacenamiento en bases de datos, y para su distribución set de datos de reanálisis y de modelación (como los de Copernicus).

Comprender cómo interaccionan estos tres factores (temperatura, salinidad, presión) permite explicar fenómenos clave como la estratificación vertical, donde el agua se organiza en capas de distinta densidad, o la circulación termohalina, impulsada por diferencias de densidad a gran escala. Estos gradientes son fundamentales para la dinámica oceánica, ya que controlan procesos como la mezcla vertical y la distribución de nutrientes y masas de agua. Incluso diferencias de densidad del orden de milésimas de kg/m³ pueden generar presiones horizontales que desencadenan movimientos de agua a lo largo de la columna oceánica (Pawlowicz, 2013).

1.4.3 Variabilidad espacial de la salinidad en el entorno ecuatorial

La configuración de corrientes y masas de agua en la zona genera un patrón espacial característico de salinidad en el Pacífico ecuatorial oriental. En términos latitudinales, existe un gradiente meridional marcado: las aguas superficiales se vuelven progresivamente más salinas hacia el sur. Las menores salinidades (< 34) se observan típicamente al norte de Ecuador (aproximadamente entre ~2°N y 8°N), en correspondencia con las ATS influenciadas por la alta pluviosidad de la ZCIT (Cucalon, 1989). Por el contrario, al sur del ecuador (entre 0–5°S) las salinidades superficiales son mayores (34–35 en promedio),

¹ TEOS-10: Ecuación termodinámica del agua de mar, por sus siglas en inglés

reflejando la presencia de AES alimentadas por la Corriente de Humboldt y el afloramiento ecuatorial (Astro Córdova, 2018).

1.4.4 Variabilidad estacional de la salinidad

El clima oceánico en la costa ecuatorial exhibe dos estaciones climáticas principales asociadas con la migración anual de la ZCIT y los cambios en la intensidad de los vientos alisios. En términos aproximados, de diciembre a mayo ocurre la estación húmeda, mientras que de junio a noviembre predomina la estación fresca y seca. Estos períodos pueden generar notables cambios en la salinidad superficial. Según las definiciones de Emilio Cuccalón en 1986. La estación seca se define de enero-abril, mientras que la húmeda de julio-octubre, los meses restantes se consideran de transición entre ambas estaciones.

La estación húmeda tiene como característica el flujo de agua cálida hacia el sur (25-27 °C) y salinidades bajas (33 – 33.8) a lo largo de la costa ecuatoriana. Esto se debe a que los vientos meridionales se debilitan paralelos a la costa, mientras que los vientos alisios del noroeste se intensifican ocasionando un transporte meridional de agua en el ecuador conocido como la Corriente del Niño y en esta época, el frente ecuatorial es impredecible, ya que se desplaza hacia el sur (frente a Perú).

La estación seca se caracteriza por presentar un marcado Frente Ecuatorial, cuya posición está determinada por el equilibrio entre la Corriente de Humboldt y el gradiente de presión horizontal. Este frente se caracteriza por marcados gradientes termohalinos superficiales entre aproximadamente 0° y 20°S, y su influencia se extiende hasta unos 30 a 40 m de profundidad. En el borde sur del Frente Ecuatorial se observan aguas frías, salinas y con alta concentración de nutrientes, originadas por la mezcla de las Aguas Subtropicales Superficiales (ASTS) transportadas por la Corriente de Humboldt con las subyacentes Aguas Ecuatoriales Subsuperficiales (AESS), las cuales afloran en torno a los 20°S. Por efecto del fortalecimiento de los vientos meridionales y de la propia Corriente de Humboldt durante la estación invernal, este frente se desplaza hacia el norte. En la zona sur de Ecuador, cerca del Golfo, hay presencia de aguas relativamente frías (21-22 °C) y salina

(34.4-35) lo que indica que no hay presencia de la descarga de los ríos, como consecuencia de la época seca (Cucalón, 1986).

CAPÍTULO 2

2. Metodología

2.1 Esquema de Análisis

En esta sección se describe la metodología que se implementó para el desarrollo del proyecto y se presenta un resumen en la ilustración 2.

Figura 2

Esquema de análisis [Autor, 2025]



Previo a la obtención y descarga de los datos, se estableció un periodo de tiempo que corresponde desde noviembre del 2022 hasta mayo del 2025. La selección del periodo se realizó considerando la calibración del equipo de medición perfilador CTD 19-Plus V la cual se hizo en junio del 2022, mes en el cual el equipo fue enviado a su casa de calibración Sea Bear en Estados Unidos y obtuvo su certificado de calibración. Con el fin, de que el conjunto de datos para la validación sea confiable.

En el paso 3 se proponen dos opciones como parte de la metodología, sin embargo, se escoge la opción 1 como método de resolución debido a que se obtuvieron mejores resultados.

Al final del documento en el apartado Apéndice, se podrá visualizar el código implementado en el software R Studio con el cual se logró realizar la validación de los datos.

Esto se describe con mayor detalle a continuación.

2.2 Obtención de datos in-situ

Como parte fundamental del análisis comparativo de la salinidad en la costa ecuatoriana, se incorporaron datos in situ obtenidos mediante campañas oceanográficas ejecutadas por el Instituto Público de Investigación de Acuicultura y Pesca (IPIAP). Esta institución proporcionó los archivos originales en formato Excel correspondientes a los perfiles de salinidad, temperatura y profundidad registrados por un sensor CTD (*Conductivity-Temperature-Depth*), utilizado en estaciones ubicadas a 10 millas náuticas de la costa.

Los datos entregados por IPIAP provenían del procesamiento inicial realizado con el software *Sea-Bird*, el cual es ampliamente reconocido en el ámbito oceanográfico por su capacidad para adquirir, calibrar y convertir las lecturas en crudo del sensor CTD en parámetros físicos corregidos y de alta calidad. En efecto, *Sea-Bird Scientific* proporciona algoritmos específicos para compensar errores derivados del arrastre térmico, respuesta del sensor de presión, alineamiento temporal entre canales, y otros factores que afectan la precisión de los datos adquiridos durante el descenso y ascenso del equipo. Esta corrección sistemática asegura la fidelidad de los registros obtenidos, permitiendo su uso en análisis científico con estándares internacionales. Además, el procesamiento y depuración de los datos siguió los lineamientos establecidos por la Comisión Permanente del Pacífico Sur (CPPS), específicamente el “Protocolo Regional de Recolección de Datos Oceanográficos y Costero-Marinos”, que establece criterios técnicos para la validación, filtrado y normalización de datos físicos marinos en los países del Pacífico suroriental (CPPS, 2015).

2.3 Descarga de datos de Copernicus

Para este estudio se obtuvieron datos diarios de salinidad superficial y subsuperficial del Servicio Marino de Copernicus (CMEMS) y se utilizó el producto *Global Ocean Physics Reanalysis* de Copernicus, cuyo ID GLOBAL_MULTIYEAR_PHY_001_030, con el set de datos cmems_mod_glo_phy_myint_0.083deg_P1D-m el cual nos proporciona datos de salinidad diaria desde los 0.49 m hasta los 109.73 m (para nuestro conjunto de datos), con

una resolución horizontal de $1/12^\circ$ lo que equivale a una rejilla regular de aproximadamente 9 km entre puntos de malla en la zona ecuatorial. Cada *grid point* representa la salinidad en un píxel oceánico de unos pocos kilómetros de lado. El periodo de tiempo del set de datos inicia el 1 de julio del 2021 hasta el 27 de mayo del 2025 y están disponibles de forma abierta (previo registro de usuario en la plataforma) en formato NetCDF. Los archivos netCDF (Formato de Datos Comunes de Red) es un tipo de formato que permite almacenar metadatos, es decir, datos científicos de múltiples dimensiones que contienen información adicional (ESRI).

La descarga de los datos se realizó inicialmente a través de la plataforma web de Copernicus Marine, utilizando su interfaz gráfica de subsetting. A través de esta se seleccionó la variable salinidad del agua de mar, delimitando el área de interés 1.4° S a 2.2° S de latitud y 80.8° W a 81.2° W de longitud, y el intervalo temporal requerido. Alternativamente, para automatizar la obtención de datos y reproducir el procedimiento, se empleó un script en Python haciendo uso de la API de Copernicus Marine. En particular, se utilizó la biblioteca oficial *copernicusmarine*, que permite conectarse al servidor de datos y descargar los archivos originales o subconjuntos especificando el ID del producto, las coordenadas geográficas y el rango de fechas deseado. Esta dualidad de métodos (portal web y script Python) aseguró tanto la verificación manual de los datos descargados como la reproducibilidad del proceso de extracción de datos.

2.4 Esquema de interpolación vertical

Los modelos oceánicos de reanálisis como los productos Copernicus proveen datos de salinidad en niveles verticales irregulares, lo que dificulta obtener valores en profundidades estándar (0, 10, 30, 50, 75 y 100 m). Para solucionar esto, es necesario interpolar los perfiles verticales de salinidad a esas profundidades. Sin embargo, la elección del método de interpolación influye en la precisión de los resultados. Estudios recientes han demostrado que usar un solo método (por ejemplo, la interpolación lineal) puede introducir sesgos sistemáticos cuando el perfil de salinidad o temperatura tiene curvaturas

pronunciadas asociadas a termoclinas o haloclinas (*Li et al., 2022*). Por tanto, se ha propuesto una estrategia metodológica combinada: aplicar y comparar múltiples métodos de interpolación unidimensional (lineal, polinómico de segundo y tercer grado, spline cúbico y LOESS) para minimizar errores y mejorar la representatividad de la salinidad interpolada. Este proceso permite estimar, mediante técnicas de interpolación unidimensional, el valor de salinidad del modelo en las mismas fechas y profundidades que los registros in situ, y evaluar cuál de las técnicas implementadas ofrece menor error de predicción.

1. Preparación y emparejamiento de los datos:

Para cada registro in situ, se identifica la fecha exacta y se selecciona el campo de salinidad del modelo de Copernicus correspondiente a esa misma fecha. A continuación, se extrae el perfil vertical de salinidad del modelo (en esa fecha y coordenadas) y se aplica interpolación en la dimensión vertical para estimar la salinidad del modelo exactamente a la profundidad de la medición in situ.

2. División de datos para entrenamiento y validación.
3. Aplicación de cinco métodos de interpolación unidimensional.
4. Evaluación del error y selección del mejor método.

A continuación, se describen los fundamentos de cada método y se analiza comparativamente su comportamiento en términos de robustez y sustentando con evidencia de la literatura científica la utilidad de su uso combinado en contextos oceanográficos.

2.4.1 Interpolación lineal

La interpolación lineal conecta los puntos adyacentes del perfil con segmentos rectos. Es el método más simple y asume cambios lineales entre mediciones conocidas. Su ventaja principal es la simplicidad y estabilidad: no genera valores fuera del rango de los datos originales evitando extrapolaciones y tiende a ser robusta frente a valores atípicos aislados. Este método se ha utilizado ampliamente para estandarizar perfiles oceanográficos a niveles fijos (*Tang et al., 2022*).

2.4.2 Interpolación polinómica de segundo y tercer grado

La interpolación polinómica utiliza un polinomio de grado fijo (cuadrático para grado 2, cúbico para grado 3) que aproxima la forma del perfil. Un polinomio de segundo grado puede ajustarse para pasar por tres puntos de datos, mientras que uno de tercer grado puede pasar por cuatro puntos o ajustarse por mínimos cuadrados si hay más puntos. Estos polinomios de bajo grado permiten cierta curvatura en la estimación, representando mejor perfiles suavemente curvados que la interpolación lineal. En efecto, capturan tendencias convexas o cóncavas del gradiente de salinidad que la lineal omitiría. Su fortaleza radica en la simplicidad del modelo matemático y la ausencia de oscilaciones erráticas asociadas a polinomios de grado muy alto (*Chapra y Canale, 2006*).

2.4.3 Interpolación spline cúbico

El spline cúbico es un método de interpolación segmentario que ajusta polinomios cúbicos por tramos entre cada par de puntos adyacentes, garantizando continuidad de primer y segundo derivadas en los nodos. A diferencia de un único polinomio global, un spline cúbico pasa exactamente por todos los datos medidos y produce una curva suave en todo el perfil (*Durán, 1994*). Un spline cúbico bien implementado minimiza la curvatura total de la función interpolada, lo que suele brindar estimaciones suaves y físicamente razonables. Su ventaja científica es que captura las variaciones graduales del perfil y puede ajustarse a cambios de pendiente (*North y Livingstone, 2013*).

2.4.4 Suavizado local LOESS (Locally Weighted Scatterplot Smoothing)

El método LOESS o regresión local ponderada, no es una interpolación estrictamente determinada, ya que no garantiza pasar exactamente por cada punto sino un método de suavizado que estima una curva ajustando polinomios de bajo grado a lo largo del perfil. En cada vecindad local, LOESS realiza una regresión generalmente lineal o cuadrática ponderando más las observaciones cercanas a la profundidad de interés (*Ridgway et al., 2002*).

2.5 Esquema de interpolación horizontal bidimensional

Con el objetivo de comparar datos satelitales del producto *Copernicus Global Ocean Physics Reanalysis* con observaciones in situ, se desarrolló un procedimiento de ajuste horizontal de salinidad. Este permite estimar, mediante técnicas de interpolación espacial bidimensional, el valor de salinidad del modelo Copernicus exactamente en las coordenadas geográficas (latitud y longitud) donde se realizó una medición in situ, para una fecha y profundidad específicas. Esta rutina se diseñó con la misma lógica que la interpolación vertical: aplicar distintos métodos de interpolación en una dimensión (en este caso, sobre el plano horizontal) y evaluar cuál proporciona menor error de estimación en cada caso.

1. Preparación y emparejamiento de datos

Para cada registro in situ, se selecciona del producto Copernicus el campo de salinidad correspondiente a la misma fecha. En caso de que la profundidad del modelo no coincida exactamente con la de la medición, se aplican primero un esquema de interpolación vertical, como se describe en el procedimiento anterior, para obtener el campo horizontal de salinidad a la profundidad deseada. Como resultado, se obtiene una malla bidimensional de salinidad en latitud y longitud, correspondiente a la fecha y profundidad del registro in situ.

2. División de datos para entrenamiento y validación.

Para cada punto del conjunto de entrenamiento, se construye el subconjunto de datos del modelo de Copernicus necesarios para realizar la interpolación en las coordenadas deseadas.

3. Aplicación de cinco métodos de interpolación bidimensional.

4. Evaluación del error y selección del mejor método.

En el análisis espacial, la interpolación bidimensional permite estimar el valor de una variable en ubicaciones no muestreadas a partir de observaciones en puntos cercanos. Todos los métodos parten del principio de autocorrelación espacial. A continuación, se

describen cinco métodos comunes de interpolación bidimensional desde los más simples hasta métodos geoestadísticos avanzados junto con sus fundamentos.

2.5.1 Vecino más cercano

La interpolación por el vecino más cercano asigna a cada ubicación desconocida el valor del punto de muestreo más próximo. En la práctica, el espacio se divide en polígonos de Thiessen o Voronoi, donde cada polígono recibe el valor de la muestra que lo origina. Este método es extremadamente simple y conserva los valores originales sin promediarlos. Dado que solo considera un punto (el más cercano) e ignora los demás vecinos, el modelo puede presentar saltos bruscos y no refleja variaciones graduales. (*Navarro Alberto et al., 2011*).

2.5.2 Interpolación por promedio de k vecinos más cercanos

El método de promedio de k vecinos más cercanos extiende la idea anterior promediando los valores de las k muestras más próximas a cada ubicación sin datos. En lugar de tomar un único punto, se calcula la media aritmética de los k valores vecinos para estimar el valor desconocido. Si $k = 1$, este método equivale exactamente al vecino más cercano, mientras que con valores mayores de k se obtiene una superficie más suavizada. La premisa es que al considerar varios puntos cercanos se reduce el efecto de valores atípicos o ruido en una sola muestra, logrando una estimación más estable localmente. Este enfoque de promedio móvil local asume que todos los k vecinos elegidos aportan influencia similar, no hay un peso diferencial por distancia en la media simple (*Morales España et al., 2008*).

2.5.3 Regresión espacial

Es un método que ajusta una función matemática (típicamente un polinomio) a todos los datos en el espacio. Es decir, se realiza una regresión de mínimos cuadrados donde las coordenadas espaciales (X , Y) actúan como variables independientes, obteniendo así una superficie polinomial que aproxima la distribución de la variable dependiente. Este método

produce una superficie suave que minimiza la varianza respecto a los puntos medidos, aunque raramente pasa exactamente por ellos, es decir, captura las tendencias globales de la variable, pero no necesariamente las fluctuaciones locales (*ESRI, Análisis de Regresión*).

2.5.4 Interpolación por Inverse Distance Weighting (IDW)

Es un método local que asigna a cada punto desconocido un valor calculado como promedio ponderado de las muestras vecinas, donde los pesos dependen de la distancia. En IDW se asume que cuanto más cerca está un punto de la ubicación de interés, más influencia (peso) debe tener en la estimación, y este peso decae inversamente con la distancia (*Bautista Zúñiga, 2007*).

2.5.5 Kriging ordinario

Es un método de interpolación geoestadístico considerado más sofisticado, que incorpora un modelo explícito de correlación espacial para estimar valores desconocidos. A diferencia de los métodos determinísticos anteriores, el kriging se basa en la teoría de las variables regionalizadas y utiliza el semivariograma (o función de variancia espacial) para ponderar de forma óptima los datos vecinos. En términos conceptuales, el kriging ordinario produce un estimador lineal insesgado de mínima varianza, esta interpolación utiliza el grado de autocorrelación entre los sitios donde se hizo el muestreo para calcular las estimaciones en los puntos donde no hay información. Una de sus grandes ventajas es que este método calcula una medida de error sobre la superficie estimada (*Bautista Zúñiga, 2007*).

A continuación, se muestra una tabla con los métodos utilizados para estimar la salinidad en las coordenadas específicas de la rutina implementada en el proyecto; en donde se puede visualizar que el mejor método de interpolación geoespacial es Kriging Ordinario.

Tabla 1

Métodos de Interpolación Geoespacial más utilizados con su porcentaje de error

Métodos de Interpolación	% Observaciones	Promedio de Error
Kriging Ordinario	93.3	0.0323
Vecino más cercano	5.37	0.0758
Promedio de k	1.34	0.0758
Vecino más Cercano		

2.6 Validación de los resultados: Análisis exploratorio y pruebas no paramétricas

Una vez generadas las superficies interpoladas, es fundamental realizar una validación de los resultados para cuantificar qué tan precisos y exactos son los valores estimados con respecto a datos reales conocidos. Esta validación consta típicamente de un análisis exploratorio de los errores (diferencias entre valores predichos y observados) y de pruebas estadísticas que corroboren la calidad del modelo. A continuación, se describen los elementos clave de este proceso.

