



Universidade de Vigo

Trabajo Fin de Máster

---

# Exploración de métodos para estimar la incertidumbre de los descartes pesqueros utilizando el formato RDBES

---

Joel Marcelo Morales Chávez

Máster en Técnicas Estadísticas

Curso 2024-2025



## Propuesta de Trabajo Fin de Máster

<p><b>Título en galego:</b> Exploración de métodos para estimar a incerteza dos descartes de pesca mediante o formato RDBES</p>
<p><b>Título en español:</b> Exploración de métodos para estimar la incertidumbre de los descartes pesqueros utilizando el formato RDBES</p>
<p><b>English title:</b> Exploration of methods to estimate the uncertainty of fishing discards using RDBES format</p>
<p><b>Modalidad:</b> Modalidad B</p>
<p><b>Autor/a:</b> Joel Marcelo Morales Chávez, Universidade de Santiago de Compostela</p>
<p><b>Director/a:</b> Manuel Alfredo Mosquera Rodríguez, Universidade de Vigo</p>
<p><b>Tutor/a:</b> José Antonio Castro Pampillón, Instituto Español de Oceanografía (IEO-CSIC); Adriana Nogueira Gassent, Instituto Español de Oceanografía (IEO-CSIC)</p>
<p><b>Breve resumen del trabajo:</b></p> <p>La estimación del descarte pesquero es importante para cuantificar la mortalidad de las poblaciones de peces requerida en los modelos de dinámica de poblaciones sobre cuyos resultados se sostienen las medidas de gestión. Dicha estimación se realiza sobre datos recogidos por los programas científicos de muestreo a bordo, de modo que debe ir acompañado de su incertidumbre asociada.</p>
<p><b>Recomendaciones:</b></p> <p>Conocimientos de estadística de muestreo y programación en R.</p>
<p><b>Otras observaciones:</b></p> <p>Propuesta propiciada por el alumno, Joel Morales y los tutores, José Castro y Adriana Nogueira.</p>



Don Manuel Alfredo Mosquera Rodríguez, Profesor titular de la Universidade de Vigo, don José Antonio Castro Pampillón, Científico Titular de Instituto Español de Oceanografía (IEO-CSIC), y doña Adriana Nogueira Gassent, Investigadora M3 de Instituto Español de Oceanografía (IEO-CSIC), informan que el Trabajo Fin de Máster titulado

**Exploración de métodos para estimar la incertidumbre de los descartes pesqueros  
utilizando el formato RDBES**

fue realizado bajo su dirección por don Joel Marcelo Morales Chávez para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal. Además, Don Manuel Alfredo Mosquera Rodríguez y don Joel Marcelo Morales Chávez

sí                       no

autorizan a la publicación de la memoria en el repositorio de acceso público asociado al Máster en Técnicas Estadísticas.

En Santiago de Compostela, a 03 de junio de 2025.

El director:  
Don Manuel Alfredo Mosquera Rodríguez

MOSQUERA  
RODRIGUEZ MANUEL  
ALFREDO -  
78738609H                      **Fecha:**  
**2025.06.03**  
**13:15:23 +02'00'**

El autor:  
Don Joel Marcelo Morales Chávez



El tutor:  
Don José Antonio Castro Pampillón

Firmado por CASTRO  
PAMPILLON JOSE  
ANTONIO - DNI

La tutora:  
Doña Adriana Nogueira Gassent

**NOGUEIRA**  
**GASSENT**  
**ADRIANA - DNI**  
**14307098V**

Firmado digitalmente por NOGUEIRA GASSENT  
ADRIANA - DNI 14307098V  
Nombre de reconocimiento (DN): c=ES,  
o=CONSEJO SUPERIOR DE INVESTIGACIONES  
CIENTIFICAS, ou=CERTIFICADO ELECTRONICO DE  
EMPLEADO PUBLICO, ou=CENTRO  
OCEANOGRAFICO DE VIGO, ou=14307098V,  
serialNumber=IDCES-14307098V, sn=NOGUEIRA  
GASSENT, givenName=ADRIANA, cn=NOGUEIRA  
GASSENT ADRIANA - DNI 14307098V  
Fecha: 2025.06.03 10:51:30 +02'00'

**Declaración responsable.** Para dar cumplimiento a la Ley 3/2022, de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster (Artículo 11, [Disposición 2978 del BOE núm. 48 de 2022](#)), **el autor declara** que el Trabajo Fin de Máster presentado es un documento original en el que se han tenido en cuenta las siguientes consideraciones relativas al uso de material de apoyo desarrollado por otros/as autores/as:

- Todas las fuentes usadas para la elaboración de este trabajo han sido citadas convenientemente (libros, artículos, apuntes de profesorado, páginas web, programas, . . . )
- Cualquier contenido copiado o traducido textualmente se ha puesto entre comillas, citando su procedencia.
- Se ha hecho constar explícitamente cuando un capítulo, sección, demostración, . . . sea una adaptación casi literal de alguna fuente existente.

Y, acepta que, si se demostrara lo contrario, se le apliquen las medidas disciplinarias que correspondan.



# Agradecimientos

*Quiero expresar mi más profundo agradecimiento a Dios y a quienes han sido mis grandes pilares a lo largo de este camino.*

*A mis padres y hermano, por su amor incondicional, su apoyo constante y por enseñarme el valor del esfuerzo. Este TFM no habría sido posible sin ustedes.*

*A mi tutor académico, Manuel, y a mis tutores del IEO-CISC, José y Adriana, por su compromiso, dedicación y orientación en cada etapa de este trabajo.*

*A mis amigos de Perú, gracias por estar siempre presentes, a pesar de la distancia, con palabras de aliento y recuerdos que me acompañaron en cada paso.*

*A mis amigos en Santiago de Compostela, por los momentos compartidos y por hacer de esta ciudad un lugar lleno de recuerdos y aprendizaje.*

*A todos, muchas gracias por ser parte de este camino.*

*Me gustaría agradecer la acogida que tuvo mi trabajo por parte del grupo de expertos en metodología FDI del Scientific, Technical and Economic Committee for Fisheries (STECF), durante la plenaria del pasado 20 de mayo de 2025. Compartir mis avances con un equipo de gran trayectoria fue una muy buena experiencia.*

*La recogida de los datos utilizados para esta tesis ha sido cofinanciada por la Unión Europea a través del Fondo Europeo Marítimo, de Pesca y de Acuicultura (FEMPA) dentro del Programa Nacional de recopilación, gestión y uso de datos del sector pesquero en apoyo al asesoramiento científico en relación con la Política Pesquera Común.*



**Cofinanciado por  
la Unión Europea**



# Índice general

<b>Resumen</b>	<b>XI</b>
<b>Introducción</b>	<b>XIII</b>
0.1. Gestión pesquera	XIII
0.2. Importancia del descarte en la UE	XIV
0.3. Landing Obligation	XIV
0.4. Stock	XV
0.5. Métièrs	XV
0.6. Planteamiento del problema	XVI
0.7. Objetivos del trabajo	XVI
<b>1. Marco teórico</b>	<b>1</b>
1.1. Regional Database and Estimation System (RDBES)	1
1.2. Diseños de muestreo en programas de observación pesquera	2
1.2.1. Reducción a un diseño de muestreo de dos etapas	3
1.3. Estimación del descarte total observado	3
<b>2. Material y métodos</b>	<b>5</b>
2.1. Estudio de caso	5
2.2. Metodología	6
2.2.1. Subpoblación	7
2.2.2. Método basado en diseño de 2 etapas	7
2.2.3. Método bootstrap	10
2.2.4. Algoritmos bootstrap	10
2.3. Herramientas estadísticas y software utilizado	12
<b>3. Resultados</b>	<b>13</b>
3.1. Estimación del descarte por método	13
3.1.1. Caso 1: En función de mareas $N$	13
3.1.2. Caso 2: En función lances $N = \frac{M_0}{M}$	14
3.2. Comparación de varianzas e intervalos de confianza	15
3.2.1. Caso 1: En función de mareas $N$	15
3.2.2. Caso 2: En función lances $N = \frac{M_0}{M}$	19
3.3. Comparación de metodologías	21
<b>4. Discusión</b>	<b>23</b>
4.1. Comparación con el método oficial actual del IEO-CSIC	23
4.2. Implicaciones para la implementación del RDBES	23
4.3. Limitaciones del estudio	24
4.4. Recomendaciones para el futuro	24

<b>5. Conclusiones</b>	<b>27</b>
<b>Bibliografía</b>	<b>29</b>
<b>Anexo</b>	<b>31</b>

# Resumen

La estimación precisa de los descartes es esencial para una gestión eficaz de las poblaciones pesqueras, ya que los datos de descartes forman parte de la mortalidad por pesca. No hacerlo podría comprometer seriamente la sostenibilidad a largo plazo de los recursos pesqueros. Su precisión se ve afectada por diversos factores, desde el diseño del muestreo hasta el proceso de estimación. Expertos en muestreo han expresado su preocupación por la escasez de indicadores de calidad de los datos en las estimaciones de descartes, especialmente cuando se deben utilizar datos de métodos de muestreo no probabilísticos como sustitutos de datos probabilísticos.