2.6.1 Análisis exploratorio de precisión y exactitud

Dos criterios fundamentales que deben examinarse en los errores de interpolación son la exactitud y la precisión de las estimaciones. En términos estadísticos, la exactitud se refiere al sesgo o error sistemático, es decir, un método exacto que produce errores promedio cercanos a cero (no tiende a sobreestimar ni subestimar en promedio). Por otro lado, la precisión se refiere a la dispersión de los errores o variabilidad de las predicciones: un método preciso tiene errores generalmente pequeños en magnitud, es decir,

predicciones muy próximas a los valores reales en la mayoría de los casos (*Ruiz Armenteros et al., 2010*).

Para cuantificar estos aspectos, suelen emplearse estadísticas resumen de los errores de validación:

2.6.2 Error medio (Bias)

Mide el sesgo promedio del modelo. Se calcula como el promedio simple de todas las diferencias entre el valor estimado y observado. Un error medio idealmente igual a 0 indica ausencia de sesgo global; si es positivo, revela una tendencia a sobreestimar, y si es negativo, a subestimar (*Basharat y Naiem, 2017*).

2.6.3 Error cuadrático medio (RMSD)

Evalúa la magnitud de los errores, combinando en una sola métrica la variabilidad de estos. Se define como la desviación estándar del promedio de las diferencias entre el valor estimado y observado. Se obtiene como la raíz cuadrada del promedio de los errores al cuadrado, un RMSD más pequeño indica predicciones más cercanas a los valores observados en términos de magnitud de error (*Olumide, 2023*).

2.6.4 Error absoluto medio (MAE)

Representa el promedio de las diferencias absolutas entre los valores observados y los estimados. Es una métrica directa, en las mismas unidades de la variable de interés (para este proyecto en PSU), que permite interpretar fácilmente la magnitud del error (*Gilbert, 2023*).

2.6.5 Error porcentual absoluto medio (MAPE)

Expresa el error en términos relativos, permitiendo comparar errores en distintos rangos de salinidad. Es particularmente útil cuando se desea interpretar el error como un porcentaje del valor real. Aunque el MAPE puede ser sensible a valores cercanos a cero o

extremos, en contextos oceanográficos donde la salinidad se mantiene dentro de rangos estables, esta métrica ofrece interpretaciones robustas y comparables (*Gilbert, 2023*).

CAPÍTULO 3

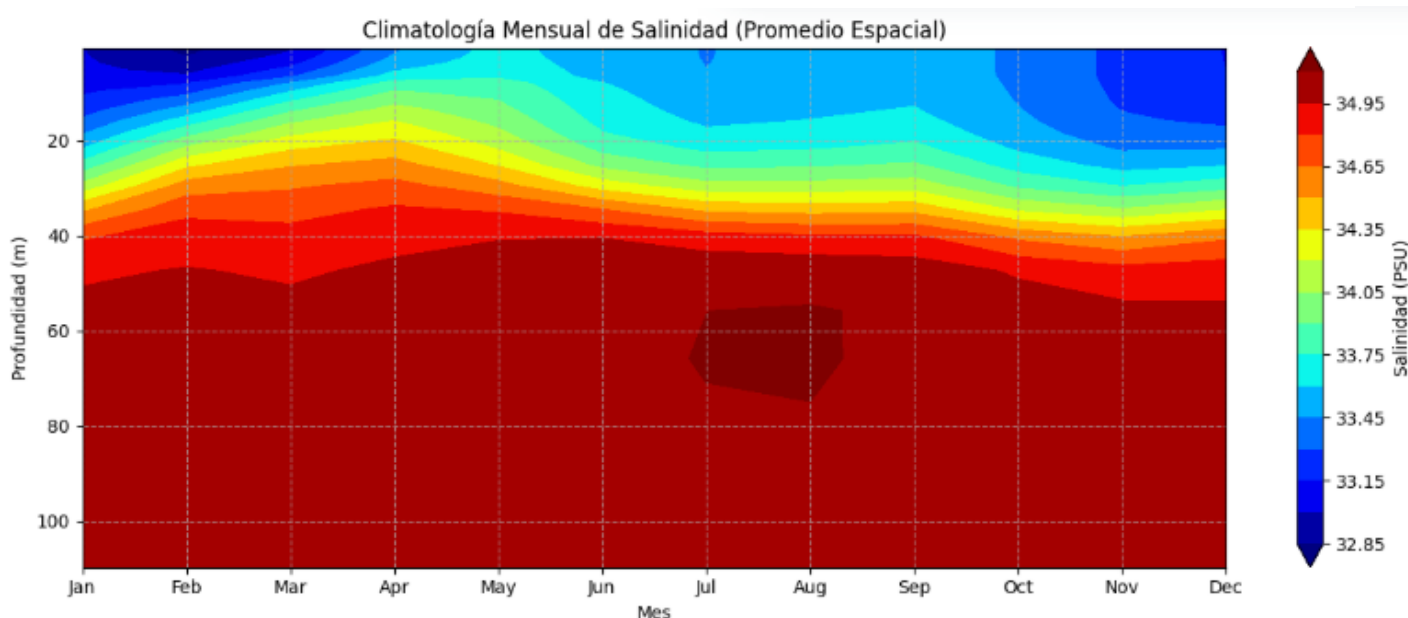
3. Resultados y Análisis

3.1 Ciclo anual de salinidad en Puerto López y Salinas (2022–2025)

El análisis comparativo de la salinidad in situ versus la salinidad modelada por Copernicus revela un ciclo anual definido en ambas estaciones costeras, Puerto López y Salinas. Entre noviembre de 2022 y mayo de 2025 se observa que la salinidad medida exhibe variaciones estacionales marcadas: valores más bajos a inicios de cada año (época húmeda) y valores más altos hacia mitad, y finales de año (época seca). En Puerto López, por ejemplo, las mediciones superficiales alcanzan mínimos típicamente entre febrero y marzo en torno a 33.0 – 33.5, coincidiendo con la época lluviosa, mientras que en Salinas los mínimos ocurren de forma similar con valores aproximados de 32.4 – 33.9. Por otro lado, los máximos de salinidad en Salinas se registran aproximadamente entre abril y julio llegando a valores entre 33.5 – 34.7 ocurriendo durante la época de transición a seca. Estas tendencias concuerdan con la climatología regional y por lo descrito por Emilio Cucalón en 1986 en donde se evidencia que, durante la época húmeda, se presentan condiciones de aguas cálidas y salinidades más bajas debido al flujo costero que baja hacia el Sur proveniente de la Bahía de Panamá, mientras que para la época seca la presencia del Frente Ecuatorial es una de las características principales debido al equilibrio entre la presencia de la Corriente de Humboldt y el gradiente de presión hidrostática horizontal, el Frente Ecuatorial se caracteriza por presentar gradientes termo-halinos elevados, lo cual influye al aumento en la salinidad de la zona durante la época (*Cucalón, 1986*). Debido a estas condiciones el Pacífico oriental ecuatorial presenta aguas superficiales menos salinas al inicio del año debido a fuertes lluvias y afluencia de aguas tropicales, mientras que a mitad del año son aguas más salinas hacia la segunda mitad por la influencia de la Corriente de Humboldt y menores precipitaciones. A continuación, se muestra la climatología mensual de la salinidad en el área de estudio, la cual corresponde desde el año 1997 hasta el 2025, elaborada con los datos del modelo de reanálisis de Copernicus.

Figura 3

Climatología Mensual de Salinidad 1997-2025. [Autor, 2025]



El modelo Copernicus, reproduce en gran medida este ciclo anual. Las series temporales modeladas en ambas estaciones muestran mínimos de salinidad en los mismos periodos lluviosos y máximos durante los meses secos, evidenciando que el reanálisis es sensible a los forzamientos estacionales. Sin embargo, se aprecia una ligera diferencia de magnitud: en los picos húmedos (mínimos de salinidad) el modelo tiende a mantenerse algo más salino que las observaciones, mientras que en los picos secos (máximos de salinidad) el modelo acompaña de cerca los valores medidos, llegando en algunos casos a valores equivalentes o apenas inferiores. Por ejemplo, durante marzo, uno de los periodos más húmedos en donde la salinidad superficial *in situ* en Puerto López cayó cerca de 33.68, mientras que el modelo reportó 33.06 en esa fecha; en contraste, en el mes de septiembre en Salinas, se evidenciaron aguas más salinas 34.21 (*in situ*), mientras que el modelo estimó 34.08, prácticamente capturando el máximo observado. Esto sugiere que el modelo reproduce correctamente la fase y la forma del ciclo anual, pero con una amplitud estacional

ligeramente atenuada (es decir, no alcanza valores tan extremos en los mínimos), es decir, el modelo representa mejor el valor de la salinidad durante la época seca.

Entre las dos estaciones, no se evidencian desfases importantes en la ocurrencia de máximos y mínimos anuales ambas siguen el mismo régimen climático regional. No obstante, Salinas presenta salinidades ligeramente más elevadas que Puerto López durante la época seca, coherente con su ubicación más al sur, más próxima a la influencia subtropical salina de la corriente de Humboldt. En la época lluviosa las diferencias entre estaciones prácticamente se reducen, pues ambas quedan bajo la influencia de las aguas tropicales bajas en salinidad que se propagan hacia el sur abarcando tanto Puerto López como Salinas. En síntesis, el modelo Copernicus logra representar el patrón anual de ambas estaciones de forma consistente, capturando las variaciones interanuales generales, por ejemplo, la ligera disminución de salinidad observada en la época húmeda, posiblemente asociada a lluvias intensas o eventos oceánicos de gran escala. Cualquier discrepancia menor en magnitud (del orden de décimas de unidad) sugiere un sesgo sistemático leve que se analiza con mayor detalle en secciones posteriores.

3.2 Contraste: Época húmeda vs. época seca

Al diferenciar explícitamente entre la estación húmeda (diciembre–mayo) y la estación seca (junio–noviembre), se destacan variaciones significativas en la salinidad y en la capacidad del modelo para seguir estos cambios. Durante la época húmeda, caracterizada por altas precipitaciones y temperaturas cálidas (verano austral). Los datos in situ confirman que la salinidad superficial presenta variaciones y muestra mínimos valores durante esta época, aunque estos no son constantes en todos los meses, reflejando la influencia de aguas tropicales diluidas por la lluvia que se desplazan hacia el sur (la llamada Corriente del Niño). El modelo Copernicus reproduce esta disminución estacional: sus salinidades simuladas también caen durante los meses lluviosos, indicando que es sensible a los cambios estacionales impulsados por la atmósfera (precipitación, vientos) y la dinámica de corrientes estacionales. Sin embargo, es evidente que el modelo no alcanza

plenamente los valores mínimos medidos, permaneciendo aproximadamente 0.2–1.0 por debajo de las observaciones en la época húmeda a nivel superficial (< 50 m). Esto apunta a que el reanálisis subestima el efecto de dilución por lluvias y esorrentía costera durante esta época, ya que también se evidencia que es donde mayor error presenta el modelo.

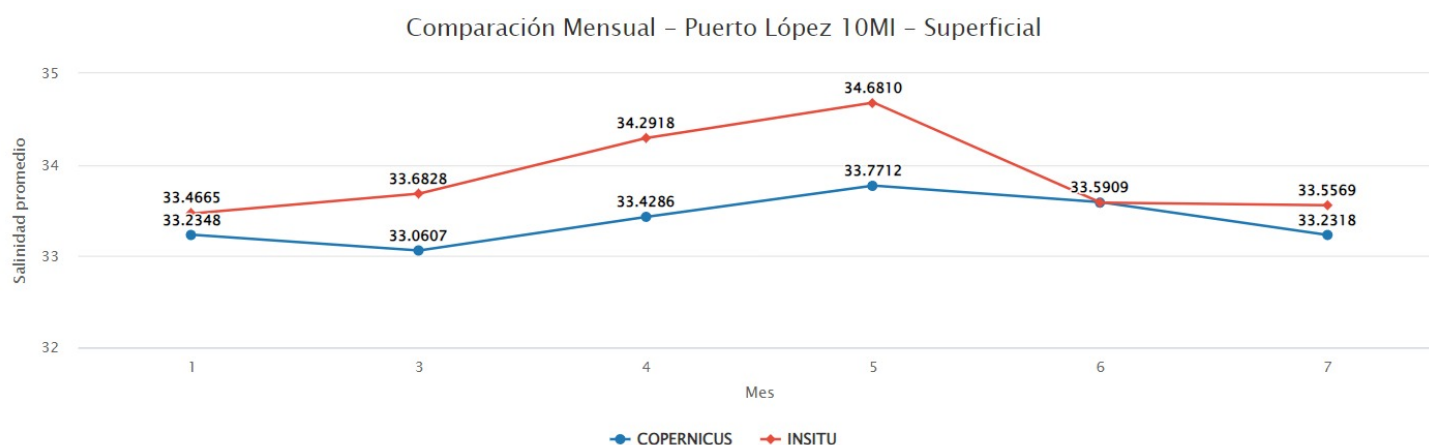
Figura 5

Promedios mensuales de salinidad superficial en Salinas (2023-2025) [Autor, 2025]



Figura 4

Promedios mensuales de salinidad superficial en Puerto López (2023-2025) [Autor, 2025]



Sin embargo, esto no ocurre a nivel subsuperficial, en donde es notable que el modelo de reanálisis se estabiliza y estima mejor el valor de la salinidad, después de los 50 metros (Quispe-Ccalluari y Quispe Ramos, 2025), en donde las diferencias son entre 0.01 –

0.2 y el modelo presenta errores muy bajos. Esto podría darse como debido a que a los 50 m es la profundidad máxima donde existe variación de la densidad (Recalde Mosquera, 2014). En los siguientes gráficos se puede observar las diferencias mínimas que existen entre los datos del modelo de Copernicus e *in-situ*.

Figura 6

Promedios mensuales de Salinidad Subsuperficial en Salinas (2023-2025) [Autor, 2025]

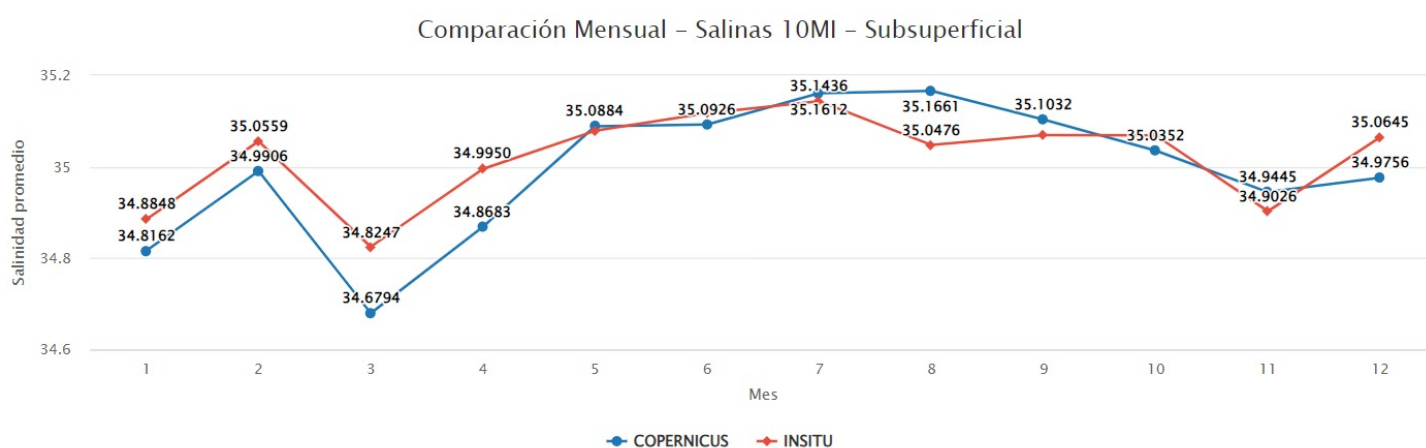
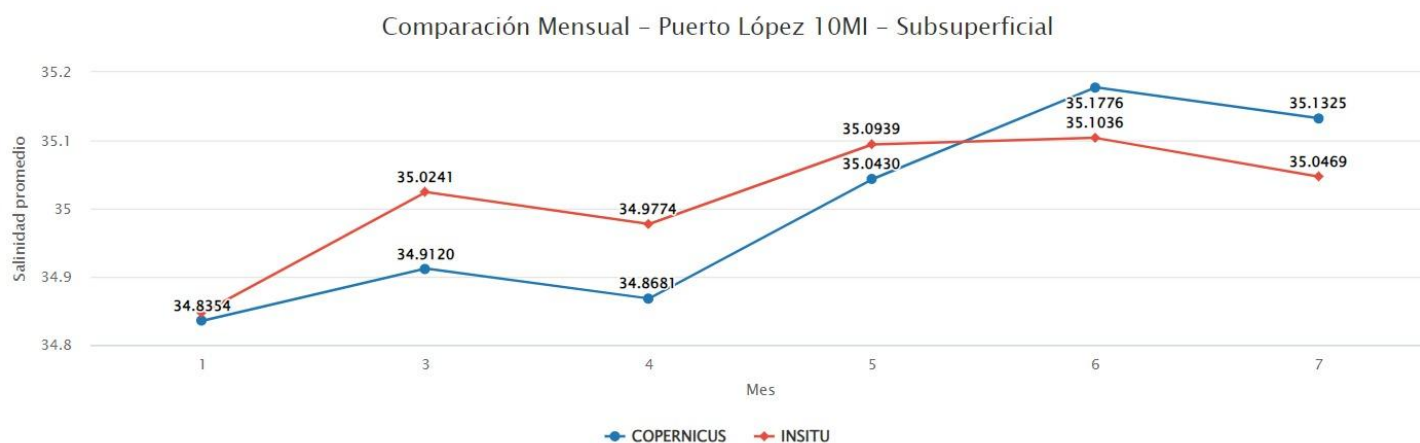


Figura 7

Promedios mensuales de Salinidad Subsuperficial en Puerto López (2023-2025) [Autor, 2025]



En contraste, durante la época seca (invierno austral), el escenario se invierte: disminuyen las lluvias y predominan vientos meridionales más intensos, favoreciendo la presencia de aguas ligeramente frías y salinas. Las mediciones muestran que la salinidad subsuperficial aumenta significativamente en ambos sitios conforme aumenta la

profundidad, con valores típicamente en el rango 34.0 – 34.6. Este enriquecimiento salino refleja la presencia de masas de agua Ecuatoriales Subsuperficiales que emergen cerca de la costa ecuatoriana bajo la intensificación de la Corriente de Humboldt (*Recalde Mosquera, 2014*) . En Salinas, se observaron salinidades ligeramente más elevadas, alrededor de 34.8–35.0 en algunos momentos de agosto–septiembre. Puerto López, algo más al norte, igualmente alcanzó ~35.0 a nivel subsuperficial durante este periodo seco. Esto sugiere que el modelo incorpora los procesos dominantes de la estación seca, como la mayor evaporación relativa y el ingreso de aguas salinas desde el sur, sin embargo, se pueden observar algunas discrepancias en aquello, ya que se visualiza un incremento de la salinidad superficial en el mes de mayo, el cual es considerado el final de la época húmeda, cercana a la transición a seca.