Como estudio de caso, utilizamos datos del programa de observadores a bordo del IEO-CSIC de la flota de arrastre española que captura peces demersales en la zona 7 del ICES. El muestreo sigue un diseño en varias etapas, desde la selección de los buques hasta la recogida de la muestra de descartes durante la maniobra de pesca. Para obtener las estimaciones más precisas de los descartes, es necesario especificar correctamente los métodos sucesivos de selección de las unidades de muestreo, así como parametrizar adecuadamente el cálculo de la varianza en cada etapa del muestreo. Para ello, se empleó un diseño de muestreo en dos etapas. En la primera etapa, las mareas fueron seleccionadas mediante muestreo aleatorio simple sin reemplazo (SRSWOR), mientras que en la segunda etapa, las operaciones de pesca (lances) se eligieron utilizando muestreo sistemático (SYSS). Para la estimación de la varianza, se adoptaron dos enfoques complementarios. En primer lugar, se utilizó la estimación clásica correspondiente al SRSWOR y, en el caso del SYSS, se compararon las aproximaciones obtenidas mediante el método de diferencias sucesivas y la estimación basada en el supuesto de SRSWOR. En segundo lugar, se implementaron técnicas de remuestreo, particularmente el método Bootstrap, con el fin de evaluar la incertidumbre de manera más flexible y robusta.

# Abstract

Accurate estimation of discards is essential for effective fish stock management, as discard data are part of the fishing mortality. Failure to do so could seriously compromise the long-term sustainability of fish stocks. However, the accuracy of discard estimates is affected by various factors, from sampling design of the sampling programs to the estimation process. Sampling experts have expressed concern about the paucity of data quality indicators in discard estimates, particularly when data from non-probability survey methods need to be used as proxies for probability data.

As a case study, we use data from observers on board the Spanish trawl fleet targeting demersal fish in ICES subarea 7. Sampling follows a multistage design, from vessel selection to the collection of the sample of discarded catch during each haul. To obtain the most accurate discard estimates, it is necessary to properly specify the successive methods for selecting sampling units, as well as to correctly parameterize the variance calculation at each sampling step. For this purpose, a two-stage sampling design was employed. In the first stage, fishing trips were selected using simple random sampling without replacement (SRSWOR), while in the second stage, fishing operations (sets) were selected using systematic sampling (SYSS). Two complementary approaches were adopted to estimate the variance. First, the classical SRSWOR estimate was used, and for SYSS, the approximations obtained using the method of successive differences and the estimate based on the SRSWOR assumption were compared. Second, resampling techniques, particularly the bootstrap method, were implemented to assess uncertainty in a more flexible and robust manner.

# Introducción

## 0.1. Gestión pesquera

La pesca es clave no solo como una fuente vital de alimento nutritivo para muchas personas en el mundo, sino también como un pilar económico que sostiene a numerosas familias y comunidades costeras. Las poblaciones de peces, siendo un recurso natural renovable, pueden mantenerse saludables a lo largo del tiempo siempre que se manejen con criterios de sostenibilidad. Sin embargo, esta capacidad de regeneración puede verse amenazada cuando la presión de la pesca se vuelve excesiva y desequilibra el ecosistema.

Durante años se creyó que los océanos ofrecían recursos inagotables, lo que llevó a una explotación sin límites, especialmente en aguas internacionales. Pero con los avances tecnológicos en las técnicas de pesca y localización de especies, se comprobó que estos recursos son finitos y pueden agotarse si no se gestionan adecuadamente. Esto motivó la creación de normativas y sistemas de manejo que regulan la actividad pesquera para proteger los ecosistemas marinos y asegurar la continuidad de la pesca en el futuro.

La sobrepesca constituye actualmente una de las amenazas más serias para la salud de los océanos. Para abordar el problema eficazmente, es necesario aplicar una combinación de estrategias como establecer límites de captura basados en evaluaciones científicas, la regulación de las artes de pesca y el esfuerzo pesquero, y la protección de zonas clave para la reproducción y el crecimiento de especies.

Como señala FAO (2022a), "La ordenación eficaz de toda la pesca constituye un objetivo esencial de la transformación azul. La mejora de la ordenación pesquera resulta esencial para recuperar las poblaciones de peces, incrementar las capturas y restaurar los ecosistemas para que su estado sea saludable y productivo, realizando al mismo tiempo una ordenación de los recursos explotados dentro de los límites de los ecosistemas. Esto requiere cambios transformadores para promover reformas de la gobernanza y las políticas, marcos de ordenación eficaces, tecnologías innovadoras y una protección social suficiente." (p. XXII).

En este contexto, es importante señalar que la Unión Europea (UE) regula la explotación de sus recursos pesqueros a través de los Totales Admisibles de Captura (TAC), que se distribuyen posteriormente entre los Estados miembros en forma de cuotas nacionales. Para garantizar que estas medidas de gestión sean efectivas, es esencial contar con datos fiables sobre el total de las capturas realizadas, incluidas tanto las que se desembarcan como las que se descartan.

Los descartes —es decir, la parte de la captura que se devuelve al mar— pueden representar una fracción considerable del total extraído alterando los ecosistemas marinos. Una estimación precisa de los descartes es clave dentro del sistema de gestión basado en TAC ya que, de lo contrario, los TAC se basarían únicamente en la fracción desembarcada, ignorando una parte significativa de las capturas reales, comprometiendo las evaluaciones y las decisiones de gestión.

## 0.2. Importancia del descarte en la UE

Los descartes se devuelven al mar ya sean vivos o muertos por ser especies sin valor comercial, por no superar la talla mínima permitida, por haber alcanzado la cuota asignada por una especie o por tratarse de especies prohibidas. En primer lugar los descartes alteran los ecosistemas marinos, alterando las cadenas tróficas, y un impacto negativo en la diversidad. Por otro lado cuando se habla de gestionar de forma sostenible los recursos marinos, el descarte pesquero juega un papel esencial. Este proceso ayuda a tener una imagen completa de la mortalidad causada por la pesca, más allá de lo que se registra en los desembarques.

Los descartes pesqueros deben registrarse obligatoriamente en los diarios de pesca. No obstante, la calidad limitada de estos datos sigue haciendo imprescindible su estimación científica. Para abordar esta necesidad, la Unión Europea (UE) lleva más de dos décadas impulsando programas nacionales de observadores a bordo. Estos programas recopilan información biológica detallada sobre la porción de capturas que normalmente es devuelta al mar pero el muestreo sólo cubre un porcentaje del total de la flota. Por lo tanto, se realiza una estimación. Al ser una estimación, su habilidad puede verse afectada por distintos elementos, como la forma en que se recoge la información o los métodos utilizados para analizarla.

Como señala Borges (2009), "Los descartes han adquirido mayor importancia a la vista del público en Europa con la creciente conciencia pública sobre la conservación de los océanos, con la intensificación de la sobreexplotación de los recursos pesqueros y, recientemente, con el reconocimiento público por parte de la industria pesquera del descarte generalizado de especies comerciales."