Es importante resaltar que el modelo muestra sensibilidad a la estacionalidad, las transiciones entre épocas húmeda y seca son más graduales en la simulación que en los datos reales. Las observaciones pueden mostrar cambios relativamente bruscos de salinidad de un mes a otro al inicio o fin de la temporada de lluvias. El reanálisis, en cambio, suele distribuir estos cambios en el tiempo, reflejando su naturaleza filtrada, es importante recordar que integra datos y forzamientos en ciclos de asimilación periódicos. A pesar de ello, no se aprecian desfases significativos en el momento de cambio estacional.

3.3 Variación con la profundidad: Capa subsuperficial (50 – 100 m)

Una de las preguntas claves es cómo difieren las comparaciones entre modelo y datos en la superficie versus a mayor profundidad. Para ello se analizaron dos estratos: la capa superficial de 0–30 m (aproximadamente equivalente a la capa de mezcla estacional) y la capa subsuperficial de 50–100 m. Este desglose permite evaluar si el desempeño del modelo Copernicus varía con la profundidad, dado que los procesos físico-oceanográficos no son los mismos en la superficie que a 50 m de profundidad.

Los resultados muestran diferencias notables, en la capa superficial (0–30 m), la salinidad tanto observada como modelada exhibe mayores fluctuaciones temporales,

estrechamente ligadas a las condiciones atmosféricas y estacionales ya descritas. La variabilidad intermensual en superficie es alta, con rangos de hasta ~ 1.0 a lo largo del año. En cambio, la capa de 30–100 m presenta una variabilidad más amortiguada, es decir, las salinidades en este estrato cambian más lentamente y en un rango menor (típicamente ~ 0.03 – 0.2 a lo largo del año). Esto concuerda con la naturaleza más estable del océano subsuperficial, donde la influencia directa de la lluvia o evaporación es reducida y las masas de agua tienden a mantener características más constantes y homogéneas. Por ejemplo, a 50 m de profundidad en ambas estaciones (Salinas y Puerto López), la salinidad se mantuvo alrededor de 34.8–35.1 durante gran parte del año, con pequeñas caídas durante el pico lluvioso. En superficie, en cambio, la salinidad osciló entre ~ 33.2 (época húmeda) y ~ 34.2 (época seca) en el mismo periodo. El modelo refleja esta estratificación vertical, en la capa superficial, las discrepancias modelo-observación son más evidentes en la época húmeda. Esto sugiere que puede estar mezclando o difuminando verticalmente la salinidad más de la cuenta, reduciendo el contraste vertical real. De hecho, globalmente se sabe que muchos modelos tienen dificultades para reproducir estratificaciones halinas muy someras asociadas a precipitaciones intensas o descargas continentales, a menos que incorporen de forma precisa esos flujos de agua dulce. En la capa subsuperficial (30–100 m), sin embargo, el modelo muestra un mejor ajuste, las salinidades modeladas en este rango de profundidad siguen muy de cerca las observadas, con diferencias promedio muy bajas. Esto indica que el reanálisis representa correctamente las propiedades de las masas de agua principales bajo la superficie, como la masa de Agua Ecuatorial Subsuperficial y Agua Subtropical Superficial. La concordancia en profundidad sugiere que el modelo captura bien la estructura de salinidad, probablemente apoyado por la asimilación de perfiles de salinidad de Argo u otras fuentes globales en esas capas (*Lellouche et al., 2021*)

Otro aspecto relevante es cómo cambian los errores con la profundidad (esto se cuantifica formalmente en la siguiente sección de métricas). De manera cualitativa, se encontró que los errores del modelo (diferencias absolutas) tienden a ser mayores en la

superficie que a 50 m. Esto es consistente con la expectativa de que las discrepancias se acentúan en la interfaz aire-mar, donde pequeños errores en la parametrización de flujos de calor, evaporación o mezcla vertical en el modelo pueden traducirse en sesgos de salinidad notables en la capa de mezcla. Además, cualquier error de fase en la respuesta estacional, por ejemplo, retraso en la simulación de la llegada de aguas frescas impactará primero en la superficie. Desde la perspectiva oceanográfica, esta evaluación sugiere que el modelo Copernicus desempeña mejor su papel en reproduciendo la estructura subsuperficial que la superficial en la costa ecuatoriana. La capa superficial está sujeta a cambios de pequeña escala (lluvia localizada, ríos costeros menores, frentes muy agudos) que un modelo global de aproximadamente 8 km difícilmente resuelve con precisión. Por otro lado, a 50 o 100 m, predominan las dinámicas de masa de agua de mayor escala que el reanálisis sí capta adecuadamente. Este hallazgo es relevante, ya que implica que las aplicaciones que dependan de la salinidad superficial, por ejemplo, estudios de intercambio aire-mar, o ecosistemas costeros sensibles a bajos de salinidad deberán tomar con cautela los datos del modelo, mientras que para caracterizar condiciones subsuperficiales el producto ofrece una confiabilidad mayor.

3.4 Indicadores de error y desempeño del modelo Copernicus por temporada

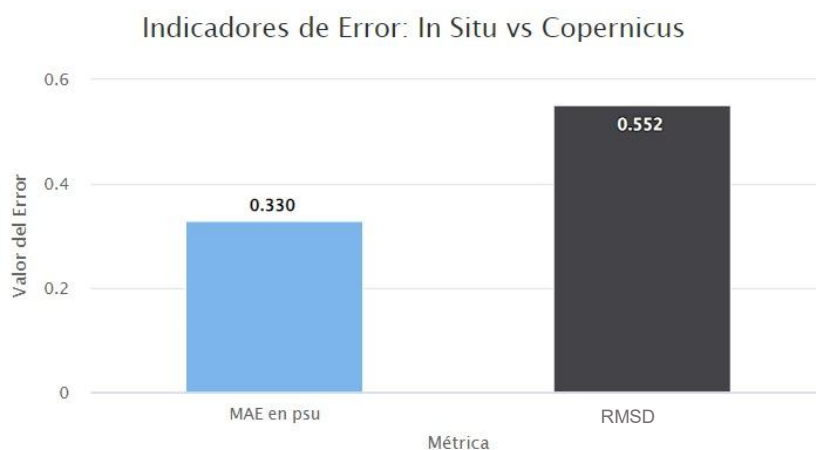
Para cuantificar formalmente el desempeño del modelo frente a las observaciones, se calcularon varios indicadores estadísticos de error: el Error Absoluto Medio (MAE), la Raíz de la Desviación Cuadrática Media (RMSD) y el Error Absoluto Porcentual Medio (MAPE), desglosados por mes, estación y capa de profundidad en cada sitio. Estos indicadores permiten evaluar no solo la magnitud de los errores en unidades prácticas de salinidad, sino también su variación relativa y estacional.

Los resultados de MAE y RMSD confirman las observaciones cualitativas descritas anteriormente. En términos generales, los errores del modelo son bajos a moderados, sin embargo, son superiores a los que establece el Manual de Uso de Copernicus, el cual menciona que el RMSD aceptable es de 0.2, mientras que el análisis realizado para la zona

ecuatoriana indica que el RMSD es de 0.55 y un MAE de 0.33, aun así, esto indica un buen desempeño promedio, que se evidencia mejor para ciertas épocas y profundidades, como se muestra a continuación.

Figura 8

Indicadores de Error: In situ vs Copernicus [Autor, 2025]



Sin embargo, estos errores no son uniformes a lo largo del año, presentan picos durante ciertas épocas y disminuyen en otras. Específicamente, los mayores errores se concentran en la época húmeda (aproximadamente diciembre–mayo). Durante estos meses, el MAE mensual llegó a superar el de la época seca. Esta estacionalidad del error refuerza la conclusión previa.

Figura 9

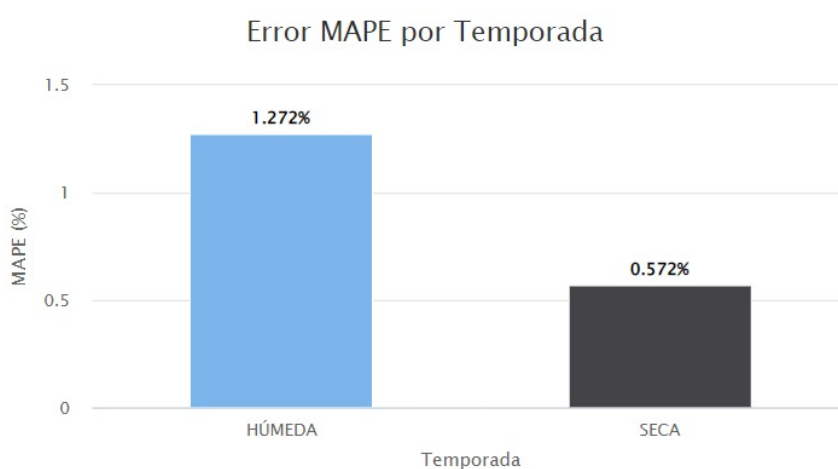
Errores por Época (Húmeda y Seca) [Autor, 2025]



El modelo tiene mayor dificultad durante la estación lluviosa para reproducir fielmente la salinidad. La dispersión de los datos en esa época (reflejada en RMSD) es mayor, evidenciando eventos puntuales donde la diferencia modelo-observación fue notable. En cambio, en la estación seca los errores se reducen y estabilizan un indicio de que el modelo se comporta de forma más consistente cuando las condiciones oceánicas están dominadas por aguas más homogéneas.

Figura 10

Error MAPE por Temporada (Húmeda y Seca) [Autor, 2025]



Cabe mencionar que estos niveles de error local concuerdan con evaluaciones de reanálisis globales en otras regiones costeras. Estudios globales señalan que la incertidumbre de salinidad de productos como GLORYS12V1 en franjas costeras puede llegar a del orden de algunas décimas PSU, tal como hallamos en Ecuador. Asimismo, la presencia de sesgos estacionales (más error en épocas lluviosas) es un patrón observado en zonas con marcado ciclo hidrológico, dado que los modelos tienden a promediar la entrada de agua dulce espacialmente. Por ejemplo, (Dossa *et al.*, 2021) reportaron que los gradientes costeros de salinidad en regiones tropicales con ríos o fuertes lluvias suelen estar suavizados en los productos de reanálisis, originando discrepancias máximas precisamente cuando las aguas costeras reales están más frescas que las simuladas.

3.5 Variabilidad del error de salinidad en función de la profundidad

La distribución vertical de la salinidad en la costa ecuatorial del Pacífico presenta una marcada estratificación halina, con variaciones importantes entre la superficie y las capas subsuperficiales. Esta estratificación, junto con procesos locales como precipitación intensa y el aporte continental, puede afectar la habilidad del modelo para reproducir correctamente la salinidad a diferentes profundidades. En aguas ecuatoriales, ya se ha evidenciado que la salinidad suele ser menor cerca de la superficie y aumenta con la profundidad, formando una haloclina (gradiente vertical de salinidad). Por ejemplo, en la costa norte de Ecuador en Esmeraldas se han registrado salinidades superficiales por debajo de 32, mientras que en capas más profundas (40–100 m) las aguas son mucho más salinas, superando las 35 unidades (*Torres et al., 2019*). La costa ecuatorial está bajo la influencia directa de la Zona de Convergencia Intertropical, particularmente hacia el norte de Ecuador, donde las precipitaciones tropicales son intensas. Las lluvias también aportan grandes volúmenes de agua dulce a la superficie del mar, creando capas superficiales de salinidad muy reducida. Estos aportes de agua dulce pueden estratificar fuertemente la capa superior del océano (*Shcherniba et al., 2019*) generando un problema para modelar correctamente el valor de la salinidad en estas zonas.

La salinidad hasta los 30 m de profundidad en la costa ecuatorial ecuatoriana, presentan un valor altamente variable e influenciado por procesos como precipitación tropical, escorrentía continental y mezcla vertical. Estas capas superficiales de baja salinidad son las más difíciles de modelar con precisión y, por tanto, el modelo Copernicus muestra mayores discrepancias relativas frente a datos *in-situ*. En cambio, a mayor profundidad por debajo de 50 m, las aguas tienden a ser más homogéneas y estables en salinidad, lo que reduce el error absoluto y también su porcentaje relativo. De hecho, capas profundas con salinidad casi constante ~34.2 presentan menores errores. Esto concuerda con que el desempeño del modelo mejora en niveles donde la variabilidad oceanográfica es menor y más predecible.

La gráfica comparativa de MAE (error medio absoluto) y RMSD (raíz del error cuadrático medio) del modelo de salinidad de Copernicus frente a datos in situ revela un patrón claro de variación del error a lo largo de la columna de agua. En las capas superficiales (0–10 m) se observa generalmente un error mayor en comparación con niveles más profundos. Esto puede atribuirse a la alta variabilidad de la salinidad superficial en la costa ecuatoriana como se mencionó anteriormente.

Figura 11

Error MAPE del modelo por profundidad [Autor, 2025]

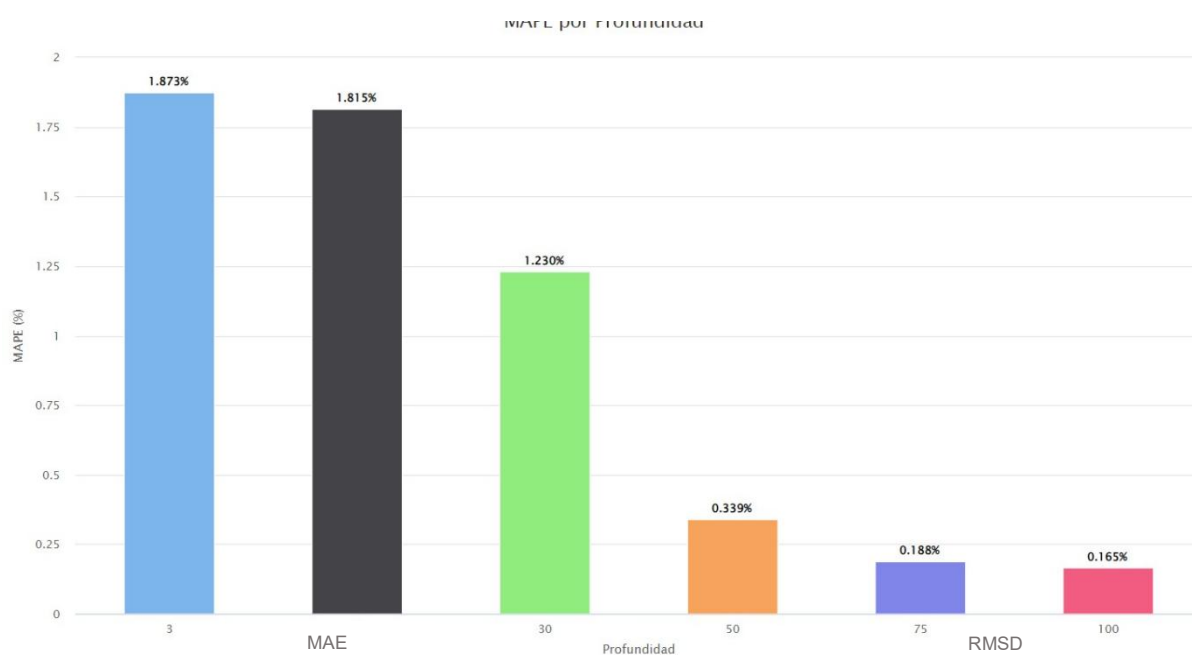
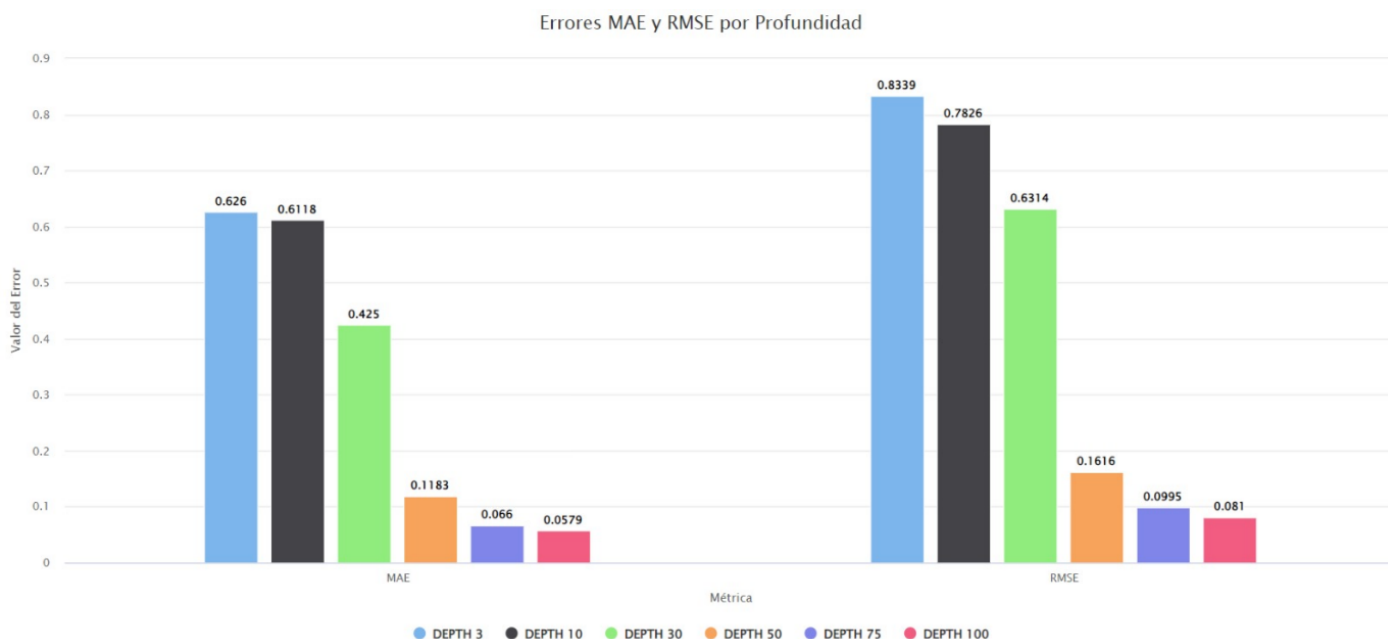


Figura 12*Errores MAE Y RMSE por profundidad [Autor, 2025]*

Si el modelo de Copernicus no capta plenamente el aporte de agua dulce (por resolución o limitaciones en la asimilación de datos de lluvia), se generan discrepancias significativas en la superficie. Estudios previos han demostrado que la variabilidad de corta escala de la salinidad superficial en los trópicos está fuertemente modulada por la lluvia, y los modelos oceánicos tienden a subestimar las fluctuaciones (*Bingham y Brodnitz, 2021*).

Para las capas más profundas (50–100 m) la gráfica suele mostrar una disminución de los errores grande, evidenciada por valores menores de MAE y RMSE. Esto indica que el modelo reproduce con mayor fidelidad las propiedades de salinidad del agua subsuperficial profunda. A esos niveles, las aguas tienden a ser más homogéneas y menos afectadas por forzantes atmosféricos instantáneos, facilitando que el modelo capture su variabilidad de forma más estable. Generalmente, los modelos de reanálisis oceánicos muestran que el error disminuye con la profundidad, según menciona el *SYNTHESIS QUALITY OVERVIEW DOCUMENT* del modelo de Copernicus, en donde resalta que el modelo muestra un RMS bajo con respecto a los datos in situ pero que es mayor en los primeros 200 m, llegando a 0.15 y por debajo de los 200 m a 0.05, sin embargo, presenta

sesgos regionales debido a la complejidad de las dinámicas en la zona del Pacífico (CMEMS, 2023). Es importante mencionar que a esta profundidad coincide la ubicación típica de la haloclina en la región costera ecuatorial, es decir, debajo de la capa de mezcla superficial (usualmente < 30 m), existe un gradiente termo-halino entre 30 y 50 m (Cucalón, 1989), donde pequeñas discrepancias en la posición o intensidad del gradiente de salinidad pueden traducirse en errores considerables. De hecho, validaciones de modelos en otras regiones han reportado los mayores RMSD de salinidad precisamente en capas subsuperficiales cercanas a la haloclina como en el mar Báltico el modelo de reanálisis de Copernicus mostró un RMSE máximo 0,8 g/kg (PSU) alrededor de 30–80 m, disminuyendo a 0,25 g/kg en aguas más profundas (Maljutenko et al., 2023).

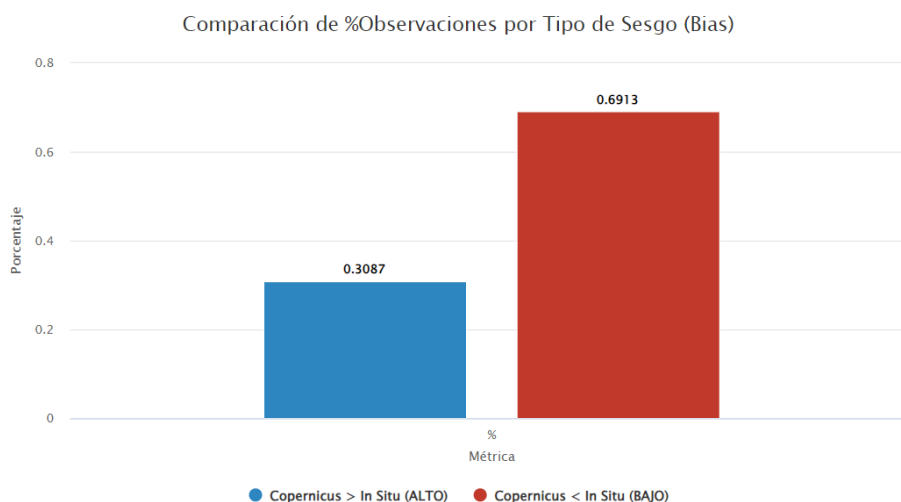
Esto concuerda con el resultado de error que se obtuvo para las primeras capas superficiales; el modelo de Copernicus podría estar posicionando incorrectamente la haloclina, resultando en mayor discrepancia con los datos in situ en ese rango de profundidad.

3.6 Estimación del Modelo

En la siguiente ilustración se observa la comparación del porcentaje de observaciones según el tipo de sesgo (Bias) entre los datos de salinidad de Copernicus y las mediciones in situ. El análisis indica que en 30,87 % de los casos, el producto Copernicus presenta valores superiores a los observados *in-situ* (sesgo positivo), mientras que en 69,13 % de las observaciones, el modelo presenta valores inferiores (sesgo negativo).

Figura 13

Comparación de Subestimación y Sobreestimación del Modelo [Autor, 2025]



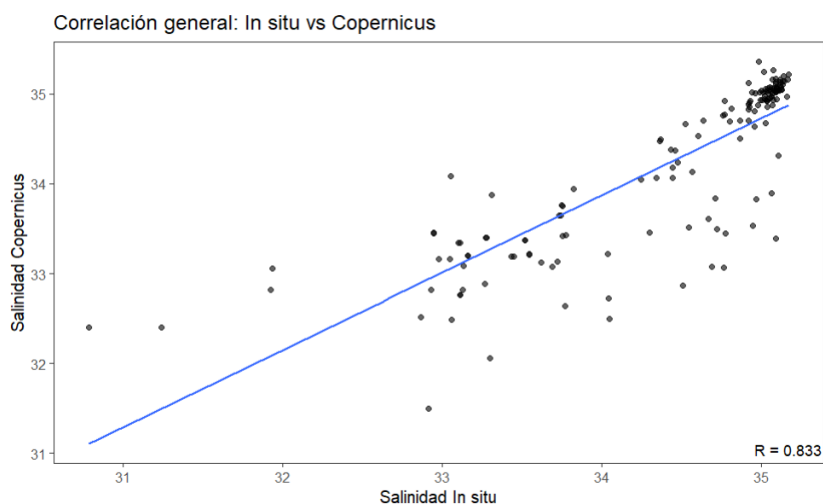
Este predominio de sesgo negativo sugiere que el reanálisis global de Copernicus tiende a subestimar la salinidad en la región costera del Ecuador, lo cual es coherente con hallazgos de investigaciones en entornos tropicales y zonas de descarga continental. Estudios como los de *Dossa et al. (2021)* reportan que los modelos globales, debido a su resolución horizontal limitada y a la escasez de datos de asimilación en áreas costeras, suelen suavizar gradientes halinos y atenuar picos de salinidad superficial, especialmente en regiones con fuerte influencia de corrientes costeras y mezcla estuarina. Por otro lado, el 30,87 % de sesgo positivo podría estar asociado a condiciones puntuales donde el modelo sobreestima la salinidad, posiblemente debido a una representación insuficiente de procesos de aporte de agua dulce (lluvia o ríos) o a diferencias temporales entre las mediciones reales y las simuladas. Tal como señalan *Gou et al. (2024)*, la precisión en la representación de salinidad costera por modelos globales depende fuertemente de la adecuada parametrización de procesos subescala y de la densidad de observaciones usadas para la asimilación de datos.

3.7 Correlación entre datos in situ y del modelo

El diagrama de dispersión global muestra una correlación lineal positiva alta entre ambas fuentes ($R \approx 0,83$), lo que indica que, en promedio, el reanálisis de Copernicus reproduce adecuadamente las variaciones de la salinidad medidas en campo en tu dominio de estudio. Sin embargo, se observa una mayor dispersión en salinidades intermedias–bajas, con un “abanico” de puntos por debajo de ~ 34 , mientras que los valores altos cercanos a ~ 35 los puntos se agrupan más cerca de la recta de ajuste. Este patrón es consistente con la mayor variabilidad halina superficial de la costa ecuatoriana durante parte del año y con el efecto de suavizado espacial propio de la resolución del producto que no siempre captura los gradientes ni aportes locales de agua dulce que son rasgos típicos de ambientes costeros tropicales. La validación local de este producto permite cuantificar su exactitud y mejorar la confianza en su uso, alineándose con recomendaciones científicas recientes. En suma, un R global de 0,833 sugiere que el reanálisis Copernicus reproduce con acierto las condiciones salinas ecuatorianas en promedio, aunque se justifica profundizar en evaluaciones más detalladas por temporada y profundidad para identificar posibles limitaciones.

Figura 14

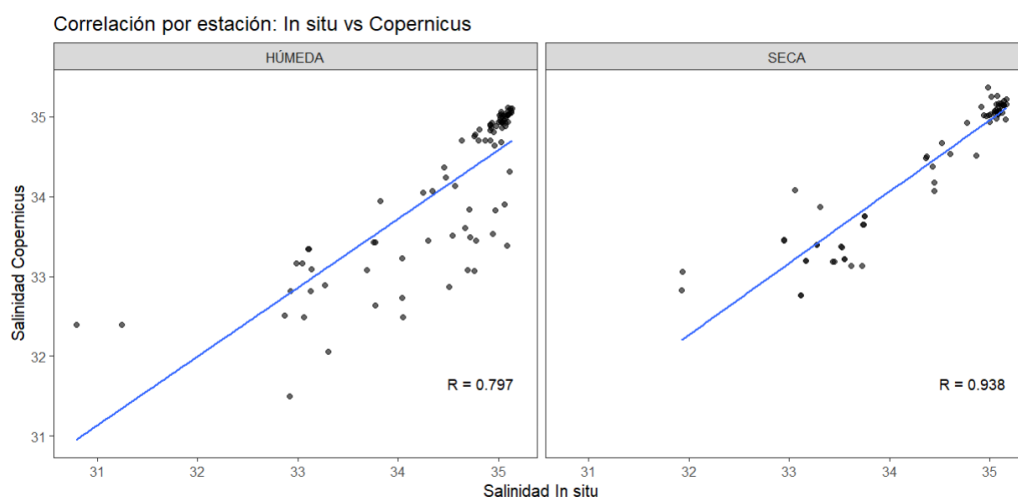
Gráfico de Correlación: Datos In Situ vs Copernicus [Autor, 2025]



Para el análisis estacional tenemos que, en la estación húmeda, $R = 0,797$. La relación sigue siendo fuerte, pero pierde ajuste respecto a la estación seca. En esta época (aprox. dic–may), la región presenta salinidades superficiales relativamente más bajas y gradientes halinos más marcados, debido a lluvias, descarga continental y cambios en la posición/estructura del frente ecuatorial. mientras que en la estación seca (junio–noviembre) aumenta notablemente ($R = 0,938$). Ambos coeficientes indican relaciones altas y positivas, pero la brecha entre ellos revela que la exactitud del modelo varía con la estacionalidad. En la época lluviosa, la dispersión de los puntos es mayor, lo que evidencia discrepancias pronunciadas. Esto sugiere que en meses húmedos el reanálisis Copernicus tiene más dificultad para reproducir ciertos eventos locales de salinidad, posiblemente asociados a entradas de agua dulce superficial por intensas lluvias costeras y escorrentía. En cambio, durante la época seca la nube de dispersión se ajusta estrechamente a la recta 1:1, reflejando una correspondencia casi directa entre las salinidades simuladas y medidas. Un $R \simeq 0,94$ en estación seca implica que el modelo logra capturar el 88% de la variabilidad salina, mostrando un desempeño sobresaliente en condiciones oceánicas más estables y homogéneas.

Figura 15

Correlación por estacionalidad [Autor, 2025]

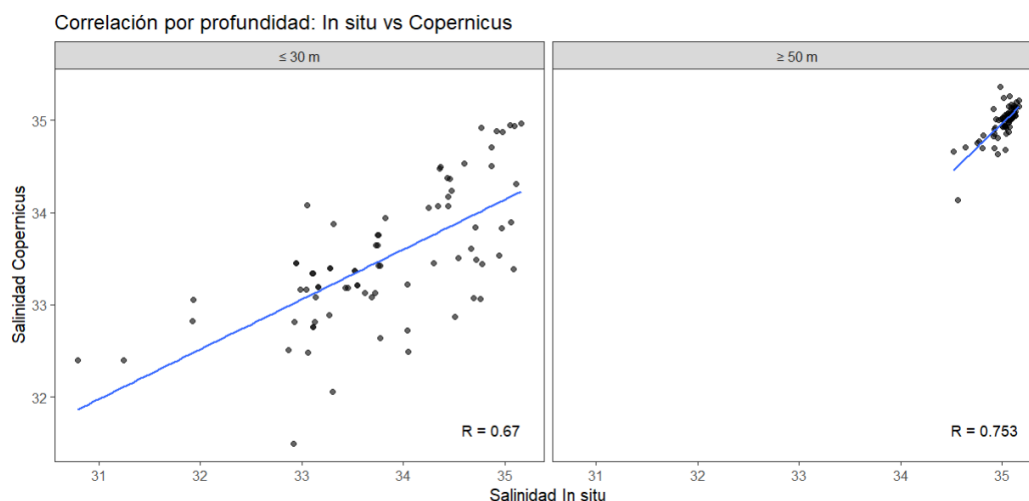


La correlación entre salinidad in situ y modelada en dos rangos de profundidad: en la capa superficial (≤ 30 m) frente a las aguas subsuperficiales (≥ 50 m). Los coeficientes obtenidos revelan una disminución notable de la concordancia cerca de la superficie ($R = 0,67$) en comparación con niveles más profundos ($R = 0,753$). Esto implica que el modelo reproduce con menor fidelidad las variaciones de salinidad superficiales, mientras que a mayor profundidad su desempeño mejora ligeramente.

Los resultados evidencian que la precisión del reanálisis varía con la profundidad debido a procesos que ocurren superficialmente. En la capa ≤ 30 m, la correlación moderada ($R \sim 0,67$) indica mayor dispersión, atribuible a la alta variabilidad y estratificación propias de la superficie. Factores como el aporte de agua dulce por lluvias intensas pueden generar fluctuaciones bruscas de salinidad superficial que el modelo, con su resolución finita y mezcla vertical parametrizada, no reproduce completamente. No obstante, el incremento de correlación en profundidad sugiere que el producto Copernicus representa razonablemente bien la estructura vertical de salinidad en aguas ecuatorianas más allá de la capa superficial.

Figura 16

Correlación por profundidad [Autor, 2025]



3.8 Costos de la validación de datos

CAPÍTULO 4

4. Conclusiones y recomendaciones

4.1 Conclusiones

Se logró completar con éxito el primer objetivo mediante la compilación de un amplio banco de datos de salinidad costera obtenidos con perfiles CTD de salidas de campo del IPIAP. Estos datos abarcan distintas profundidades y ubicaciones representativas de la costa ecuatoriana, cubriendo tanto la columna de agua superficial como niveles medios. Los registros de salinidad se obtuvieron con equipos calibrados, aprovechando que los sensores CTD miden la conductividad eléctrica del agua con muy alta precisión. En conjunto, la base de datos in situ ofrece una referencia de calidad alta para la comparación, al asegurar mediciones precisas y controles de calidad en laboratorio y campo.

La comparación de los datos de salinidad del reanálisis Copernicus con las mediciones in situ mostró concordancia general en la tendencia, pero también discrepancias sistemáticas cuantificadas mediante métricas estadísticas. Se calcularon el sesgo, el error absoluto medio (MAE), la raíz del error cuadrático medio (RMSD) y el error porcentual medio (MAPE) como medidas de validación del conjunto de datos. A nivel general, los valores promedio de MAE fueron del orden de 0.3 y el RMSD de 0.5, con MAPE la diferencia es del 1%. Sin embargo, estos indicadores aumentan y disminuyen cuando clasificamos las diferencias por profundidades y estacionalidad. Para las capas más profundas (a partir de los 50 m) disminuyeron notablemente, por ejemplo, MAE ~ 0.058 a 100 m, reflejando una mejor coincidencia en niveles subsuperficiales, mientras que para las capas superficiales 10 m se evidencia un MAE ~ 0.612 . Lo mismo ocurre por estacionalidad, en donde para la época seca el MAE es de 0.192 y para la época húmeda es de 0.433.

El análisis estadístico (por ejemplo, regresiones lineales y coeficiente de correlación) evidenció correlaciones moderadas-altas entre el producto Copernicus y los datos de campo, aunque con una dispersión significativa en extremos de salinidad. En síntesis, aunque el reanálisis reproduce los valores promedio de salinidad costera, existen sesgos medios (positivos o negativos) y una variabilidad residual que se concreta en los errores

MAE, RMSE y MAPE señalados, variando según profundidad y estacionalidad típica de la costa ecuatoriana.

El nivel de exactitud (sesgo sistemático) y precisión (variabilidad de los errores) del producto Copernicus muestra dependencia con la profundidad y los procesos oceanográficos locales ocurridos durante los diferentes meses del año. En general, el reanálisis exhibió un desempeño más robusto en capas medias y profundas de la columna de agua que en superficie. En los primeros 10–30 m se observó una mayor dispersión de los datos, atribuible a la presencia de haloclinas estacionales y entrada de agua dulce, lo que incrementó los errores (por ejemplo, MAE superficial 0.5–0.7). A medida que aumenta la profundidad, los errores disminuyeron (MAE ~0.05–0.1 entre 50–100 m), indicando mejor “precisión y exactitud” en aguas más profundas. Estos resultados sugieren que procesos como las lluvias intensas (que forman una capa menos salina en superficie) y la entrada de la Corriente de Humboldt que crea gradientes verticales marcados influyeron en el desempeño del modelo. De hecho, el reanálisis tiende a presentar un sesgo negativo en superficie durante épocas lluviosas (subestimando la salinidad debido a que no captura del todo la dilución por esorrentía). Aunque la comparación de los datos de salinidad del reanálisis Copernicus con las mediciones *in situ* mostró concordancia general en la tendencia, pero también discrepancias sistemáticas, aunque el reanálisis reproduce los valores promedio de salinidad costera, existen sesgos medios (positivos o negativos) y una variabilidad residual que se concreta en los errores MAE, RMSE y MAPE señalados, variando según profundidad y estacionalidad típica de la costa ecuatoriana, lo cual nos permite concluir que el modelo presenta confiabilidad en las estimaciones a partir de los 50 metros, por lo que se destaca la capacidad del modelo para captar estructuras subsuperficiales de salinidad (por ejemplo, gradientes de haloclina profunda) gracias a la asimilación de perfiles *in-situ* y a su esquema de corrección de sesgos, aunque la respuesta en la capa de mezcla superficial no es confiable.

La validación reveló que el error del reanálisis Copernicus en la costa ecuatoriana es dependiente de la temporada climática y la profundidad. La región presenta una estación húmeda (diciembre-mayo) con lluvias intensas y una estación seca (junio-noviembre) con temperaturas más bajas e intensificación de la corriente de Humboldt. Durante la época húmeda, el modelo mostró diferencias mayores. En contraste, en la época seca se observó una reducción del error, pero para ambas épocas el modelo tiende a subestimar ligeramente la salinidad real, posiblemente por no captar los cambios que ocurren a nivel superficial. En consecuencia, la exactitud y precisión del reanálisis es mayor en situaciones de condiciones oceánicas “estacionarias” por ejemplo, período seco estable y menor durante eventos extremos o transitorios (época de lluvias intensas), lo que destaca la sensibilidad del modelo a la estacionalidad regional.

En conjunto, el reanálisis global GLORYS12V1 de Copernicus demostró ser una herramienta para el monitoreo costero, pero no exenta de limitaciones. Gracias a su alta resolución horizontal ($1/12^\circ$, ~ 8 km) y 50 niveles verticales, así como a la asimilación de perfiles *in-situ* de salinidad, el producto logra un desempeño superior al de reanálisis de menor resolución. De hecho, la literatura destaca que los sistemas como GLORYS12 tienden a mostrar un mayor desempeño costero comparado con otros productos de reanálisis (*Feng et al., 2024*). No obstante, a pesar de estas ventajas, persisten limitaciones relevantes: la resolución de $1/12^\circ$ sigue siendo insuficiente para representar plenamente los procesos de mesoescala costera (como corrientes intensas cerca de la costa, pequeñas corrientes fluviales y discontinuidades abruptas de salinidad), y el modelo puede extrapolar erróneamente condiciones de agua profunda en zonas someras. Además, la frecuencia temporal (diaria/mensual) limita la captura de cambios rápidos en respuesta a precipitaciones puntuales.