Figura 1: Captura de peces durante una maniobra de pesca (fuente: <https://www.oceansoffun.org/overfishing>)

## 0.3. Landing Obligation

La Unión Europea ha abordado el problema de los descartes con la reforma de la Política Pesquera Común (PPC), en 2013. Este reglamento base de la gestión de los recursos pesqueros europeos, establece en su Artículo 15 la obligación del desembarque de los descartes, con el objetivo de mejorar la cuantificación del descarte y avanzar hasta la eliminación de esta práctica, fomentando una explotación más sostenible de los recursos marinos. Esta legislación pretende reducir la mortalidad no

documentada y fomentar una pesca más selectiva y sostenible. Sin embargo, su implementación ha enfrentado algunos problemas como limitaciones técnicas o por conflictos socioeconómicos. Estudios recientes (FAO 2019) destacan que, diversos elementos relacionados con la implementación del sistema de obligación de desembarque deben revisarse de forma constante para asegurar tanto su eficacia como la sostenibilidad de las pesquerías en la UE bajo este nuevo enfoque.

De esta forma, cabe destacar que una estimación precisa de los descartes no solo contribuye a optimizar la gestión de los recursos pesqueros, ajustando los límites de capturas a la realidad biológica sino que también permite evaluar la eficacia de medidas clave en la legislación pesquera europea, como la Landing Obligation. Sin datos fiables de descartes, resulta difícil medir el grado de cumplimiento y efectividad de esta normativa, así como su impacto en las prácticas pesqueras y en la conservación de los recursos marinos.

## 0.4. Stock

Un stock pesquero es un grupo de peces de la misma especie que conforman una población con características propias, que se gestiona de manera independiente, desde el punto de vista ecológico, biológico y económico. Estas poblaciones presentan sus propias dinámicas, como el crecimiento, la reproducción, la mortalidad y el reclutamiento (nuevos individuos que ingresan en una población explotada), y están influenciadas por factores ambientales comunes y prácticas pesqueras en esa área de distribución. Por lo tanto, un stock se define como una unidad de gestión en la que los procesos demográficos internos (reclutamiento, el crecimiento y la mortalidad), tienen mayor influencia que los intercambios con otras poblaciones cercanas FAO (2022b).

La regulación de descartes afecta directamente a la biomasa de los stocks, ya que obliga a contabilizar todas las capturas (incluidas las no comerciales), en los modelos y en las evaluaciones científicas. De esta forma, se mejora la precisión de las estimaciones de estado del stock y se refuerza la transparencia en gestión pesquera.

## 0.5. Métiers

Un métier es una categoría específica dentro de una pesquería o flota, de definida por el tipo de embarcación o por el tipo de marea o viaje realizado.

Este concepto fue introducido formalmente en la legislación de la UE en 2008, principalmente para mejorar la planificación del muestreo, la gestión pesquera y la recolección de datos en el marco del Reglamento (CE) N<sup>o</sup> 199/2008 (UE 2008) sobre el establecimiento de un marco comunitario para la recogida, gestión y uso de datos en el sector pesquero. Desde entonces se ha venido usando como estrato estadístico del muestreo biológico (Acosta-Yepes et al. 2019).

Un métier agruparía las operaciones de pesca que comparten:

- Arte o técnica de pesca
- Área geográfica
- Especie objetivo
- Periodo del año

En los programas de recopilación de datos pesqueros (Data Collection Framework), lo métiers se utilizan como estratos de muestreo.

## 0.6. Planteamiento del problema

Contar con estimaciones fiables de los descartes por especie es fundamental para evaluar correctamente el estado de las pesquerías y tomar decisiones de gestión basadas en evidencia. En el contexto español, el muestreo de descartes se realiza de forma diferenciada: mientras que el Centro de Investigación Marina y Alimentaria (AZTI) se encarga exclusivamente de la flota vasca, el Instituto Español de Oceanografía (IEO-CSIC) cubre el resto de la flota española, que representa la mayor parte de la actividad pesquera nacional. Ambas instituciones aplican metodologías establecidas para generar estas estimaciones. Sin embargo, el proceso de estimación no está exento de dificultades, especialmente en lo que respecta a la evaluación de la incertidumbre asociada a los datos, como la varianza o la precisión de los resultados.

Los programas de muestreo utilizados suelen estar basados en diseños multietápicos complejos, lo que implica distintos niveles de variabilidad. A pesar de esto, las fórmulas habitualmente empleadas para calcular la varianza o los indicadores de calidad no siempre reflejan adecuadamente esta complejidad.

En los últimos años, varios informes técnicos han puesto de manifiesto la necesidad de mejorar los métodos actuales. En particular, el Comité Científico, Técnico y Económico de Pesca (con siglas en inglés, STECF) ha señalado la ausencia de directrices claras y métodos validados para calcular e interpretar indicadores de calidad como los coeficientes de variación o los intervalos de confianza. Tal como se señala en su informe (STECF 2023), los indicadores de calidad solicitados el año pasado requieren más pruebas y análisis antes de ser adoptados como estándar. Por lo tanto, actualmente, los indicadores se proporcionen de forma voluntaria y no existan normas oficiales de validación para su presentación. El Grupo de Trabajo de Expertos (EWG) destacó la falta de una base teórica sólida en torno a las estadísticas de calidad y mencionó la necesidad de avanzar en la aplicación de la teoría de muestreo y de la metodología bootstrap en este contexto.

Esta falta de estandarización y claridad metodológica plantea la necesidad de revisar el enfoque actual y compararlo con otras alternativas basadas en diseños de muestreo multietápico y técnicas estadísticas más flexibles, como el bootstrap. Estos métodos podrían ofrecer estimaciones más precisas y realistas de la variabilidad en los datos, lo cual es clave para mejorar la calidad de la información que se utiliza en la gestión de los recursos pesqueros. Por tanto, en este trabajo se propone analizar y comparar distintos enfoques metodológicos, con el fin de identificar cuáles ofrecen mejores resultados y podrían integrarse en futuras prácticas oficiales de estimación de descartes.

Adicionalmente, se prevé que en los próximos años entre en vigor un nuevo formato de base de datos, desarrollado por Regional Advisory Committes e implementado por la Unión Europea para coordinación científica técnica de los programas nacionales de muestreo pesquero. Este sistema, es la Regional Database and Estimation System (RDBES) –se explicará en capítulo siguiente–, que sustituirá progresivamente al actual. Este nuevo esquema propone cambios estructurales importantes en la forma en que se recogen, procesan y reportan los datos biológicos de muestreo. Ante este escenario de transición, resulta más relevante garantizar que las metodologías de estimación sean sólidas, adaptables y compatibles con los nuevos requerimientos técnicos y normativos.

## 0.7. Objetivos del trabajo

El objetivo principal de este trabajo es evaluar la estimación del descarte por especie y la varianza asociada, aplicando distintos enfoques metodológicos. En particular, se utilizarán fórmulas clásicas derivadas del diseño de muestreo multietápico, así como técnicas de remuestreo mediante el método bootstrap. Estos enfoques se compararán con el procedimiento actualmente empleado por el Instituto Español de Oceanografía (IEO-CSIC), organismo responsable de la elaboración oficial de estas estima-

ciones del descarte de la flota española no vasca. A través de esta comparación, se busca determinar cuál de los métodos ofrece una mayor precisión y robustez en la estimación de descartes, con el objetivo de identificar posibles mejoras en los protocolos existentes. Asimismo, el análisis pretende ofrecer recomendaciones metodológicas sólidas que contribuyan a optimizar la evaluación de descartes en el marco de la gestión sostenible de los recursos pesqueros.



# Capítulo 1

## Marco teórico

### 1.1. Regional Database and Estimation System (RDBES)

La base de datos RDBES es una herramienta desarrollada por la Unión Europea con el propósito de armonizar la recopilación de datos pesqueros entre los Estados Miembros. Su función principal es garantizar una mayor calidad, trazabilidad y accesibilidad de la información relacionada con las capturas, el esfuerzo pesquero y los muestreos biológicos. Gracias a su estructura jerárquica y adaptable, RDBES permite representar diversos esquemas de muestreo, abarcando desde mareas hasta los datos específicos de cada especie capturada. Este sistema sustituye al anterior (RDB) y está en proceso de consolidarse como el nuevo estándar para el reporte de datos en el marco del Reglamento de Datos de la Unión Europea.

Según el Consejo Internacional para la Exploración del Mar (ICES [2025](#)), los grupos de expertos en evaluación de poblaciones de peces del ICES sustituirán el sistema actual por el sistema RDBES para respaldar el asesoramiento sobre el estado de los recursos pesqueros tanto en países de la UE como fuera de ella, así como para definir y coordinar estrategias regionales de muestreo.