4.2 Recomendaciones

- A partir de los errores identificados (MAE, RMSE, MAPE), es pertinente implementar un modelo de corrección empírica que ajuste los datos de Copernicus frente a las

observaciones *in-situ*. Este modelo podría basarse en técnicas de regresión espacial, análisis multivariado o enfoques de *machine learning*, con el fin de reducir los sesgos sistemáticos detectados en superficie y mejorar la precisión del producto en aguas costeras.

- Copernicus provee un panorama coherente de la salinidad litoral y facilita la identificación de tendencias generales, debe usarse con cautela como complemento de las observaciones in situ, especialmente cuando se requiera alta precisión local. En términos prácticos, estos resultados indican que el producto es adecuado para seguimiento y planificación a escala regional y para profundidades a partir de los 50 m, ya que, sus limitaciones en capas superficiales y procesos costeros sugieren validar siempre sus estimaciones con datos de campo adicionales.
- El IPIAP, como instituto de investigación, debe promover un muestreo mensual de la variabilidad climática con el fin de disponer de valores reales y actualizados. Esta información permitirá construir rutinas de corrección que posibiliten la aplicación de modelos ajustados a dichas correcciones, especialmente en el análisis de las condiciones a nivel superficie

Bibliografía

- Alory, G., Maes, C., Delcroix, T., Reul, N., y Illig, S. (17 de abril de 2012). Seasonal dynamics of sea surface salinity off Panama: The far Eastern Pacific Fresh Pool. <https://doi.org/https://doi.org/10.1029/2011JC007802>
- Astro Córdova, C. (2018). Variabilidad Espacio Temporal del Frente Ecuatorial en el Océano Pacífico Oriental. Lima, Perú. https://repositorio.upch.edu.pe/bitstream/handle/20.500.12866/1425/Variabilidad_AstroCordova_Cristhian.pdf?sequence=1
- Basharat , J., y Naiem, A. (October de 2017). Comparative analysis of diffuse solar radiation models based on sky-clearness index and sunshine period for humid-subtropical climatic region of India: A case study. 78. <https://doi.org/https://doi.org/10.1016/j.rser.2017.04.073>
- Bautista Zúñiga, F. (2007). Técnicas de Muestreo para Manejadores de Recursos Naturales. https://d1wqtxts1xzle7.cloudfront.net/98726382/tmuestreo.pdf?1676513291=&response-content-disposition=inline%3B+filename%3DTecnicas_de_muestreo_para_manejadores_de.pdf&Expires=1754026673&Signature=DV-5Y-4YqH6Ur39k9Yy3dq~mnzAaAW1ri7jy17GpymFKQrxvhKyELEAfSx
- Bingham, F., y Brodnitz, S. (21 de October de 2021). Sea surface salinity short-term variability in the tropics. 17. <https://doi.org/https://doi.org/10.5194/os-17-1437-2021>
- Chapra, S., y Canale, R. (2006). Métodos Numéricos para Ingenieros . McGrall-Hill.
- CMEMS. (22 de December de 2023). SYNTHESIS QUALITY OVERVIEW DOCUMENT . 1.6. <https://documentation.marine.copernicus.eu/SQO/CMEMS-GLO-SQO-001-030.pdf>

- CPPS, C. P. (2015). Protocolo del Uso del CTD Sea-Bird y Procesamiento de Datos.
https://repository.oceanbestpractices.org/bitstream/handle/11329/1597/cpps_gte_bd_002.pdf?sequence=1&isAllowed=y
- Cucalón, E. (1986). Bases Biológicas y Marco Conceptual para el Manejo de los Recursos Pelágicos en el Pacífico Sur Oriental. En *Sinopsis de la Oceanografía de la Región Frente a las Costas de Ecuador* (págs. 30-33). Lima-Perú.
- Cucalón, E. (1989). Oceanographic Characteristics Off the Coast of Ecuador. (S. Olsen, y L. Arriaga, Edits.) *A Sustainable Shrimp Mariculture Industry for Ecuador*.
https://www.crc.uri.edu/download/ShrimpBook_Ocean_Cucalon.pdf#:~:text=Also%20at%20this%20time%2C%20the,is%20also%20characterized%20by%20high
- Cucalon, E. (1989). Oceanographic Characteristics Off the Coast of Ecuador.
- Dossa, A., Alory, G., Costa Da Silva, A., Dahunsi, A., y Bertrand, A. (26 de June de 2021).
 Global Analysis of Coastal Gradients of Sea Surface Salinity.
<https://doi.org/https://doi.org/10.3390/rs13132507>
- Durán, C. M. (1994). *Interpolación y Suavizamiento con Splines*. Universidad Autónoma Nacional de México.
<https://ru.dgb.unam.mx/jspui/bitstream/20.500.14330/TES01000208200/3/0208200.pdf>
- ESRI. (s.f.). ¿Qué son los datos netCDF?
<https://desktop.arcgis.com/es/arcmap/latest/manage-data/netcdf/what-is-netcdf-data.htm>
- ESRI. (s.f.). Análisis de Regresión. Retrieved 25 de julio de 2025, from
<https://pro.arcgis.com/es/pro-app/latest/tool-reference/spatial-statistics/regression-analysis-basics.htm>
- Feng, X., Widlansky, M., Balmaseda, M., Zuo, H., Spillman, C., Smith, G., . . . Sweet, W. (21 de February de 2024). Improved capabilities of global ocean reanalyses for analysing sea level variability near the Atlantic and Gulf of Mexico Coastal U.S.
<https://doi.org/https://doi.org/10.3389/fmars.2024.1338626>

Gilbert, L. (28 de Agosto de 2023). MAE Y MAPE: ¿Qué métrica es mejor?

<https://medium.com/trusted-data-science-haleon/mape-vs-mae-which-metric-is-better-68dd559cbfb1>

Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R., Garric, G., Melet, A., Drévillon, M., . . .

Regnier, C. (20 de July de 2021). The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis. 9. <https://doi.org/https://doi.org/10.3389/feart.2021.698876>

Lewis, E. L. (1980). The Practical Salinity Scale 1978 and Its Antecedents. *IEEE Journal of Oceanic Engineering*, 1.

<https://www.sciencedirect.com/science/article/abs/pii/0198014981900029>

Li, Y., Church, J., McDougall, T., y Barker, P. (December de 2022). Sensitivity of

Observationally Based Estimates of Ocean Heat Content and Thermal Expansion to Vertical Interpolation Schemes.

<https://doi.org/http://dx.doi.org/10.1029/2022GL101079>

Maljutenko, I., Raudsepp, U., y Lagema, P. (2023). *Ocean Monitoring Indicator (OMI) of the Copernicus Marine Environment Monitoring Service Baltic Sea Cod reproductive volume*. Quality Information Document, CMEMS.

https://documentation.marine.copernicus.eu/QUID/CMEMS-OMI-QUID-BALTIC_HEALTH_codt_volume.pdf#:~:text=of%20reanalysis%20show%20only%20small,with%20the%20observations%20from%20different

McDougall, T., Feistel, R., Millero, F., Jackett, D., Wright, D., King, B., . . . Seitz, S. (2009).

The International Thermodynamic Equation Of Seawater 2010 (TEOS-10):

Calculation and Use of Thermodynamic Properties. [https://scor-int.org/Publications/Thermodynamic_TEOS-](https://scor-int.org/Publications/Thermodynamic_TEOS-10_Manual_GOSHIP_19Aug09.pdf#:~:text=The%20Practical%20Salinity%20Scale%2C%20PSS,algorithm%20for%20the%20specific)

[10_Manual_GOSHIP_19Aug09.pdf#:~:text=The%20Practical%20Salinity%20Scale%2C%20PSS,algorithm%20for%20the%20specific](https://scor-int.org/Publications/Thermodynamic_TEOS-10_Manual_GOSHIP_19Aug09.pdf#:~:text=The%20Practical%20Salinity%20Scale%2C%20PSS,algorithm%20for%20the%20specific)

McDougall, T., Feistel, R., y Pawlowicz, R. (2013). Ocean Circulation & Climate. En *Chapter 6 - Thermodynamics of Seawater* (Vol. 103, págs. 141-158). International Geophysics.

<https://www.sciencedirect.com/science/article/abs/pii/B9780123918512000064>

1.pdf?sequence=1&isAllowed=y#:~:text=TEOS,agua%20de%20mar%20en%20el

Millero, F., Feistel, R., y Wright, D. (2008). The composition of Standard Seawater and the definition of the Reference-Composition Salinity Scale. *Deep Sea Research Part I: Oceanographic Research Papers*, 55, 50-72.

<https://www.sciencedirect.com/science/article/abs/pii/S0967063707002282>

Mitchell, T. P., y Wallace, J. M. (1 de October de 1992). The annual cycle in equatorial convection and sea surface temperature. *Journal of Climate*, 5(10), 1140–1156.

[https://doi.org/ https://doi.org/10.1175/1520-0442\(1992\)005<1140:TACIEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1992)005<1140:TACIEC>2.0.CO;2)

Morales España, G., Mora Florez, J., y Vargas Torres, H. (Septiembre de 2008). Estrategia de regresión basada en el método de los k vecinos más cercanos para la estimación de la distancia de falla en sistemas radiales.

[http://www.scielo.org.co/scielo.php?pid=S0120-](http://www.scielo.org.co/scielo.php?pid=S0120-62302008000300009&script=sci_arttext)

[62302008000300009&script=sci_arttext](http://www.scielo.org.co/scielo.php?pid=S0120-62302008000300009&script=sci_arttext)

Naciones Unidas. (2015). <https://sdgs.un.org/goals/goal14>

Navarro Alberto, J., Pacheco Ávila, J., Cabrera Sansores, A., y Luna Pabello, V. (2011).

Técnicas de Muestreo para Manejadores de Recursos Naturales. México.

[https://d1wqtxts1xzle7.cloudfront.net/98726382/tmuestreo.pdf?1676513291=&response-content-](https://d1wqtxts1xzle7.cloudfront.net/98726382/tmuestreo.pdf?1676513291=&response-content-disposition=inline%3B+filename%3DTecnicas_de_muestreo_para_manejadores_de.pdf&Expires=1754026673&Signature=DV-5Y-4YqH6Ur39k9Yy3dq~mnzAaAW1ri7jy17GpymFKQrxvhKyELEAfSx)

[disposition=inline%3B+filename%3DTecnicas_de_muestreo_para_manejadores_de.pdf&Expires=1754026673&Signature=DV-5Y-](https://d1wqtxts1xzle7.cloudfront.net/98726382/tmuestreo.pdf?1676513291=&response-content-disposition=inline%3B+filename%3DTecnicas_de_muestreo_para_manejadores_de.pdf&Expires=1754026673&Signature=DV-5Y-4YqH6Ur39k9Yy3dq~mnzAaAW1ri7jy17GpymFKQrxvhKyELEAfSx)

[4YqH6Ur39k9Yy3dq~mnzAaAW1ri7jy17GpymFKQrxvhKyELEAfSx](https://d1wqtxts1xzle7.cloudfront.net/98726382/tmuestreo.pdf?1676513291=&response-content-disposition=inline%3B+filename%3DTecnicas_de_muestreo_para_manejadores_de.pdf&Expires=1754026673&Signature=DV-5Y-4YqH6Ur39k9Yy3dq~mnzAaAW1ri7jy17GpymFKQrxvhKyELEAfSx)

North, R. P., y Livingstone, D. M. (April de 2013). Comparison of linear and cubic spline methods of interpolating lake profiles.

<https://doi.org/http://dx.doi.org/10.4319/lom.2013.11.213>

Olumide, S. (8 de Agosto de 2023). Error cuadrático medio (RMSE) en IA.

[https://arize.com/blog-course/root-mean-square-error-rmse-what-you-need-to-](https://arize.com/blog-course/root-mean-square-error-rmse-what-you-need-to-know/#:~:text=El%20error%20cuadr%C3%A1tico%20medio%20(RMSE,significativos%20y%20pron%C3%B3sticos%20menos%20precisos.)

[know/#:~:text=El%20error%20cuadr%C3%A1tico%20medio%20\(RMSE,significativos%20y%20pron%C3%B3sticos%20menos%20precisos.](https://arize.com/blog-course/root-mean-square-error-rmse-what-you-need-to-know/#:~:text=El%20error%20cuadr%C3%A1tico%20medio%20(RMSE,significativos%20y%20pron%C3%B3sticos%20menos%20precisos.)

Pawlowicz, R. (2013). Key Physical Variables in the Ocean: Temperature, Salinity, and Density. *Nature Education* .

Perugachi-Salamea, C., González-Naváez, M., Pambabay-Calero, J., García-Arévalo, I., Vargas-Ayala, F., Shigla-Cuji, G., & Nath-Nieto, J. (2014). Caracterización oceanográfica del perfil costero ecuatoriano a través de métodos estadísticos aplicados a las estaciones fijas de monitoreo del INOCAR, serie A partir de 1981. *Acta Oceanográfica del Pacífico*, 19(2), 13–33.

https://www.inocar.mil.ec/web/phocadownloadpap/actas_oceanograficas/acta19_2/OCE1902_2.pdf

Quispe-Ccalluari, C., y Quispe Ramos, D. (2025). Validación del producto Glorys12v1 (Mercator Ocean) para monitoreo de temperatura y salinidad del mar peruano. Perú. <https://repositorio.imarpe.gob.pe/bitstream/20.500.12958/9995/5/Informe%2052-1%20c%286%29.pdf>

Recalde Mosquera, S. (2014). *Variación estacional de la salinidad de la capa de mezcla en una estación costera del Pacífico Ecuatorial Oriental*. Universidad de Concepción. Facultad de Ciencias Naturales y Oceanográficas., Departamento de Oceanografía. <https://repositorio.udec.cl/server/api/core/bitstreams/630c6703-aaaf-4f62-8d23-ba2502dd23df/content>

Ridgway, K., Dunn, J., y Wilkin, J. (September de 2002). Ocean interpolation by four-dimensional weighted least squares - Application to the waters around Australasia. [https://doi.org/https://doi.org/10.1175/1520-0426\(2002\)019%3C1357:OIBFDW%3E2.0.CO;2](https://doi.org/https://doi.org/10.1175/1520-0426(2002)019%3C1357:OIBFDW%3E2.0.CO;2)

Riofrío, M., & Chavarría, J. (2004). Caracterización termohalina de la estación oceanográfica El Pelado [Informe técnico]. Centro Nacional de Acuicultura e Investigaciones Marinas (CENAIM-ESPOL). <http://www.cenaim.espol.edu.ec/acuiclim/index.html>

Ruiz Armenteros, A. M., García Balboa , J. L., y Mesa Mingorance, J. L. (2010). ERROR, INCERTIDUMBRE, PRECISIÓN Y EXACTITUD, TÉRMINOS ASOCIADOS A LA CALIDAD ESPACIAL DEL DATO GEOGRÁFICO.

Sánchez, R. (2015). Prueba de Wilcoxon-Mann-Whitney: mitos y realidades.

<https://biblat.unam.mx/hevila/Revistamexicanadeendocrinologiametabolismo&nutricion/2015/vol2/no1/3.pdf>

Shcherniba, A., D'Asaro, E., y Harcourt, R. (14 de June de 2019). Rain and Sun Create Slippery Layers in the Eastern Pacific Fresh Pool.

<https://doi.org/https://doi.org/10.5670/oceanog.2019.217>

Sonnenholzner, S. (1993). Variabilidad Oceanográfica Subsuperficial Frente a la Costa Ecuatorial Asociada a la Subcorriente Ecuatorial (1984-1990). Guayaquil.

<https://www.dspace.espol.edu.ec/handle/123456789/59587>

Tang, B., Zhao, D., Chaoran, C., y Zhao, X. (September de 2022). Reconstruction of ocean temperature and salinity profiles in the Northern South China Sea using satellite observations. *Frontiers*. <https://doi.org/https://doi.org/10.3389/fmars.2022.945835>

Torres, G., Carnicer, O., Canepa, A., De La Fuente, P., Recalde, S., Narea, R., . . . Borbor-Córdova, M. (26 de marzo de 2019). Spatio-Temporal Pattern of Dinoflagellates Along the Tropical Eastern Pacific Coast (Ecuador). 6.

<https://doi.org/https://doi.org/10.3389/fmars.2019.00145>

Villarreal, A. (2022). *Análisis del comportamiento climatológico anual y estacional del agua subsuperficial subtropical en el Atlántico Norte y su influencia en el mar Caribe*. Barranquilla, Colombia. <https://manglar.uninorte.edu.co/bitstream/handle/10584/11457/108517102>

Wieser, M. (2006). ATOMIC WEIGHTS OF THE ELEMENTS 2005 (IUPAC TECHNICAL REPORT). En *INORGANIC CHEMISTRY DIVISION COMMISSION ON ISOTOPIC ABUNDANCES AND ATOMIC WEIGHTS** (Vol. 78, págs. 2051–2066).