Este formato técnico, ya definitivo, se está implementando progresivamente. Los Estados miembros dispondrán de un periodo de adaptación hasta 2027, cuando su uso será obligatorio para toda transmisión de datos biológico-pesqueros en Europa.

La Figura 1.1 muestra la estructura del esquema del formato RDBES, diseñado para organizar datos de muestreo a bordo de forma detallada, representa una de las trece diferentes jerarquías disponibles en RDBES. La jerarquía comienza con el diseño del muestreo (Design - DE) y se despliega hacia niveles más específicos: detalles del muestreo (Sampling Details - SD), selección de embarcaciones (Vessel Selection - VS), viajes de pesca (Fishing Trip - FT), operaciones de pesca (Fishing Operation - FO), selección de especies (Species Selection - SS) y finalmente las muestras biológicas (Sample - SA). Estas 7 tablas son obligatorias, y están indicadas con un borde rojo. Cabe destacar que la tabla Sampling Details (SD), aunque también obligatoria, no pertenece a la jerarquía, y está representada solo con un borde rojo, a diferencia de las que pertenecen a la jerarquía, tienen un fondo azul. Adicionalmente, el esquema incluye referencias externas a tablas auxiliares como VD (datos de embarcaciones) y SL (lista de especies), que enriquecen la información sin alterar la estructura jerárquica. Todas las tablas están conectadas principalmente mediante relaciones de uno a muchos (1:n), lo que permite capturar múltiples niveles de detalle en el proceso de muestreo.

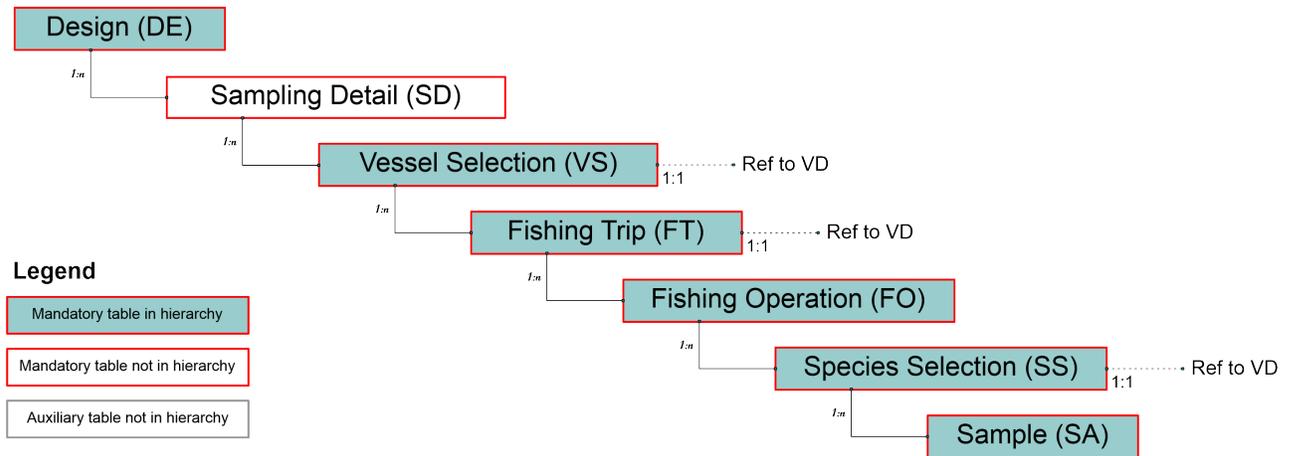


Figura 1.1: Esquema de formato RDBES para la matriz CS (commercial sampling) de un programa de muestreo a bordo. Adaptado de ICES (2024)

## 1.2. Diseños de muestreo en programas de observación pesquera

El diseño del muestreo seguido en el Programa Nacional de Observación a Bordo empleado por IEO-CSIC, en formato RDBES, responde a un esquema multietápico. Este busca garantizar la representatividad operativa y eficiencia logística en la recolección de datos de descartes. A continuación, se describen las etapas del diseño de muestreo aplicadas, en orden:

- Selección de buques (Vessel Selection):** Se empleó un enfoque de tipo *Non-Probabilistic Quasi Simple Random Sampling With Replacement (NPQRSWOR)*. Aunque no es estrictamente probabilístico, este método intenta aproximarse a una selección aleatoria simple con reemplazo, garantizando cierta heterogeneidad en la muestra de buques observados. El NPQRSWOR sustituye al SRSWOR en el muestreo de la flota de altura del IEO debido a las dificultades logísticas de coordinar barcos aleatoriamente en este tipo de pesca. En lugar de una selección puramente aleatoria, este diseño incorpora la participación de las cofradías para facilitar la selección de embarcaciones, permitiendo cierta flexibilidad operativa. Este método resulta más viable en contextos donde la disponibilidad de los barcos es limitada y su coordinación depende de la colaboración del sector pesquero.
- Selección de mareas (Fishing Trip):** Dentro de cada buque seleccionado, se aplicó un esquema de muestreo *aleatorio simple sin reemplazo (SRSWOR)* entre las mareas disponibles, con el objetivo de evitar repeticiones y cubrir distintos momentos de la actividad pesquera.
- Selección de operaciones de pesca o lances (Fishing Operation):** Durante cada marea, se utilizó un muestreo *sistemático (SYSS)* en el cual se seleccionaron dos de cada tres lances efectuados. Esta estrategia permite mantener una cobertura continua del esfuerzo pesquero con una carga de trabajo asumible para el observador.
- Selección de especies descartadas (Species Selection):** Se realizó un *censo* completo de todas las especies descartadas, registrando de forma exhaustiva la composición específica de los descartes en cada lance muestreado.
- Selección de individuos para medición biométrica (Sample Selection):** Para los individuos descartados, se aplicó un muestreo *aleatorio simple sin reemplazo (SRSWOR)* con el fin de seleccionar ejemplares a los que se les realizaron mediciones biométricas (longitud, peso, etc).

El enfoque de muestreo multietápico permite abarcar de manera efectiva las distintas fases del proceso pesquero, distribuir equitativamente las tareas entre los observadores y facilitar la implementación de técnicas estadísticas coherentes con la estructura del diseño adoptado.

### 1.2.1. Reducción a un diseño de muestreo de dos etapas

La simplificación del diseño muestral a un esquema de dos etapas se justifica principalmente por la baja influencia que ciertas etapas intermedias ejercen sobre la variabilidad total del estimador. En primer lugar, dado que el método NPQRSWOR prioriza los barcos disponibles en puerto, prácticamente toda la flota resulta seleccionada al final del año, lo que implica que no hay variabilidad significativa en la inclusión de unidades muestrales. Por esta razón, según los expertos del formato RDBES, la etapa de buques en este diseño es solo un mecanismo logístico para acceder a las mareas, y puede obviarse en el análisis sin afectar la representatividad de los datos.

En segundo lugar, la cuarta etapa (la selección de especies desarcadas) se basa en un censo, por lo que no introduce variabilidad muestral. Finalmente, aunque el artículo de referencia (Tamsett et al. 1999) ofrece orientaciones útiles para simplificar una quinta etapa, estas no son directamente aplicables en nuestro contexto, ya que el diseño implementado contempla únicamente una muestra. En conjunto, estas consideraciones permiten reducir el esquema a dos etapas sin comprometer la validez de las estimaciones ni subestimar su variabilidad.

Así, el diseño final conserva únicamente dos niveles de muestreo:

1. Selección de mareas (Fishing Trip Selection) como primera etapa.
2. Selección de operaciones de pesca o lances (Fishing Operation) como segunda etapa.



Figura 1.2: Buque comercial de la flota española de arrastre especializado en la captura de gallo en aguas atlánticas europeas no ibéricas.

## 1.3. Estimación del descarte total observado

Como no es posible observar cada operación de pesca, ya sea por limitaciones de coste o de tiempo, se recopilan datos en una parte representativa de dichas operaciones. Esto se logra observando una

selección de embarcaciones y registrando cuantos peces se descartan en los lances muestreados, que constituye nuestra muestra.