<https://old.iupac.org/publications/pac/2006/pdf/7811x2051.pdf>

Apéndice

Código de implementación en R Studio

```
library(readxl)
library(gstat)
library(sp)
library(dplyr)
library(openxlsx)
library(geosphere)

rm(list = ls())

# CONSUMIENDO INFO DE COPERNICUS EN BRUTO, TAL CUAL COMO VIENE
copernicus_bruto <- read_excel("C:/Users/israe/Downloads/COPERNICUS DATOS EN
BRUTO.xlsx",
                             col_types = c("numeric", "numeric", "numeric",
                                           "date", "numeric"))

# CONSUMIENDO DATOS IN SITU
insitu = read_excel("C:/Users/israe/Desktop/datos/Tesis/dato_insitu.xlsx",
                   col_types = c("text", "text", "date",
                                "numeric", "numeric", "numeric",
                                "numeric")) %>%
  filter(TYPE == "Variabilidad Climática" & SALINITY > 30)

colnames(insitu)[3] = "FECHA"

##### INFO DE COPERNICUS QUE SEA DE LA MISMA FECHA Y POSICION DE
INSITU #####
copernicus_bruto = copernicus_bruto %>%
  mutate(ID = paste(Fechas, Latitud, Longitud, sep = " "))

insitu = insitu %>%
  mutate(ID = paste(FECHA, LATITUDE, LONGITUDE, sep = " "))

copernicus_bruto = copernicus_bruto %>%
  inner_join(
    insitu %>% distinct(FECHA) %>% rename(Fechas = FECHA)
  )

insitu %>% distinct(DEPTH)

estimar_profundidad = copernicus_bruto %>%
  filter(ID == "2022-11-24 -4 -81.91666") %>%
  print(n = 23)

View(
  copernicus_bruto %>%
    group_by(ID) %>%
    summarise(N = n()) #>%
```

```

#group_by(N) %>%
#summarise(n()) %>%
#rename(N_PROFUNDIDADES = N)
)

```

ESTIMANDO SALINIDAD DE COPERNICUS PARA DIFERENTES PROFUNDIDADES - CASO PARTICULAR

```

# Cargar librerías necesarias
library(dplyr)
library(highcharter)
library(tidyr)

```

```

# Paso 0: Definir profundidades objetivo
prof_objetivo <- c(3, 10, 30, 50, 75, 100)

```

```

# Paso 1: Separar datos de entrenamiento y test
set.seed(123)
caso_particular <- copernicus_bruto %>%
  filter(ID == "2022-11-24 -3.75 -81.5") %>%
  arrange(Profundidad)

```

```

n <- nrow(caso_particular)
idx_test <- sample(1:n, size = floor(0.2 * n))
datos_test <- caso_particular[idx_test, ]
datos_train <- caso_particular[-idx_test, ]

```

```

# Paso 2: Aplicar los 5 métodos de interpolación
# 1. Lineal
pred_lineal <- approx(x = datos_train$Profundidad, y = datos_train$Salinidad,
  xout = datos_test$Profundidad, rule = 2)$y

```

```

# 2. Spline
pred_spline <- spline(x = datos_train$Profundidad, y = datos_train$Salinidad,
  xout = datos_test$Profundidad)$y

```

```

# 3. Loess
modelo_loess <- loess(Salinidad ~ Profundidad, data = datos_train, span = 0.5)
pred_loess <- predict(modelo_loess, newdata = data.frame(Profundidad =
  datos_test$Profundidad))

```

```

# 4. Polinomio grado 2
modelo_poly2 <- lm(Salinidad ~ poly(Profundidad, 2), data = datos_train)
pred_poly2 <- predict(modelo_poly2, newdata = datos_test)

```

```

# 5. Polinomio grado 3
modelo_poly3 <- lm(Salinidad ~ poly(Profundidad, 3), data = datos_train)
pred_poly3 <- predict(modelo_poly3, newdata = datos_test)

```

```

# Paso 3: Comparar errores
resultados <- tibble(
  Profundidad = datos_test$Profundidad,
  Real = datos_test$Salinidad,
  Lineal = pred_lineal,
  Spline = pred_spline,

```

```

Loess = pred_loess,
Poly2 = pred_poly2,
Poly3 = pred_poly3
) %>%
mutate(
  error_lineal = abs(Real - Lineal),
  error_spline = abs(Real - Spline),
  error_loess = abs(Real - Loess),
  error_poly2 = abs(Real - Poly2),
  error_poly3 = abs(Real - Poly3)
)

# Paso 4: Determinar el mejor método
error_total <- resultados %>%
  summarise(
    Lineal = mean(error_lineal, na.rm = TRUE),
    Spline = mean(error_spline, na.rm = TRUE),
    Loess = mean(error_loess, na.rm = TRUE),
    Poly2 = mean(error_poly2, na.rm = TRUE),
    Poly3 = mean(error_poly3, na.rm = TRUE)
  )

mejor_metodo <- names(which.min(error_total))

# Paso 5: Visualizar con highcharter
hc <- highchart() %>%
  hc_title(text = "Selección del Mejor Modelo") %>%
  hc_xAxis(title = list(text = "Profundidad (m)")) %>%
  hc_yAxis(title = list(text = "Salinidad")) %>%
  hc_add_series(data = list_parse2(datos_train %>% select(Profundidad, Salinidad)),
    name = "Entrenamiento", type = "scatter",
    hcaes = list(x = "Profundidad", y = "Salinidad")) %>%
  hc_add_series(data = list_parse2(datos_test %>% select(Profundidad, Salinidad)),
    name = "Test Real", type = "scatter",
    hcaes = list(x = "Profundidad", y = "Salinidad")) %>%
  hc_add_series(data = list_parse2(resultados %>% select(Profundidad, Lineal)),
    name = "Lineal", type = "line",
    hcaes = list(x = "Profundidad", y = "Lineal")) %>%
  hc_add_series(data = list_parse2(resultados %>% select(Profundidad, Spline)),
    name = "Spline", type = "line",
    hcaes = list(x = "Profundidad", y = "Spline")) %>%
  hc_add_series(data = list_parse2(resultados %>% select(Profundidad, Loess)),
    name = "Loess", type = "line",
    hcaes = list(x = "Profundidad", y = "Loess")) %>%
  hc_add_series(data = list_parse2(resultados %>% select(Profundidad, Poly2)),
    name = "Polinomio Grado 2", type = "line",
    hcaes = list(x = "Profundidad", y = "Poly2")) %>%
  hc_add_series(data = list_parse2(resultados %>% select(Profundidad, Poly3)),
    name = "Polinomio Grado 3", type = "line",
    hcaes = list(x = "Profundidad", y = "Poly3"))

print(hc)

# Paso 6: Estimar en profundidades objetivo usando el mejor método
if (mejor_metodo == "Lineal") {

```

```

    est <- approx(x = datos_train$Profundidad, y = datos_train$Salinidad,
                  xout = prof_objetivo, rule = 2)
  } else if (mejor_metodo == "Spline") {
    est <- spline(x = datos_train$Profundidad, y = datos_train$Salinidad,
                  xout = prof_objetivo)
  } else if (mejor_metodo == "Loess") {
    est <- data.frame(
      x = prof_objetivo,
      y = predict(modelo_loess, newdata = data.frame(Profundidad = prof_objetivo))
    )
  } else if (mejor_metodo == "Poly2") {
    est <- data.frame(
      x = prof_objetivo,
      y = predict(modelo_poly2, newdata = data.frame(Profundidad = prof_objetivo))
    )
  } else {
    est <- data.frame(
      x = prof_objetivo,
      y = predict(modelo_poly3, newdata = data.frame(Profundidad = prof_objetivo))
    )
  }
}

```

```

estimacion_df <- tibble(
  Profundidad = est$x,
  Salinidad_estimada = est$y
)

```

Paso 7: Gráfico final con estimaciones

```

hc2 <- highchart() %>%
  hc_title(text = paste("Estimaciones finales con método:", mejor_metodo)) %>%
  hc_xAxis(title = list(text = "Profundidad (m)")) %>%
  hc_yAxis(title = list(text = "Salinidad")) %>%
  hc_add_series(data = list_parse2(caso_particular %>% select(Profundidad, Salinidad)),
                name = "Datos Reales", type = "scatter",
                hcaes = list(x = "Profundidad", y = "Salinidad")) %>%
  hc_add_series(data = list_parse2(estimacion_df %>% select(Profundidad,
Salinidad_estimada)),
                name = "Estimado", type = "line",
                hcaes = list(x = "Profundidad", y = "Salinidad_estimada"))

```

```

hc2 <- highchart() %>%
  hc_title(text = "Interpolación con el Mejor Modelo") %>%
  hc_xAxis(title = list(text = "Profundidad (m)")) %>%
  hc_yAxis(title = list(text = "Salinidad")) %>%
  hc_add_series(data = list_parse2(caso_particular %>% select(Profundidad, Salinidad)),
                name = "Datos Reales", type = "scatter",
                hcaes = list(x = "Profundidad", y = "Salinidad")) %>%
  hc_add_series(data = list_parse2(estimacion_df %>% select(Profundidad,
Salinidad_estimada)),
                name = "Estimado", type = "line",
                hcaes = list(x = "Profundidad", y = "Salinidad_estimada"))

```

```

print(hc2)

```

```
# Mostrar tabla final
estimacion_df
```

```
##### FUNCIONES PARA ESTIMAR SALINIDAD DE COPERNICUS PARA
DIFERENTES PROFUNDIDADES - CASO GENERAL
#####
```

```
library(dplyr)
library(purrr)
library(tibble)
```

```
# Vector global de profundidades a estimar
prof_objetivo_global <- c(3, 10, 30, 50, 75, 100)
```

```
# Función para interpolar por ID
interpolar_por_id <- function(id_actual, df_total) {
  datos_id <- df_total %>% filter(ID == id_actual) %>% arrange(Profundidad)
```

```
  # Validación: mínimo 6 puntos
  if (nrow(datos_id) < 6) return(NULL)
```

```
  # Separar entrenamiento y testeo (20%)
  set.seed(123)
  n <- nrow(datos_id)
  idx_test <- sample(1:n, size = floor(0.2 * n))
  datos_test <- datos_id[idx_test, ]
  datos_train <- datos_id[-idx_test, ]
```

```
  # Aplicar 5 métodos
  pred_lineal <- approx(datos_train$Profundidad, datos_train$Salinidad,
    xout = datos_test$Profundidad, rule = 2)$y
  pred_spline <- spline(datos_train$Profundidad, datos_train$Salinidad,
    xout = datos_test$Profundidad)$y
  modelo_loess <- loess(Salinidad ~ Profundidad, data = datos_train, span = 0.5)
  pred_loess <- predict(modelo_loess, newdata = datos_test)
  modelo_poly2 <- lm(Salinidad ~ poly(Profundidad, 2), data = datos_train)
  pred_poly2 <- predict(modelo_poly2, newdata = datos_test)
  modelo_poly3 <- lm(Salinidad ~ poly(Profundidad, 3), data = datos_train)
  pred_poly3 <- predict(modelo_poly3, newdata = datos_test)
```

```
  # Evaluar errores porcentuales (MAPE)
  errores <- tibble(
    Real = datos_test$Salinidad,
    Lineal = pred_lineal,
    Spline = pred_spline,
    Loess = pred_loess,
    Poly2 = pred_poly2,
    Poly3 = pred_poly3
  ) %>%
```

```

summarise(
  Lineal = mean(abs((Real - Lineal) / Real) * 100, na.rm = TRUE),
  Spline = mean(abs((Real - Spline) / Real) * 100, na.rm = TRUE),
  Loess = mean(abs((Real - Loess) / Real) * 100, na.rm = TRUE),
  Poly2 = mean(abs((Real - Poly2) / Real) * 100, na.rm = TRUE),
  Poly3 = mean(abs((Real - Poly3) / Real) * 100, na.rm = TRUE)
)

mejor_metodo <- names(which.minerrores))
mejor_error <- errores[[mejor_metodo]]

# Profundidades válidas para interpolar
max_prof <- max(datos_id$Profundidad, na.rm = TRUE)
prof_objetivo_validas <- prof_objetivo_global[prof_objetivo_global <= max_prof]

if (length(prof_objetivo_validas) == 0) return(NULL)

# Estimar salinidad en las profundidades válidas usando el mejor método
predicciones <- switch(mejor_metodo,
  "Lineal" = approx(datos_train$Profundidad, datos_train$Salinidad,
    xout = prof_objetivo_validas, rule = 2),
  "Spline" = spline(datos_train$Profundidad, datos_train$Salinidad,
    xout = prof_objetivo_validas),
  "Loess" = data.frame(x = prof_objetivo_validas,
    y = predict(modelo_loess, newdata = data.frame(Profundidad
= prof_objetivo_validas))),
  "Poly2" = data.frame(x = prof_objetivo_validas,
    y = predict(modelo_poly2, newdata = data.frame(Profundidad
= prof_objetivo_validas))),
  "Poly3" = data.frame(x = prof_objetivo_validas,
    y = predict(modelo_poly3, newdata = data.frame(Profundidad
= prof_objetivo_validas)))
)

tibble(
  Longitud = unique(datos_id$Longitud),
  Latitud = unique(datos_id$Latitud),
  Fechas = unique(datos_id$Fechas),
  ID = id_actual,
  Profundidad = predicciones$x,
  Salinidad = predicciones$y,
  Modelo = mejor_metodo,
  Error_Porcentual = mejor_error
)
}

# Función general para muchos IDs
interpolar_grupo_ids <- function(ids_df, df_total) {
  resultados <- ids_df$ID %>%
    map_dfr(~ interpolar_por_id(.x, df_total))
  return(resultados)
}

##### EJECUCION PARA IDS DEFINIDOS - HALLANDO COPERNICUS
PARA LAS PROFUNDIDADES REQUERIDAS #####

```



```
k=1000
```

```
k = copernicus_bruto %>%  
  distinct(ID) %>%  
  summarise(n()) %>% pull
```

```
#k=1000
```

```
ids_definidos = copernicus_bruto %>%  
  distinct(Latitud,Longitud,Fechas,ID) %>%  
  sample_n(k)
```

```
ids_definidos
```

```
start_time <- Sys.time()
```

```
# Ejecutar interpolación sobre grupo  
resultados_estimacion <- interpolar_grupo_ids(ids_definidos, copernicus_bruto)
```

```
end_time <- Sys.time()  
duracion <- end_time - start_time
```

```
cat("Tiempo de ejecución total:", round(as.numeric(duracion, units = "secs")), "segundos\n")  
cat("≈", round(as.numeric(duracion, units = "mins"), 2), "minutos\n")
```

```
resultados_estimacion %>% summarise(n_distinct(ID))/k
```

```
copernicus_bruto %>%  
  distinct(Latitud,Longitud,Fechas,ID) %>%  
  summarise(n())
```

```
copernicus_requerido = resultados_estimacion %>%  
  filter(!is.na(Salinidad))
```

```
info_copernicus_requerido = copernicus_requerido
```

```
##### 2DA PARTE - 2DA INTERPOLACION  
#####
```

```
library(dplyr)  
library(gstat)  
library(sp)  
library(FNN)  
library(tidyr)
```

```
# CONSUMIENDO INFO DE COPENICUS EN BRUTO, TAL CUAL COMO VIENE  
copernicus_bruto <- read_excel("C:/Users/israe/Downloads/COPERNICUS DATOS EN  
BRUTO.xlsx",
```

```
  col_types = c("numeric", "numeric", "numeric",  
                "date", "numeric"))
```

```

# CONSUMIENDO DATOS IN SITU
insitu = read_excel("C:/Users/israe/Desktop/datos/Tesis/dato_insitu.xlsx",
  col_types = c("text", "text", "date",
    "numeric", "numeric", "numeric",
    "numeric"))

colnames(insitu)[3] = "FECHA"

insitu = insitu %>%
  filter(TYPE == "Variabilidad Climática" & SALINITY > 30 & FECHA < as.Date("2025-05-
28"))

insitu = insitu %>%
  group_by(TYPE,STATION,FECHA,LATITUDE,LONGITUDE,DEPTH) %>%
  summarise(SALINITY = mean(SALINITY)) %>%
  mutate(ID = paste(FECHA,DEPTH))

info_total_pares = data.frame()

start_time <- Sys.time()

registros = insitu %>%
  ungroup() %>%
  head(1) %>%
  #sample_n(5) %>%
  select(ID) %>% pull

for (registro in registros){

  combinacion_insitu_particular = insitu %>%
    filter(ID == registro)

  df_particular = copernicus_requerido %>%
    mutate(ID = paste(Fechas,Profundidad)) %>%
    inner_join(combinacion_insitu_particular) %>%
    ungroup() %>%
    select(Latitud,Longitud,Salinidad)

  m <- df_particular %>% nrow()
  Z <- 1.645
  p <- 0.5
  e <- 0.1
  n0 <- (Z^2 * p * (1 - p)) / e^2
  n_muestral <- ceiling(n0 / (1 + (n0 - 1) / m))

  # 1. Selección de n_muestral observaciones aleatorias
  set.seed(123)
  test_points <- df_particular %>% sample_n(n_muestral)

  # Data de entrenamiento excluyendo los puntos de test
  train_data <- df_particular %>% anti_join(test_points)

```

```

# Función de error porcentual
error_pct <- function(real, estimado) {
  if (anyNA(c(real, estimado))) return(NA_real_)
  abs(real - estimado) / real * 100
}

# Función evaluación de un punto
evaluar_modelos_seguro <- function(punto, train) {
  lat <- punto$Latitud
  lon <- punto$Longitud
  real <- punto$Salinidad

  # Calcular distancias
  distancias <- tryCatch(
    sqrt((train$Latitud - lat)^2 + (train$Longitud - lon)^2),
    error = function(e) rep(NA, nrow(train))
  )

  k <- 4
  idx_nn <- tryCatch(which.min(distancias), error = function(e) NA_integer_)
  idx_k <- tryCatch(order(distancias)[1:k], error = function(e) rep(NA_integer_, k))

  pred_nn <- tryCatch(train$Salinidad[idx_nn], error = function(e) NA_real_)
  pred_knn_avg <- tryCatch(mean(train$Salinidad[idx_k]), error = function(e) NA_real_)
  pred_knn_med <- tryCatch(median(train$Salinidad[idx_k]), error = function(e) NA_real_)

  pred_lm <- tryCatch({
    modelo_lm <- lm(Salinidad ~ Latitud + Longitud, data = train)
    predict(modelo_lm, newdata = punto)
  }, error = function(e) NA_real_)

  pred_krig <- tryCatch({
    sp_train <- train
    coordinates(sp_train) <- ~Longitud + Latitud
    variograma <- variogram(Salinidad ~ 1, sp_train)
    modelo_vario <- fit.variogram(variograma, vgm("Sph"))
    punto_sp <- data.frame(Longitud = lon, Latitud = lat)
    coordinates(punto_sp) <- ~Longitud + Latitud
    krige(Salinidad ~ 1, sp_train, punto_sp, modelo_vario)$var1.pred
  }, error = function(e) NA_real_)

  pred_idw <- tryCatch({
    sp_train <- train
    coordinates(sp_train) <- ~Longitud + Latitud
    punto_sp <- data.frame(Longitud = lon, Latitud = lat)
    coordinates(punto_sp) <- ~Longitud + Latitud
    idw(Salinidad ~ 1, sp_train, punto_sp, idp = 2)$var1.pred
  }, error = function(e) NA_real_)

  # Calcular errores solo si hay estimaciones válidas
  error_pct <- function(real, estimado) {
    if (is.na(real) || is.na(estimado)) return(NA_real_)

```

```

    abs(real - estimado) / real * 100
  }

  tibble(
    Latitud = lat,
    Longitud = lon,
    Real = real,
    Estimado_Vecino = pred_nn,
    Error_Vecino = error_pct(real, pred_nn),
    Estimado_KNN_Prom = pred_knn_avg,
    Error_KNN_Prom = error_pct(real, pred_knn_avg),
    Estimado_KNN_Med = pred_knn_med,
    Error_KNN_Med = error_pct(real, pred_knn_med),
    Estimado_Regresion = pred_lm,
    Error_Regresion = error_pct(real, pred_lm),
    Estimado_Kriging = pred_krig,
    Error_Kriging = error_pct(real, pred_krig),
    Estimado_IDW = pred_idw,
    Error_IDW = error_pct(real, pred_idw)
  )
}

library(purrr)
resultados_detallados <- map_dfr(
  1:nrow(test_points),
  function(i) {
    evaluar_modelos_seguro(test_points[i, ], train_data) %>%
      mutate(ID = combinacion_insitu_particular$ID)
  }
)

errores_promedio_detallados <- resultados_detallados %>%
  summarise(
    Vecino = mean(Error_Vecino, na.rm = TRUE),
    KNN_Prom = mean(Error_KNN_Prom, na.rm = TRUE),
    KNN_Med = mean(Error_KNN_Med, na.rm = TRUE),
    Regresion = mean(Error_Regresion, na.rm = TRUE),
    Kriging = mean(Error_Kriging, na.rm = TRUE),
    IDW = mean(Error_IDW, na.rm = TRUE)
  ) %>%
  pivot_longer(everything(), names_to = "MODELO", values_to =
"ERROR_PROMEDIO") %>%
  arrange(ERROR_PROMEDIO)

# Elegimos el modelo con menor error
mejor_modelo <- errores_promedio_detallados %>% head(1)

# Extraer lat y lon del punto objetivo
punto_obj <- insitu %>%
  filter(ID == combinacion_insitu_particular$ID) %>%
  rename(Latitud = LATITUDE, Longitud = LONGITUDE)