Para estimar el descarte total, primero se analizan los datos recogidos en los lances de pesca durante cada marea. Estos valores se ajustan según el esfuerzo pesquero total de la flota. A continuación, los resultados se extrapolan al conjunto del dominio espacio-temporal (la combinación de trimestre y división ICES), tomando como referencia el número total de mareas realizadas por el métier en esa zona y periodo correspondiente.

## Capítulo 2

# Material y métodos

### 2.1. Estudio de caso

El presente análisis se basa en datos obtenidos por medio de un marco de recopilación de datos, la Data Collection Framework estructurado conforme al formato RDBES. La base de datos empleada comprende tanto información de muestreo primario como datos censales relacionados con la actividad pesquera, correspondientes al año 2022. La intensidad de muestreo fue del 3.21 %, correspondiente a 9 mareas muestreadas de un total de 280.

Los datos empleados se agrupan en dos matrices principales:

- CS (Commercial Sampling): Datos brutos de muestreo sobre la actividad pesquera comercial, los cuales incluyen información detallada sobre las capturas y descartes en unidades de muestreo individuales.
- CE (Commercial Effort): Datos censales sobre el esfuerzo pesquero. Estos datos proceden de los registros oficiales recopilados por el Ministerio de Agricultura, Pesca y Alimentación, que cubren la totalidad de la flota pesquera española.

El estudio se centra en el estrato IEO\_P1\_S\_S7\_OTB, correspondiente a la flota española de arrastre dirigida gallo en aguas atlánticas europeas no ibéricas. Este estrato fue seleccionado específicamente debido a que en él se realizan mareas de media y larga duración, lo que conlleva una mayor cantidad de lances por marea. Esta característica es particularmente valiosa para el análisis, ya que proporciona un volumen de datos más amplio y una representatividad estadística más robusta. Al contrario que la flota de litoral, en la cual se muestrea prácticamente la totalidad de los lances por marea, el estrato IEO\_P1\_S\_S7\_OTB ofrece una oportunidad adicional para evaluar el efecto del número de lances en la estimación de la varianza.

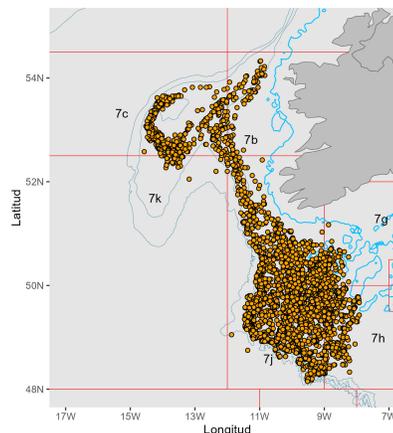


Figura 2.1: Mapa de la distribución del esfuerzo de la flota española de arrastre de zona 7 de ICES.

Este estrato, corresponde al metier de pesca de arrastre de fondo con puertas dirigido a especies de peces demersales en aguas irlandesas e inglesas, identificado como OTB\_DEF\_70-99\_0\_0, que realiza

operaciones de pesca de arrastre demersal (OTB) dirigidas a peces demersales (DEF), con redes de apertura de malla entre 70 y 99 mm. La especie objetivo del estudio es el gallo del norte, científicamente conocido como *Lepidorhombus whiffi agonis* (Walbaum 1792), una especie de elevado interés comercial para España.

Metier: OTB\_DEF\_70-99\_0\_0 que se define como:

- Arte: Arrastre de fondo
- Zona ICES: 7
- Especie Objetivo: Gallo (*Lepidorhombus whiffi agonis*)
- Época del año: todo el año

## 2.2. Metodología

Para asegurar una presentación clara y coherente de los métodos estadísticos utilizados en este estudio, a continuación se introducen las principales notaciones empleadas a lo largo del documento. Estas definiciones permiten expresar con precisión los procedimientos de estimación y ayudan a interpretar los resultados, especialmente considerando que se trabaja con un diseño de muestreo complejo y multietápico. Usar una notación consistente no solo facilita la lectura, sino que también es clave para garantizar la reproducibilidad del análisis y la correcta aplicación de los modelos estadísticos planteados.

Descripción	Símbolo	
	Población	Muestra
Número total de mareas.	$N$	$n$
Número total de lances por cada marea $i$ .	$M_i$	$m_i$
Descarte observado en un lance $j$ por cada marea $i$ .	$Y_{ij}$	$y_{ij}$
Media de descartes por lance en la marea $i$ .	$\bar{Y}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} Y_{ij}$	$\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$
Media de descartes por lance.	$\bar{\bar{Y}} = \frac{1}{N} \sum_{i=1}^N \bar{Y}_i$	$\bar{\bar{y}} = \frac{1}{n} \sum_{i=1}^n \bar{y}_i$
Total de descartes por cada marea $i$ .	$Y_i = M_i \bar{Y}_i$	$\hat{y}_i = M_i \bar{y}_i$
Media de descartes por cada marea $i$ .	$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$	$\bar{y} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i$
Estimación total de descarte observado.	$\hat{y} = N \bar{y} = \frac{N}{n} \sum_{i=1}^n \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij}$	

### 2.2.1. Subpoblación

Una subpoblación, también llamado dominio, es un grupo más pequeño dentro de una población más amplia que tiene características propias que las distinguen del resto. El estudio de subpoblaciones es importante en el enfoque de descarte, ya que permitiría desagregar en grupos el muestreo del estrato, lo que facilitaría identificar el descarte pesquero y evaluar el impacto de la pesca de manera más precisa en áreas geográficas específicas, así diseñar medidas de mitigación específicas.

#### Estimador del total en subpoblaciones

Para estimar el total de una subpoblación, se usa el estimador de Hansen y Hurwitz, donde tendríamos:

$$\hat{T}_D = \sum_{i \in D} \frac{y_i}{\pi_i}$$

Donde:

- $D$  : Conjunto de unidades muestreadas.
- $y_i$  : Valor en la unidad  $i$ .
- $\pi_i$  : Probabilidad de inclusión de la unidad  $i$ .

Este enfoque se desarrolla con más detalle en Thompson (2012) y Cochran (1977).

### 2.2.2. Método basado en diseño de 2 etapas

Este tipo de diseño se caracteriza por un proceso de selección estructurado en dos niveles jerárquicos. En una primera etapa, se eligen unidades grandes o agrupaciones, conocidas como unidades primarias (mareas). Posteriormente, en una segunda etapa, se seleccionan los elementos individuales de estudio dentro de esas unidades previamente elegidas, también llamados unidades secundarias (lances).

Lohr (2009) señala que “El muestreo en dos etapas puede ahorrar tiempo y dinero si la información auxiliar es relativamente barata de obtener y si contar con esa información auxiliar puede aumentar la precisión de las estimaciones para las cantidades de interés.”

#### Primera etapa

En esta etapa, la unidad de muestreo corresponde a las mareas. Se utiliza un muestreo aleatorio **simple sin reemplazo (SRSSWOR)** para seleccionar las **mareas** de la población total. Este método implica que cada marea tiene la misma probabilidad de ser seleccionada y que las unidades se extraen sin devolverlas a la población.

Para esta etapa, se aplican las fórmulas clásicas de muestreo aleatorio simple para el cálculo de estimadores y sus varianzas, lo que facilita la inferencia estadística sobre las unidades primarias seleccionadas.

#### Segunda etapa

En esta etapa, la unidad de muestreo corresponde a la operación de pesca, denominada **lance**. La selección de estas unidades se realiza mediante un **muestreo sistemático (SYSS)** dentro de cada **marea**.