```

```

lat <- punto_obj$Latitud
lon <- punto_obj$Longitud

# Spatial
train_sp <- train_data
coordinates(train_sp) <- ~Longitud + Latitud
punto_sp <- data.frame(Longitud = lon, Latitud = lat)
coordinates(punto_sp) <- ~Longitud + Latitud

# Aplicamos modelo
salinidad_estim <- switch(
  mejor_modelo$MODELO,
  "Vecino" = {
    d <- sqrt((train_data$Latitud - lat)^2 + (train_data$Longitud - lon)^2)
    train_data$Salinidad[which.min(d)]
  },
  "KNN_Prom" = {
    d <- sqrt((train_data$Latitud - lat)^2 + (train_data$Longitud - lon)^2)
    mean(train_data$Salinidad[order(d)[1:4]])
  },
  "KNN_Med" = {
    d <- sqrt((train_data$Latitud - lat)^2 + (train_data$Longitud - lon)^2)
    median(train_data$Salinidad[order(d)[1:4]])
  },
  "Regresion" = {
    modelo_lm <- lm(Salinidad ~ Latitud + Longitud, data = train_data)
    predict(modelo_lm, newdata = punto_obj)
  },
  "Kriging" = {
    variograma <- variogram(Salinidad ~ 1, train_sp)
    modelo_vario <- fit.variogram(variograma, vgm("Sph"))
    krige(Salinidad ~ 1, train_sp, punto_sp, modelo_vario)$var1.pred
  },
  "IDW" = {
    idw(Salinidad ~ 1, train_sp, punto_sp, idp = 2)$var1.pred
  }
)

# Resultado final
resultado_final <- tibble(
  ID = combinacion_insitu_particular$ID,
  MODELO_ELEGIDO = mejor_modelo$MODELO,
  ERROR = mejor_modelo$ERROR_PROMEDIO,
  SALINIDAD_ESTIMADA = salinidad_estim
)

info_total_pares = info_total_pares %>% rbind(
  resultado_final
)

}

end_time <- Sys.time()
duracion <- end_time - start_time
duracion

```

```

m <- df_particular %>% nrow()
Z <- 1.645
p <- 0.5
e <- 0.1
n0 <- (Z^2 * p * (1 - p)) / e^2
n_muestral <- ceiling(n0 / (1 + (n0 - 1) / m))

# 1. Selección de n_muestral observaciones aleatorias
set.seed(123)
test_points <- df_particular %>% sample_n(n_muestral)

# Data de entrenamiento excluyendo los puntos de test
train_data <- df_particular %>% anti_join(test_points)

# Función de error porcentual
error_pct <- function(real, estimado) {
  if (anyNA(c(real, estimado))) return(NA_real_)
  abs(real - estimado) / real * 100
}

# Función evaluación de un punto
evaluar_modelos_seguro <- function(punto, train) {
  lat <- punto$Latitud
  lon <- punto$Longitud
  real <- punto$Salinidad

  # Calcular distancias
  distancias <- tryCatch(
    sqrt((train$Latitud - lat)^2 + (train$Longitud - lon)^2),
    error = function(e) rep(NA, nrow(train))
  )

  k <- 4
  idx_nn <- tryCatch(which.min(distancias), error = function(e) NA_integer_)
  idx_k <- tryCatch(order(distancias)[1:k], error = function(e) rep(NA_integer_, k))

  pred_nn <- tryCatch(train$Salinidad[idx_nn], error = function(e) NA_real_)
  pred_knn_avg <- tryCatch(mean(train$Salinidad[idx_k]), error = function(e) NA_real_)
  pred_knn_med <- tryCatch(median(train$Salinidad[idx_k]), error = function(e) NA_real_)

  pred_lm <- tryCatch({
    modelo_lm <- lm(Salinidad ~ Latitud + Longitud, data = train)
    predict(modelo_lm, newdata = punto)
  }, error = function(e) NA_real_)

  pred_krig <- tryCatch({
    sp_train <- train
    coordinates(sp_train) <- ~Longitud + Latitud
    variograma <- variogram(Salinidad ~ 1, sp_train)

```

```

modelo_vario <- fit.variogram(variograma, vgm("Sph"))
punto_sp <- data.frame(Longitud = lon, Latitud = lat)
coordinates(punto_sp) <- ~Longitud + Latitud
krige(Salinidad ~ 1, sp_train, punto_sp, modelo_vario)$var1.pred
}, error = function(e) NA_real_)

```

```

pred_idw <- tryCatch({
  sp_train <- train
  coordinates(sp_train) <- ~Longitud + Latitud
  punto_sp <- data.frame(Longitud = lon, Latitud = lat)
  coordinates(punto_sp) <- ~Longitud + Latitud
  idw(Salinidad ~ 1, sp_train, punto_sp, idp = 2)$var1.pred
}, error = function(e) NA_real_)

```

```

# Calcular errores solo si hay estimaciones válidas
error_pct <- function(real, estimado) {
  if (is.na(real) || is.na(estimado)) return(NA_real_)
  abs(real - estimado) / real * 100
}

```

```

tibble(
  Latitud = lat,
  Longitud = lon,
  Real = real,
  Estimado_Vecino = pred_nn,
  Error_Vecino = error_pct(real, pred_nn),
  Estimado_KNN_Prom = pred_knn_avg,
  Error_KNN_Prom = error_pct(real, pred_knn_avg),
  Estimado_KNN_Med = pred_knn_med,
  Error_KNN_Med = error_pct(real, pred_knn_med),
  Estimado_Regresion = pred_lm,
  Error_Regresion = error_pct(real, pred_lm),
  Estimado_Kriging = pred_krig,
  Error_Kriging = error_pct(real, pred_krig),
  Estimado_IDW = pred_idw,
  Error_IDW = error_pct(real, pred_idw)
)
}

```

```

library(purrr)
resultados_detallados <- map_dfr(
  1:nrow(test_points),
  function(i) {
    evaluar_modelos_seguro(test_points[i, ], train_data) %>%
      mutate(ID = combinacion_insitu_particular$ID)
  }
)

```

```

errores_promedio_detallados <- resultados_detallados %>%
  summarise(
    Vecino = mean(Error_Vecino, na.rm = TRUE),
    KNN_Prom = mean(Error_KNN_Prom, na.rm = TRUE),
    KNN_Med = mean(Error_KNN_Med, na.rm = TRUE),
    Regresion = mean(Error_Regresion, na.rm = TRUE),
  )

```

```

    Kriging = mean(Error_Kriging, na.rm = TRUE),
    IDW = mean(Error_IDW, na.rm = TRUE)
  ) %>%
  pivot_longer(everything(), names_to = "MODELO", values_to =
"ERROR_PROMEDIO") %>%
  arrange(ERROR_PROMEDIO)

# Elegimos el modelo con menor error
mejor_modelo <- errores_promedio_detallados %>% head(1)

# Extraer lat y lon del punto objetivo
punto_obj <- insitu %>%
  filter(ID == combinacion_insitu_particular$ID) %>%
  rename(Latitud = LATITUDE, Longitud = LONGITUDE)

lat <- punto_obj$Latitud
lon <- punto_obj$Longitud

# Spatial
train_sp <- train_data
coordinates(train_sp) <- ~Longitud + Latitud
punto_sp <- data.frame(Longitud = lon, Latitud = lat)
coordinates(punto_sp) <- ~Longitud + Latitud

# Aplicamos modelo
salinidad_estim <- switch(
  mejor_modelo$MODELO,
  "Vecino" = {
    d <- sqrt((train_data$Latitud - lat)^2 + (train_data$Longitud - lon)^2)
    train_data$Salinidad[which.min(d)]
  },
  "KNN_Prom" = {
    d <- sqrt((train_data$Latitud - lat)^2 + (train_data$Longitud - lon)^2)
    mean(train_data$Salinidad[order(d)[1:4]])
  },
  "KNN_Med" = {
    d <- sqrt((train_data$Latitud - lat)^2 + (train_data$Longitud - lon)^2)
    median(train_data$Salinidad[order(d)[1:4]])
  },
  "Regresion" = {
    modelo_lm <- lm(Salinidad ~ Latitud + Longitud, data = train_data)
    predict(modelo_lm, newdata = punto_obj)
  },
  "Kriging" = {
    variograma <- variogram(Salinidad ~ 1, train_sp)
    modelo_vario <- fit.variogram(variograma, vgm("Sph"))
    krige(Salinidad ~ 1, train_sp, punto_sp, modelo_vario)$var1.pred
  },
  "IDW" = {
    idw(Salinidad ~ 1, train_sp, punto_sp, idp = 2)$var1.pred
  }
)

```



```

# Resultado final
resultado_final <- tibble(
  ID = combinacion_insitu_particular$ID,
  MODELO_ELEGIDO = mejor_modelo$MODELO,
  ERROR = mejor_modelo$ERROR_PROMEDIO,
  SALINIDAD_ESTIMADA = salinidad_estim
)

info_total_pares = info_total_pares %>% rbind(
  resultado_final
)

end_time <- Sys.time()
duracion <- end_time - start_time
duracion

```

```

##### 3RA PARTE -
COMPARACION COPERNICUS VS INSITU
#####

```

```

pares_copernicus <- info_total_pares

```

```

pares_copernicus %>%
  group_by(MODELO_ELEGIDO) %>%
  summarise(OBS = n(), "ERROR PORCENTUAL PROMEDIO" = mean(ERROR)) %>%
  mutate("%OBS" = OBS/sum(OBS)*100) %>%
  rename(METODO = MODELO_ELEGIDO) %>%
  select(1,2,4,3)

```

```

insitu = read_excel("C:/Users/israe/Desktop/datos/Tesis/dato_insitu.xlsx",
  col_types = c("text", "text", "date",
    "numeric", "numeric", "numeric",
    "numeric")) %>%
  filter(TYPE == "Variabilidad Climática" & SALINITY > 30)

```

```

colnames(insitu)[3] = "FECHA"

```

```

insitu = insitu %>%
  group_by(TYPE,STATION,FECHA,LATITUDE,LONGITUDE,DEPTH) %>%
  summarise(SALINITY = mean(SALINITY)) %>%
  filter(SALINITY > 30) %>%
  filter(FECHA < as.Date("2025-05-28"))

```

```

df_comparacion = pares_copernicus %>%
  inner_join(
    insitu %>%
      mutate(ID = paste(FECHA,DEPTH))
  ) %>%

```

```

  rename(COPERNICUS = SALINIDAD_ESTIMADA, INSITU = SALINITY, LATITUD =
LATITUDE, LONGITUD = LONGITUDE ) %>%
  select(TYPE,STATION,FECHA,LATITUD,LONGITUD,DEPTH,INSITU,COPERNICUS)

```

```

df_errores <- df_comparacion %>%
  filter(INSITU > 30) %>%
  mutate(
    ABS_ERROR = abs(INSITU - COPERNICUS),
    REL_ERROR = abs(INSITU - COPERNICUS) / INSITU * 100,
    SQ_ERROR = (INSITU - COPERNICUS)^2
  ) %>%
  mutate(TEMPORADA = if_else(month(FECHA) %in% c(12,1,2,3,4,5) , "HÚMEDA" ,
"SECA"))

```

```

df_errores <- df_errores %>% mutate(ID = row_number())

```

```

##### COMPARACION CON PROMEDIOS POR TIPO DE
PROFUNDIDAD #####

```

```

datos_grafico <- df_comparacion %>%
  mutate(
    TIPO_PROFUNDIDAD = if_else(DEPTH <= 30, "Superficial", "Subsuperficial"),
    MES = month(FECHA)
  ) %>%
  group_by(STATION, MES, TIPO_PROFUNDIDAD) %>%
  summarise(
    COPERNICUS = round(mean(COPERNICUS),4),
    INSITU = round(mean(INSITU),4),
    .groups = "drop"
  )

```

```

# Lista única de combinaciones
combos <- datos_grafico %>%
  distinct(STATION, TIPO_PROFUNDIDAD)

```

```

graficos <- purrr::pmap(
  combos,
  function(STATION, TIPO_PROFUNDIDAD) {
    datos_filtrados <- datos_grafico %>%
      filter(STATION == !!STATION, TIPO_PROFUNDIDAD == !!TIPO_PROFUNDIDAD) %>%
      #right_join(tibble(MES = 1:12), by = "MES") %>%
      arrange(MES)

```

```

    highchart() %>%
      hc_chart(type = "bar") %>%
      hc_title(text = paste("Comparación Mensual -", STATION, "-",
TIPO_PROFUNDIDAD)) %>%
      hc_xAxis(categories = datos_filtrados$MES, title = list(text = "Mes")) %>%
      hc_yAxis(title = list(text = "Salinidad promedio")) %>%
      hc_tooltip(shared = TRUE, valueDecimals = 4, valueSuffix = "", pointFormat = "<span
style='color:{series.color}'>{series.name}</span>: <b>{point.y:.2f}</b><br/>") %>%
      hc_plotOptions(bar = list(
        dataLabels = list(
          enabled = TRUE,

```

```

    format = "{point.y:.4f}" # ← Esto asegura que salgan dos decimales
  )
  )) %>%
  hc_add_series(name = "COPERNICUS", data = datos_filtrados$COPERNICUS, color =
"#1f77b4") %>%
  hc_add_series(name = "INSITU", data = datos_filtrados$INSITU, color = "#e74c3c")
}
)

```

```

graficos[[1]] # Primer gráfico
graficos[[2]] # Segundo gráfico
graficos[[3]] # Tercer gráfico
graficos[[4]] # Cuarto gráfico

```

```

##### VISUAL PARA SABER HISTOGRAMA DE COPERNICUS
#####

```

```

copernicus_bruto_filtrado = copernicus_bruto %>%
  filter(Salinidad > 31)

```

```

n <- length(copernicus_bruto_filtrado$Salinidad)
breaks_sturges <- ceiling(log2(n) + 1)

```

```

hist_data <- hist(copernicus_bruto_filtrado$Salinidad, plot = FALSE, breaks =
breaks_sturges)

```

```

df_hist <- data.frame(
  x = hist_data$mids,
  y = hist_data$counts
)

```

```

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Histograma de Salinidad - Datos Copernicus") %>%
  hc_xAxis(
    title = list(text = "Salinidad"),
    categories = round(df_hist$x, 2)
  ) %>%
  hc_yAxis(title = list(text = "Frecuencia")) %>%
  hc_add_series(name = "Frecuencia", data = df_hist$y, color = "#7FB3D5") %>%
  hc_tooltip(pointFormat = "Salinidad: {point.category}<br>Frecuencia: {point.y}") %>%
  hc_plotOptions(column = list(
    borderWidth = 1,
    borderColor = "#2C3E50",
    pointPadding = 0,
    groupPadding = 0
  )) %>%
  hc_legend(enabled = FALSE)

```

```

##### VISUAL PARA SABER HISTOGRAMA DE COPERNICUS
PARA PROFUNDIDAD < 30 #####

```

```

copernicus_bruto_filtrado = copernicus_requerido %>%
  filter(Profundidad < 30)

n <- length(copernicus_bruto_filtrado$Salinidad)
breaks_sturges <- ceiling(log2(n) + 1)

hist_data <- hist(copernicus_bruto_filtrado$Salinidad, plot = FALSE, breaks =
breaks_sturges)

df_hist <- data.frame(
  x = hist_data$mids,
  y = hist_data$counts
)

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Histograma de Salinidad - Datos Copernicus - Profundidad < 30m") %>%
  hc_xAxis(
    title = list(text = "Salinidad"),
    categories = round(df_hist$x, 2)
  ) %>%
  hc_yAxis(title = list(text = "Frecuencia")) %>%
  hc_add_series(name = "Frecuencia", data = df_hist$y, color = "#7FB3D5") %>%
  hc_tooltip(pointFormat = "Salinidad: {point.category}<br>Frecuencia: {point.y}") %>%
  hc_plotOptions(column = list(
    borderWidth = 1,
    borderColor = "#2C3E50",
    pointPadding = 0,
    groupPadding = 0
  )) %>%
  hc_legend(enabled = FALSE)

##### VISUAL PARA SABER HISTOGRAMA DE COPERNICUS
PARA PROFUNDIDAD > 30 #####

copernicus_bruto_filtrado = copernicus_requerido %>%
  filter(Profundidad > 30)

n <- length(copernicus_bruto_filtrado$Salinidad)
breaks_sturges <- ceiling(log2(n) + 1)

hist_data <- hist(copernicus_bruto_filtrado$Salinidad, plot = FALSE, breaks =
breaks_sturges)

df_hist <- data.frame(
  x = hist_data$mids,
  y = hist_data$counts
)

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Histograma de Salinidad - Datos Copernicus - Profundidad > 30m") %>%
  hc_xAxis(
    title = list(text = "Salinidad"),

```

```

    categories = round(df_hist$x, 2)
  ) %>%
  hc_yAxis(title = list(text = "Frecuencia")) %>%
  hc_add_series(name = "Frecuencia", data = df_hist$y, color = "green") %>%
  hc_tooltip(pointFormat = "Salinidad: {point.category}<br>Frecuencia: {point.y}") %>%
  hc_plotOptions(column = list(
    borderWidth = 1,
    borderColor = "#2C3E50",
    pointPadding = 0,
    groupPadding = 0
  )) %>%
  hc_legend(enabled = FALSE)

```

VISUAL POR REGISTRO/OBSERVACION CUAL ES EL DATO DE COPERNICUS VS INSITU

GRAFICO DE LINEAS COMPARANDO IN SITU VS COPERNICUS

```

df_errores %>%
  mutate(ID = row_number()) %>%
  hchart("line", hcaes(x = ID, y = INSITU), name = "In Situ") %>%
  hc_add_series(df_errores$COPERNICUS, name = "Copernicus") %>%
  hc_title(text = "Comparación por Observación de Salinidad In Situ vs Copernicus") %>%
  hc_xAxis(title = list(text = "Registro")) %>%
  hc_yAxis(title = list(text = "Salinidad"))