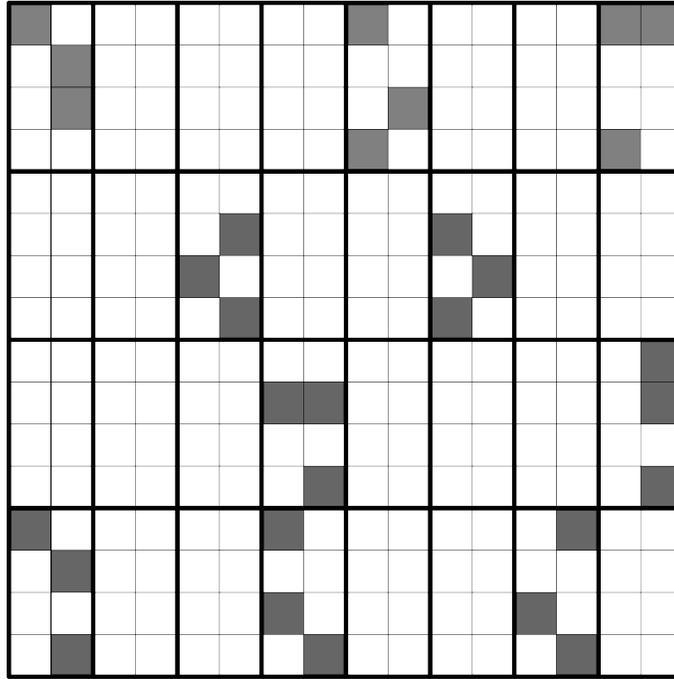


Figura 2.2: Muestra en dos etapas con 10 unidades primarias y 3 unidades secundarias por cada unidad primaria. Adaptado de (Thompson 2012)

Uno de los principales desafíos asociados al muestreo sistemático radica en la estimación de la varianza, debido a la dependencia estructural entre las unidades seleccionadas. Para la estimación de la varianza en esta etapa, se aplicarán dos métodos complementarios, según lo indicado por Arnab (2017):

- **Asumiendo muestreo aleatorio simple sin reemplazo (SRSWOR):** Consiste en la selección aleatoria de unidades sin reemplazo, aplicando las fórmulas clásicas para el cálculo de la varianza.
- **Aproximación mediante el método de las diferencias:** Este método considera la posible correlación entre los lances con el fin de mejorar la precisión de las estimaciones.

### Estimador del total

Las probabilidades de seleccionar una marea en la primera etapa se denotan como  $\pi_i = \frac{n}{N}$ , mientras que la probabilidad de seleccionar un lance  $j$  dentro del grupo  $i$  en la segunda etapa se denota como  $\pi_{j|i} = \frac{m_i}{M_i}$ . Por lo tanto, la probabilidad total de seleccionar el lance  $j$  en la marea  $i$  es el producto  $\pi_{ij} = \pi_i \times \pi_{j|i}$ .

Por lo tanto nuestra formula estaría dado por:

$$\hat{t} = \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{1}{\pi_{ij}} y_{ij} = \frac{N}{n} \sum_{i=1}^n \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij}$$

Este estimador se desarrolla en Lohr (2009).

### Formula de la varianza

Según Cochran (1977) y más recientemente Vigneau (2006) y Lohr (2009), la varianza del estimador del total en el diseño en 2 etapas sin reemplazo se calcula mediante la siguiente expresión:

$$V(\hat{y}) = \overbrace{\frac{N^2}{n} \left(1 - \frac{n}{N}\right) \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}^{\text{Representa la varianza entre mareas}} + \overbrace{\frac{N}{n} \sum_{i=1}^n \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) \frac{\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2}{m_i - 1}}^{\text{Representa la varianza dentro de cada marea}} \quad (2.1)$$

En la segunda etapa del muestreo bietápico, es posible mejorar la estimación de la varianza mediante el método de las diferencias. Este enfoque, propuesto por Arnab (2017) y adaptado a nuestro enfoque de diseño de muestra en dos etapas, considerando tanto la variabilidad entre las unidades primarias como la variabilidad dentro de cada unidad primaria.

$$V(\hat{y}) = \overbrace{\frac{N^2}{n} \left(1 - \frac{n}{N}\right) \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}^{\text{Representa la varianza entre mareas}} + \overbrace{\frac{N}{n} \sum_{i=1}^n \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) \frac{\sum_{j=1}^{m_i-1} (y_{ij+1} - y_{ij})^2}{2(m_i - 1)}}^{\text{Representa la varianza dentro de cada marea}} \quad (2.2)$$

En Cochran (1977) y Vigneau (2006) se propone una adaptación para el cálculo de la varianza en diseños de muestreo en dos etapas, en la cual se sustituye el valor de  $N$ , correspondiente al número total de unidades primarias, por la expresión  $\frac{M_0}{\bar{M}}$ , donde  $M_0$  representa el total de unidades secundarias en la población y  $\bar{M}$  es el promedio de unidades secundarias por unidad primaria en la muestra.

Esta sustitución permite expresar la varianza en función del tamaño efectivo de la población secundaria, lo que resulta útil cuando el interés principal recae en estimar totales o medias relacionados con las unidades secundarias.

En consecuencia, la formulación actualizada se expresa de la siguiente manera:

- Asumiendo muestreo aleatorio simple sin reemplazo

$$V(\hat{y}) = \overbrace{\frac{M_0^2}{n\bar{M}^2} \left(1 - \frac{n}{N}\right) \frac{\sum_{i=1}^n (\bar{y}_i - \bar{\bar{y}})^2}{n-1}}^{\text{Representa la varianza entre mareas}} + \overbrace{\frac{M_0}{n\bar{M}} \sum_{i=1}^n \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) \frac{\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2}{m_i - 1}}^{\text{Representa la varianza dentro de cada marea}} \quad (2.3)$$

- Aproximación mediante el método de las diferencias

$$V(\hat{y}) = \overbrace{\frac{M_0^2}{n\bar{M}^2} \left(1 - \frac{n}{N}\right) \frac{\sum_{i=1}^n (\bar{y}_i - \bar{\bar{y}})^2}{n-1}}^{\text{Representa la varianza entre mareas}} + \overbrace{\frac{M_0}{n\bar{M}} \sum_{i=1}^n \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) \frac{\sum_{j=1}^{m_i-1} (y_{ij+1} - y_{ij})^2}{2(m_i - 1)}}^{\text{Representa la varianza dentro de cada marea}} \quad (2.4)$$

### Intervalo de confianza

En el análisis estadístico, el intervalo de confianza es una herramienta clave para hacer inferencias sobre los parámetros de una población a partir de los datos obtenidos en una muestra. Su función principal es ofrecer un rango dentro del cual es razonable suponer que se encuentra el valor real del parámetro poblacional, aportando así un nivel de confianza que respalda la estimación realizada.

- **Para muestras grandes:** La distribución muestral del estimador del total se aproxima a una distribución normal debido al Teorema Central del Límite.

$$IC = \left( \hat{y} \pm z_{\alpha/2} \cdot \sqrt{V(\hat{y})} \right)$$

Donde  $z_{\alpha/2}$  es el valor crítico de la distribución normal estándar para el nivel de confianza deseado.

- **Para muestras pequeñas:** Se asume que la variable estudiada sigue una distribución aproximadamente normal en la población. En este caso, la distribución muestral del estimador del total se ajusta a una distribución  $t$  de Student con  $n - 1$  grados de libertad.

$$IC = \left( \hat{y} \pm t_{\alpha/2, n-1} \cdot \sqrt{V(\hat{y})} \right)$$

Donde  $t_{\alpha/2, n-1}$  representa el valor crítico de la distribución  $t$  para el nivel de confianza  $1 - \alpha$  y  $n - 1$  grados de libertad.

Este desarrollo se realiza siguiendo los fundamentos presentados en Cochran (1977) y Lohr (2009).

### 2.2.3. Método bootstrap

El método bootstrap es una técnica estadística flexible y poderosa que permite estimar la distribución de un estadístico a partir de los propios datos, utilizando un proceso de remuestreo con reemplazo. A diferencia de los métodos tradicionales, que suelen basarse en supuestos estrictos sobre la distribución de la población, el bootstrap se apoya directamente en la muestra disponible para generar múltiples muestras simuladas.

Esto lo convierte en una herramienta especialmente útil cuando no se cumplen las condiciones clásicas, como la normalidad de los datos o el uso de muestras grandes. En estos casos, donde los métodos convencionales pueden no ser fiables, el bootstrap ofrece una alternativa robusta para calcular intervalos de confianza, errores estándar y otros estadísticos clave, sin necesidad de conocer la forma exacta de la distribución real de los datos.

Al respecto, Efron y Tibshirani (1993) señalan que "Bootstrap puede utilizarse de forma no paramétrica o paramétrica. En el modo no paramétrico, evita suposiciones paramétricas restrictivas y, a veces, peligrosas, sobre la forma de las poblaciones subyacentes. En el modo paramétrico, puede proporcionar estimaciones de error más precisas que los métodos tradicionales basados en información de Fisher."(p. 394).

### 2.2.4. Algoritmos bootstrap

#### Bootstrap en segunda etapa

Este algoritmo mantiene fijas las mareas, realizando un remuestreo con reemplazo de los lances dentro de cada marea fija.

1. Identificar el conjunto de mareas en la muestra original. Denotamos este conjunto como  $FT_1, FT_2, \dots, FT_n$ .
2. Para cada réplica bootstrap  $b = 1, 2, \dots, B$ :
  - a) Inicializar una muestra bootstrap vacía  $\mathcal{D}_b$ .
  - b) Para cada marea  $FT_i$ , con  $i = 1, \dots, n$ :

- 1) Extraer los lances asociadas a  $FT_i$ , denotadas como  $FO_{i1}, FO_{i2}, \dots, FO_{im}$ .
- 2) Realizar un muestreo con reemplazo de estos lances para obtener una muestra bootstrap  $\{FO_{i1}^*, FO_{i2}^*, \dots, FO_{im}^*\}$ .
- 3) Agregar los lances remuestreados a  $\mathcal{D}_b$ .
- c) Calcular el estimador usando la muestra  $\mathcal{D}_b$ .

Este método estima la variabilidad intra-marea, pero no captura la variabilidad entre mareas, lo que puede llevar a una subestimación de la varianza total.

El algoritmo de **Bootstrap en primera etapa** y en **Bootstrap en segunda etapa** fue adaptado del algoritmo de Bootstrap Naive desarrollado por Rao y Wu (1988).

### Bootstrap en primera etapa

Este algoritmo realiza un remuestreo con reemplazo de las mareas, manteniendo fijos los lances dentro de cada marea seleccionada.

1. Identificar el conjunto de mareas en la muestra original. Denotamos este conjunto como  $FT_1, FT_2, \dots, FT_n$ .
2. Para cada réplica bootstrap  $b = 1, 2, \dots, B$ :
  - a) Inicializar una muestra bootstrap vacía  $\mathcal{D}_b$ .
  - b) Seleccionar con reemplazo  $n$  mareas del conjunto original. Denotamos el conjunto remuestreado como  $\{FT_1^*, FT_2^*, \dots, FT_n^*\}$ .
  - c) Para cada marea seleccionada  $FT_i^*$ , con  $i = 1, \dots, n$ :
    - 1) Extraer todas los lances asociados a  $FT_i^*$  desde la muestra original.
    - 2) Asignar un nuevo identificador a la marea.
    - 3) Agregar los lances a la muestra bootstrap  $\mathcal{D}_b$ .
  - d) Calcular el estimador total usando la muestra  $\mathcal{D}_b$ .

Este método captura la variabilidad entre las mareas, pero no remuestrea los lances dentro de cada marea. Por lo tanto, no refleja la variabilidad intra-marea.

### Bootstrap reescalado

Este algoritmo realiza un remuestreo sin reemplazo tanto las mareas y en los lances dentro de cada marea de la muestra original, con el fin de modificar los pesos originales en la muestra.

1. Identificar el conjunto de mareas en la muestra original. Denotamos este conjunto como  $FT_1, FT_2, \dots, FT_n$ .
2. Para cada réplica bootstrap  $b = 1, 2, \dots, B$ :
  - a) Seleccionar sin reemplazo  $n^* = \lfloor n/2 \rfloor$  mareas del conjunto original. Denotamos el conjunto remuestreado como  $FT_1^*, FT_2^*, \dots, FT_{n^*}^*$ .
  - b) Calcular los pesos de Bootstrap para la marea, sea  $\delta_{1i}$  igual a 1 si se selecciona la marea  $i = 1, \dots, n^*$  y 0 en caso contrario.

$$w_{1i}^* = w_{1i} \left( 1 - \lambda_1 + \lambda_1 \frac{n}{n^*} \delta_{1i} \right), \text{ Donde } \lambda_1 = \sqrt{n^* \frac{(1-f_1)}{n-n^*}}, f_1 = \frac{n}{N}, w_1 = \frac{N}{n}$$

- c) Para cada marea  $FT_i$ , con  $i = 1, \dots, n$ :

- 1) Extraer los lances asociadas a  $FT_i$ , denotadas como  $FO_{i1}, FO_{i2}, \dots, FO_{im}$ .
- 2) Realizar un muestreo sin reposición  $m^* = \lceil m/2 \rceil_{\text{parte entera}}$  lances para obtener una muestra bootstrap  $\{FO_{i1}^*, FO_{i2}^*, \dots, FO_{im^*}^*\}$ .
- 3) Calcular los pesos de Bootstrap para los lances, Sea  $\delta_{2i}$  igual a 1 si se selecciona el lance  $j$  en la marea  $i$  y 0 en caso contrario.

$$w_{2ij}^* = w_{2ij} \frac{w_{1i}}{w_{1i}^*} \left( 1 - \lambda_1 + \lambda_1 \frac{n}{n^*} \delta_{1i} - \lambda_{2i} \sqrt{\frac{n}{n^*}} \delta_{1i} + \lambda_{2i} \sqrt{\frac{n}{n^*}} \delta_{1i} \frac{m_i}{m_i^*} \delta_{2ij} \right)$$

$$\text{Donde } \lambda_{2i} = \sqrt{m_i^* f_1 \frac{(1-f_{2i})}{m_i - m_i^*}}, f_{2i} = \frac{m_i}{M_i}, w_{2i} = \frac{M_i}{m_i}$$

- d) Calcular el estimador total como  $\sum_{i=1}^n \sum_{j=1}^m w_{ij}^* y_{ij}$ , donde  $w_{ij}^* = w_{1i}^* w_{2ij}^*$

Este algoritmo fue desarrollado por Preston (2009).

### Intervalo de Confianza por Percentiles

El método de percentiles es una técnica bootstrap para construir intervalos de confianza directamente a partir de la distribución empírica del estimador obtenida por remuestreo. Dados  $B$  valores bootstrap  $\{\hat{\theta}_1^*, \dots, \hat{\theta}_B^*\}$ , se ordenan y se toman los percentiles  $\frac{\alpha}{2}$  y  $1 - \frac{\alpha}{2}$  para formar el intervalo:

$$IC_{1-\alpha}^{\text{percentil}} = \left[ \hat{\theta}_{(\frac{\alpha}{2})}^*, \hat{\theta}_{(1-\frac{\alpha}{2})}^* \right] \quad (2.5)$$

Efron y Tibshirani (1993) destacan su simplicidad y aplicabilidad en contextos donde los métodos clásicos resultan limitados.

## 2.3. Herramientas estadísticas y software utilizado

El procesamiento, análisis de datos y desarrollo de todas las estimaciones se realizaron utilizando el lenguaje y entorno estadístico R. La implementación de los procedimientos estadísticos se realizó mediante scripts personalizados desarrollados específicamente para el presente estudio, sin el uso de paquetes estadísticos especializados para muestreo como *survey* o *sampling*. Esto permitió un mayor control sobre la formulación de los estimadores, el cálculo de varianzas y la replicación detallada del diseño muestral en dos etapas.

El enfoque empleado incluyó programación estructurada en R, manipulación de datos mediante funciones del paquete *dplyr*, y validación de los resultados mediante comparaciones con estimaciones obtenidas utilizando la librería *survey*, lo que permitió verificar la correcta implementación de los procedimientos desarrollados manualmente.

## Capítulo 3

# Resultados

En este capítulo se presentan y analizan los resultados obtenidos tras aplicar los métodos propuestos para la estimación del total de descarte. El objetivo principal es evaluar la precisión y la estabilidad de cada método, así como comparar su desempeño en términos de varianza e los intervalos de confianza. Además, se muestran los resultados desglosados por dominios específicos definidos por subzonas de pesca, trimestres del año y metiers.