```

GRAFICO DE LINEAS COMPARANDO IN SITU VS COPERNICUS - COLORES MAS FUERTES - LEYENDA CLARA

```

highchart() %>%
  hc_add_series(
    data = df_errores,
    type = "line",
    hcaes(x = ID, y = INSITU),
    name = "In Situ",
    color = "#1f77b4"
  ) %>%
  hc_add_series(
    data = df_errores,
    type = "line",
    hcaes(x = ID, y = COPERNICUS),
    name = "Copernicus",
    color = "#e74c3c"
  ) %>%
  hc_title(text = "Comparación por Observación de Salinidad In Situ vs Copernicus") %>%
  hc_xAxis(title = list(text = "Observación")) %>%
  hc_yAxis(title = list(text = "Salinidad")) %>%
  hc_legend(enabled = TRUE) %>%
  hc_tooltip(shared = TRUE, valueDecimals = 3)

```

GRAFICO DE PUNTOS COMPARANDO IN SITU VS COPERNICUS - COLORES MAS FUERTES - LEYENDA CLARA

```

highchart() %>%
  hc_add_series(
    data = df_errores,

```

```

    type = "scatter",
    hcaes(x = ID, y = INSITU),
    name = "In Situ",
    color = "#1f77b4",
    marker = list(symbol = "circle", radius = 5)
) %>%
hc_add_series(
  data = df_errores,
  type = "scatter",
  hcaes(x = ID, y = COPENICUS),
  name = "Copernicus",
  color = "#e74c3c",
  marker = list(symbol = "triangle", radius = 5)
) %>%
hc_title(text = "Comparación por Observación de Salinidad In Situ vs Copernicus") %>%
hc_xAxis(title = list(text = "Registro")) %>%
hc_yAxis(title = list(text = "Salinidad")) %>%
hc_legend(enabled = TRUE) %>%
hc_tooltip(shared = TRUE, valueDecimals = 3)

```

VISUAL POR REGISTRO/OBSERVACION CUAL ES EL DATO DE COPENICUS VS INSITU - PARA 3M

GRAFICO DE LINEAS COMPARANDO IN SITU VS COPENICUS - COLORES MAS FUERTES - LEYENDA CLARA

```

df_filtrado <- df_errores %>%
  filter(DEPTH == 3) %>%
  mutate(FECHA = as.Date(FECHA))

highchart() %>%
  hc_add_series(
    data = df_filtrado,
    type = "point",
    hcaes(x = FECHA, y = INSITU),
    name = "In Situ",
    color = "#1f77b4"
) %>%
  hc_add_series(
    data = df_filtrado,
    type = "point",
    hcaes(x = FECHA, y = COPENICUS),
    name = "Copernicus",
    color = "#e74c3c"
) %>%
  hc_title(text = "Comparación temporal de Salinidad In Situ vs Copernicus (Profundidad 3m)") %>%
  hc_xAxis(
    type = "datetime",
    title = list(text = "Fecha"),
    dateTimeLabelFormats = list(day = '%e %b %Y')
) %>%
  hc_yAxis(title = list(text = "Salinidad")) %>%

```

```
hc_legend(enabled = TRUE) %>%
hc_tooltip(shared = TRUE, valueDecimals = 3,
  xDateFormat = "%e %b %Y")
```

```
##### VISUAL RESUMEN GENERAL DE
INDICADORES DE ERROR #####
```

```
resumen_global <- df_errores %>%
  mutate(
    ABS_ERROR = abs(INSITU - COPERNICUS),
    REL_ERROR = abs(INSITU - COPERNICUS) / INSITU * 100,
    SQ_ERROR = (INSITU - COPERNICUS)^2
  ) %>%
  summarise(
    MAE = mean(ABS_ERROR),
    MAPE = mean(REL_ERROR, na.rm = TRUE),
    RMSE = sqrt(mean(SQ_ERROR))
  ) %>%
  select(-MAPE)
```

```
print(resumen_global)
```

```
resumen_largo <- resumen_global %>%
  pivot_longer(cols = everything(), names_to = "MÉTRICA", values_to = "VALOR") %>%
  mutate(MÉTRICA = ifelse(MÉTRICA == "MAPE", "MAPE (%)", MÉTRICA),
    MÉTRICA = ifelse(MÉTRICA != "MAPE", paste(MÉTRICA, "[psu]"), MÉTRICA))
```

```
resumen_largo
```

```
# Crear gráfico de columnas
highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Indicadores de Error: In Situ vs Copernicus") %>%
  hc_xAxis(categories = resumen_largo$MÉTRICA,
    title = list(text = "Métrica")) %>%
  hc_yAxis(title = list(text = "Valor del Error")) %>%
  hc_add_series(
    name = "Error",
    data = resumen_largo$VALOR,
    colorByPoint = TRUE,
    showInLegend = FALSE
  ) %>%
  hc_tooltip(pointFormat = "<b>{point.y:.3f}</b>") %>%
  hc_plotOptions(column = list(
    dataLabels = list(enabled = TRUE, format = '{point.y:.3f}')
  ))
```

```
###SOLO MAPE
```

```
resumen_global <- df_errores %>%
  mutate(REL_ERROR = abs(INSITU - COPERNICUS) / INSITU * 100) %>%
  summarise(`MAPE (%)` = mean(REL_ERROR, na.rm = TRUE))
```

```
metricas <- colnames(resumen_global)
valores <- resumen_global[[1]]
```

```
highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Indicadores de Error: In Situ vs Copernicus") %>%
  hc_xAxis(categories = "XD", title = list(text = "MAPE %")) %>%
  hc_yAxis(title = list(text = "Valor del Error (%))) %>%
  hc_add_series(
    name = "Error",
    data = valores,
    colorByPoint = TRUE,
    showInLegend = FALSE
  ) %>%
  hc_tooltip(pointFormat = "<b>{point.y:.3f}%</b>") %>%
  hc_plotOptions(column = list(
    dataLabels = list(enabled = TRUE, format = '{point.y:.3f}%')
  ))
```

```
valores <- df_errores %>%
  summarise(MAPE = mean(abs(INSITU - COPERNICUS) / INSITU * 100, na.rm =
TRUE)) %>%
  mutate(name = "MAPE (%)", y = MAPE) %>%
  select(name, y) %>%
  list_parse2()
```

```
highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Indicadores de Error: In Situ vs Copernicus") %>%
  hc_xAxis(type = "category", title = list(text = "Métrica")) %>%
  hc_yAxis(title = list(text = "Valor del Error (%))) %>%
  hc_add_series(
    name = "Error",
    data = valores,
    colorByPoint = TRUE,
    showInLegend = FALSE
  ) %>%
  hc_tooltip(pointFormat = "<b>{point.y:.3f}%</b>") %>%
  hc_plotOptions(column = list(
    dataLabels = list(enabled = TRUE, format = '{point.y:.3f}%')
  ))
```

VISUAL RESUMEN POR TEMPORADA DE INDICADORES DE ERROR

```
resumen_por_temporada <- df_errores %>%
  mutate(
```



```

    ABS_ERROR = abs(INSITU - COPENICUS),
    REL_ERROR = abs(INSITU - COPENICUS) / INSITU * 100,
    SQ_ERROR = (INSITU - COPENICUS)^2
  ) %>%
group_by(TEMPORADA) %>%
summarise(
  MAE = mean(ABS_ERROR),
  MAPE = mean(REL_ERROR, na.rm = TRUE),
  RMSE = sqrt(mean(SQ_ERROR)),
  .groups = "drop"
)

resumen_largo <- resumen_por_temporada %>%
pivot_longer(cols = -TEMPORADA, names_to = "MÉTRICA", values_to = "VALOR") %>%
mutate(MÉTRICA = ifelse(MÉTRICA == "MAPE", "MAPE (%)", MÉTRICA),
      VALOR = round(VALOR,4))

metricas <- unique(resumen_largo$MÉTRICA)

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Errores por Temporada") %>%
  hc_xAxis(categories = metricas, title = list(text = "Métrica")) %>%
  hc_yAxis(title = list(text = "Valor del Error")) %>%
  hc_tooltip(shared = TRUE, valueDecimals = 3) %>%
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE))) %>%
  hc_add_series(
    name = "HÚMEDA",
    data = resumen_largo %>% filter(TEMPORADA == "HÚMEDA") %>% pull(VALOR),
    color = "#1f77b4"
  ) %>%
  hc_add_series(
    name = "SECA",
    data = resumen_largo %>% filter(TEMPORADA == "SECA") %>% pull(VALOR),
    color = "#e74c3c"
  )

resumen_sin_mape <- resumen_largo %>%
  filter(MÉTRICA != "MAPE (%)")

metricas_sin_mape <- unique(resumen_sin_mape$MÉTRICA)

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Errores por Temporada") %>%
  hc_xAxis(categories = metricas_sin_mape, title = list(text = "Métrica")) %>%
  hc_yAxis(title = list(text = "Valor del Error")) %>%
  hc_tooltip(shared = TRUE, valueDecimals = 3) %>%
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE))) %>%
  hc_add_series(
    name = "HÚMEDA",
    data = resumen_sin_mape %>% filter(TEMPORADA == "HÚMEDA") %>% pull(VALOR),
    color = "#1f77b4"
  ) %>%

```

```

hc_add_series(
  name = "SECA",
  data = resumen_sin_mape %>% filter(TEMPORADA == "SECA") %>% pull(VALOR),
  color = "#e74c3c"
)

resumen_solo_mape <- resumen_largo %>%
  filter(MÉTRICA == "MAPE (%)")

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Error MAPE por Temporada") %>%
  hc_xAxis(categories = resumen_solo_mape$TEMPORADA, title = list(text =
"Temporada")) %>%
  hc_yAxis(title = list(text = "MAPE (%)")) %>%
  hc_tooltip(pointFormat = "<b>{point.y:.3f}%</b>") %>%
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE, format =
"{point.y:.3f}%"))) %>%
  hc_add_series(
    name = "MAPE (%)",
    data = resumen_solo_mape %>% mutate(name = TEMPORADA, y = VALOR) %>%
select(name, y) %>% list_parse2(),
    colorByPoint = TRUE,
    showInLegend = FALSE
  )

```

VISUAL RESUMEN POR PROFUNDIDAD DE INDICADORES DE ERROR

```

resumen_por_depth <- df_errores %>%
  mutate(
    ABS_ERROR = abs(INSITU - COPENNICUS),
    REL_ERROR = abs(INSITU - COPENNICUS) / INSITU * 100,
    SQ_ERROR = (INSITU - COPENNICUS)^2
  ) %>%
  group_by(DEPTH) %>%
  summarise(
    MAE = mean(ABS_ERROR),
    MAPE = mean(REL_ERROR, na.rm = TRUE),
    RMSE = sqrt(mean(SQ_ERROR)),
    .groups = "drop"
  )

resumen_largo <- resumen_por_depth %>%
  pivot_longer(cols = -DEPTH, names_to = "MÉTRICA", values_to = "VALOR") %>%
  mutate(MÉTRICA = ifelse(MÉTRICA == "MAPE", "MAPE (%)", MÉTRICA),
    VALOR = round(VALOR, 4))

metricas <- unique(resumen_largo$MÉTRICA)
profundidades <- unique(resumen_largo$DEPTH)

```

```

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Errores por Profundidad") %>%
  hc_xAxis(categories = metricas, title = list(text = "Métrica")) %>%
  hc_yAxis(title = list(text = "Valor del Error")) %>%
  hc_tooltip(shared = TRUE, valueDecimals = 3) %>%
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE))) %>%
  purrr::reduce(profundidades, function(hc, d) {
    hc %>%
      hc_add_series(
        name = paste0("DEPTH ", d),
        data = resumen_largo %>% filter(DEPTH == d) %>% pull(VALOR)
      )
  }, .init = .)

```

```

resumen_solo_mape <- resumen_largo %>% filter(MÉTRICA == "MAPE (%)")

```

```

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "MAPE por Profundidad") %>%
  hc_xAxis(categories = as.character(resumen_solo_mape$DEPTH), title = list(text =
"Profundidad")) %>%
  hc_yAxis(title = list(text = "MAPE (%)")) %>%
  hc_tooltip(pointFormat = "<b>{point.y:.3f}%</b>") %>%
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE, format =
"{point.y:.3f}%")) %>%
  hc_add_series(
    name = "MAPE (%)",
    data = resumen_solo_mape %>% mutate(name = as.character(DEPTH), y =
VALOR) %>% select(name, y) %>% list_parse2(),
    colorByPoint = TRUE,
    showInLegend = FALSE
  )

```

```

resumen_sin_mape <- resumen_largo %>%
  filter(MÉTRICA != "MAPE (%)")

```

```

metricas_sin_mape <- unique(resumen_sin_mape$MÉTRICA)
profundidades <- unique(resumen_sin_mape$DEPTH)

```

```

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Errores MAE y RMSE por Profundidad") %>%
  hc_xAxis(categories = metricas_sin_mape, title = list(text = "Métrica")) %>%
  hc_yAxis(title = list(text = "Valor del Error")) %>%
  hc_tooltip(shared = TRUE, valueDecimals = 3) %>%
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE))) %>%
  purrr::reduce(profundidades, function(hc, d) {
    hc %>%
      hc_add_series(
        name = paste0("DEPTH ", d),
        data = resumen_sin_mape %>% filter(DEPTH == d) %>% pull(VALOR)
      )
  }, .init = .)

```

```
}, .init = .)
```

```
##### VISUAL RESUMEN POR ESTACION DE  
INDICADORES DE ERROR #####
```

```
resumen_por_estacion <- df_errores %>%  
  mutate(  
    ABS_ERROR = abs(INSITU - COPERNICUS),  
    REL_ERROR = abs(INSITU - COPERNICUS) / INSITU * 100,  
    SQ_ERROR = (INSITU - COPERNICUS)^2  
  ) %>%  
  group_by(STATION) %>%  
  summarise(  
    MAE = mean(ABS_ERROR),  
    MAPE = mean(REL_ERROR, na.rm = TRUE),  
    RMSE = sqrt(mean(SQ_ERROR)),  
    .groups = "drop"  
  )  
  
resumen_largo_estacion <- resumen_por_estacion %>%  
  pivot_longer(cols = -STATION, names_to = "MÉTRICA", values_to = "VALOR") %>%  
  mutate(MÉTRICA = ifelse(MÉTRICA == "MAPE", "MAPE (%)", MÉTRICA),  
    VALOR = round(VALOR,4))  
  
metricas_estacion <- unique(resumen_largo_estacion$MÉTRICA)  
  
highchart() %>%  
  hc_chart(type = "column") %>%  
  hc_title(text = "Errores por Estación") %>%  
  hc_xAxis(categories = metricas_estacion, title = list(text = "Métrica")) %>%  
  hc_yAxis(title = list(text = "Valor del Error")) %>%  
  hc_tooltip(shared = TRUE, valueDecimals = 3) %>%  
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE))) %>%  
  hc_add_series(  
    name = "Puerto López 10MI",  
    data = resumen_largo_estacion %>% filter(STATION == "Puerto López 10MI") %>%  
    pull(VALOR),  
    color = "#1f77b4"  
  ) %>%  
  hc_add_series(  
    name = "Salinas 10MI",  
    data = resumen_largo_estacion %>% filter(STATION == "Salinas 10MI") %>%  
    pull(VALOR),  
    color = "#e74c3c"  
  )  
  
resumen_sin_mape_estacion <- resumen_largo_estacion %>%  
  filter(MÉTRICA != "MAPE (%)")  
  
metricas_sin_mape_estacion <- unique(resumen_sin_mape_estacion$MÉTRICA)
```

```

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Errores por Estación") %>%
  hc_xAxis(categories = metricas_sin_mape_estacion, title = list(text = "Métrica")) %>%
  hc_yAxis(title = list(text = "Valor del Error")) %>%
  hc_tooltip(shared = TRUE, valueDecimals = 3) %>%
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE))) %>%
  hc_add_series(
    name = "Puerto López 10MI",
    data = resumen_sin_mape_estacion %>% filter(STATION == "Puerto López 10MI") %>%
pull(VLOR),
    color = "#1f77b4"
  ) %>%
  hc_add_series(
    name = "Salinas 10MI",
    data = resumen_sin_mape_estacion %>% filter(STATION == "Salinas 10MI") %>%
pull(VLOR),
    color = "#e74c3c"
  )

```

```

resumen_solo_mape_estacion <- resumen_largo_estacion %>%
  filter(MÉTRICA == "MAPE (%)")

```

```

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Error MAPE por Estación") %>%
  hc_xAxis(categories = resumen_solo_mape_estacion$STATION, title = list(text =
"Estación")) %>%
  hc_yAxis(title = list(text = "MAPE (%)")) %>%
  hc_tooltip(pointFormat = "<b>{point.y:.3f}%</b>") %>%
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE, format =
"{point.y:.3f}%")) %>%
  hc_add_series(
    name = "MAPE (%)",
    data = resumen_solo_mape_estacion %>% mutate(name = STATION, y = VLOR) %>%
select(name, y) %>% list_parse2(),
    colorByPoint = TRUE,
    showInLegend = FALSE
  )

```

```

##### SOBRESTIMA O SUBESTIMA ?
#####

```

```

comparacion_tipo_bias = df_errores %>%
  mutate(TIPO_BIAS = if_else(COPERNICUS > INSITU, "ALTO", "BAJO")) %>%
  group_by(TIPO_BIAS) %>%
  summarise(OBS = n(),
    MAE = mean(ABS_ERROR),
    MAPE = mean(REL_ERROR, na.rm = TRUE),

```

```

      RMSE = sqrt(mean(SQ_ERROR)),
      .groups = "drop") %>%
mutate("%OBS" = OBS/sum(OBS)) %>%
select(1,6)

comparacion_largo_bias <- comparacion_tipo_bias %>%
  pivot_longer(cols = -TIPO_BIAS, names_to = "MÉTRICA", values_to = "VALOR") %>%
  mutate(MÉTRICA = ifelse(MÉTRICA == "MAPE", "MAPE (%)", MÉTRICA),
         VALOR = round(VALOR,4))

metricas_bias <- unique(comparacion_largo_bias$MÉTRICA)

highchart() %>%
  hc_chart(type = "column") %>%
  hc_title(text = "Comparación de %Observaciones por Tipo de Sesgo (Bias)") %>%
  hc_xAxis(categories = metricas_bias, title = list(text = "Métrica")) %>%
  hc_yAxis(title = list(text = "Porcentaje")) %>%
  hc_tooltip(shared = TRUE, valueDecimals = 3) %>%
  hc_plotOptions(column = list(dataLabels = list(enabled = TRUE))) %>%
  hc_add_series(
    name = "Copernicus > In Situ (ALTO)",
    data = comparacion_largo_bias %>% filter(TIPO_BIAS == "ALTO") %>%
pull(VALOR),
    color = "#2E86C1"
  ) %>%
  hc_add_series(
    name = "Copernicus < In Situ (BAJO)",
    data = comparacion_largo_bias %>% filter(TIPO_BIAS == "BAJO") %>%
pull(VALOR),
    color = "#C0392B"
  )

```