### 3.1. Estimación del descarte por método

#### 3.1.1. Caso 1: En función de mareas N

DOMINIO		N° MAREAS (N° LANCES)	ESTIMACIÓN (TONELADAS)			
TRIMESTRE	AREA		DISEÑO EN 2 ETAPAS	BOOTSTRAP		
				LANCES	MAREAS	REESCALADO
1	27.7.g	1 (1)	1.91	2.98	2.88	1.83
1	27.7.h	1 (10)	9.72	9.53	14.69	9.66
1	27.7.j	1 (6)	5.43	5.49	8.21	5.78
2	27.7.b	1 (1)	0.21	0.35	0.32	0.20
2	27.7.c	3 (71)	92.52	92.76	94.57	96.14
2	27.7.h	1 (7)	9.31	9.36	14.15	9.44
2	27.7.j	2 (17)	51.94	52.07	56.82	58.94
3	27.7.b	1 (2)	4.96	5.62	7.52	4.76
3	27.7.h	2 (52)	98.04	98.83	112.44	106.16
3	27.7.j	1 (29)	51.30	51.90	78.24	52.74
4	27.7.j	1 (6)	28.32	28.49	43.63	28.74
<b>TOTAL</b>			<b>353.66</b>	<b>357.39</b>	<b>433.53</b>	<b>374.29</b>

Cuadro 3.1: Estimaciones por trimestre, subáreas 7 y metier según el método de diseño y el método Bootstrap.

La estimación basada en los métodos de diseño de muestreo en dos etapas produce un único valor estimado para el total de descarte por dominio, determinado exclusivamente por las probabilidades de selección definidas en el diseño. En cambio, los métodos bootstrap generan una estimación empírica del total de descarte en cada réplica, ya que cada réplica se construye a partir de re-muestreos de la

muestra original, lo que da lugar a valores estimados distintos en cada iteración.

En la Tabla 3.1, se puede observar que las estimaciones basadas en el diseño en dos etapas tienden a ser más conservadoras, mientras que las obtenidas por métodos bootstrap, especialmente a nivel mareas, suelen ser mayores. Los métodos bootstrap reescalado y a nivel de lances, tienden a aproximarse más a las estimaciones del diseño original, por lo tanto, serían útiles cuando se busca mantener consistencia con el enfoque tradicional.

En zonas con altos niveles de descarte, como la subárea 27.7.c durante el segundo trimestre, las estimaciones obtenidas por los distintos métodos muestran una alta concordancia, con valores que varían entre 92.52 y 96.14. Esta coincidencia indica que, en situaciones de captura elevada, el método de estimación utilizado incluye poco en los resultados.

En cambio, en áreas con valores bajos, como la subárea 27.7.b en el mismo trimestre, las diferencias entre métodos son pequeñas en números absolutos (entre 0.21 y 0.35), pero más notables términos relativos. Esto implica que, aunque las cifras sean bajas, cualquier variación entre métodos puede influir en la interpretación.

### 3.1.2. Caso 2: En función lances $N = \frac{M_0}{\bar{M}}$

Según Vigneau (2006), es posible prescindir del conocimiento del parámetro poblacional  $N$  (número total de mareas) mediante la sustitución:  $N \approx \frac{M_0}{\bar{M}}$ .

Donde:

- $\bar{M}$  representa el número promedio de lances por marea.
- $M_0$  denota el total de lances en la población.

Cuando  $\bar{M}$  se conoce a nivel poblacional,  $N$  resulta directamente determinable, lo que hace redundante este enfoque. Sin embargo, en casos donde  $N$  no es observable, se puede estimar  $\bar{M}$  mediante muestreo, con  $\bar{M} \approx \frac{\sum_{i=1}^n M_i}{n}$ .

El valor de  $M_0$  fue proporcionado oficialmente por el Instituto Español de Oceanografía (IEO), obtenido mediante sus sistemas de monitorización estandarizados. Los datos fueron validados asegurando su confiabilidad para la estimación. Esto se debe a que los datos censales que proporcionados por la matriz CE no aportan el número total de lances realizados por la flota, ya que, aunque RDBES dispone de un campo para aportar ese dato, no es obligatorio sino opcional.

Los resultados presentados en la Tabla 3.2 son consistentes con los del caso anterior, ya que nuevamente se observa que las estimaciones obtenidas mediante el diseño en dos etapas tienden a ser más conservadoras. En contraste los métodos bootstrap, especialmente el basado en mareas, arrojan valores más elevados. Nuevamente los métodos bootstrap reescalado y a nivel de lances, tiene una buena aproximación a las estimaciones del diseño tradicional, lo que refuerza su utilidad cuando se busca mantener coherencia con dicho enfoque.

DOMINIO		N° MAREAS (N° LANCES)	ESTIMACIÓN (TONELADAS)			
TRIMESTRE	AREA		DISEÑO EN 2 ETAPAS	BOOTSTRAP		
				LANCES	MAREAS	REESCALADO
1	27.7.g	1 (1)	2.14	3.47	3.24	1.96
1	27.7.h	1 (10)	10.90	10.97	16.53	10.34
1	27.7.j	1 (6)	6.09	6.12	9.24	6.16
2	27.7.b	1 (1)	0.24	0.39	0.37	0.24
2	27.7.c	3 (71)	103.76	103.44	107.20	109.74
2	27.7.h	1 (7)	10.44	10.70	15.82	10.58
2	27.7.j	2 (17)	58.25	57.67	63.99	65.67
3	27.7.b	1 (2)	5.56	6.43	8.71	5.52
3	27.7.h	2 (52)	109.95	110.12	122.65	111.43
3	27.7.j	1 (29)	57.54	56.61	89.36	59.03
4	27.7.j	1 (6)	31.77	31.04	47.70	35.33
<b>TOTAL</b>			<b>396.59</b>	<b>397.94</b>	<b>483.78</b>	<b>406.80</b>

Cuadro 3.2: Estimaciones por trimestre, subáreas 7 y metier según los métodos de diseño de muestreo y método bootstrap (Enfoque en lances).

## 3.2. Comparación de varianzas e intervalos de confianza

Para evaluar la precisión y confiabilidad de las estimaciones obtenidas mediante los distintos métodos analizados, es importante analizar las varianzas asociadas a cada estimación. La varianza es una medida estadística que indica el grado de dispersión de los valores estimados, y permite cuantificar la incertidumbre inherente al proceso de muestreo.

### 3.2.1. Caso 1: En función de mareas N

DOMINIO		N° MAREAS (N° LANCES)	VARIANZA				
TRIMESTRE	AREA		DISEÑO EN 2 ETAPAS		BOOTSTRAP		
			SRSWOR/SRSWOR	SRSWOR/DIF.	LANCES	MAREAS	REESCALADO
1	27.7.g	1 (1)	0	0	2.4328	1.9866	3.2306
1	27.7.h	1 (10)	0.2038	0.0536	18.2444	51.6684	91.2312
1	27.7.j	1 (6)	0.0531	0.0076	5.8170	16.1177	33.1709
2	27.7.b	1 (1)	0	0	0.0369	0.0237	0.0403
2	27.7.c	3 (71)	1868.4096	1868.1917	182.2828	2174.7775	2820.0405
2	27.7.h	1 (7)	0.1285	0.0477	14.5964	41.9670	81.7218
2	27.7.j	2 (17)	2345.7744	2343.7836	288.5622	2113.9738	3168.8654
3	27.7.b	1 (2)	0.1666	0.0068	9.7143	12.9523	21.9855
3	27.7.h	2 (52)	3318.8195	3318.7705	180.0221	3020.2547	4295.3206
3	27.7.j	1 (29)	0.6302	0.4043	135.0205	1359.8695	2974.5340
4	27.7.j	1 (6)	2.4442	1.7433	221.5329	459.6345	865.8705

Cuadro 3.3: Varianza por trimestre, subáreas 7 y metier según los métodos de diseño de muestreo y de Bootstrap.

Un primer punto destacable es que los dos métodos de diseño en dos etapas (SRSWOR/SRSWOR y SRSWOR/Diferencias) presentan estimaciones de varianza muy similares entre sí en la mayoría de

los dominios analizados, lo que sugiere una buena consistencia entre ambos enfoques.

En términos generales, se observa que los métodos basados en bootstrap, especialmente el enfoque de reescalado, tienden a generar valores de varianza considerablemente más elevados que el diseño en dos etapas. Esto sugiere que los métodos bootstrap capturan de forma más amplia la variabilidad muestral, aunque en algunos casos podrían sobreestimar la incertidumbre.

En dominios con mayor número de mareas y elevados niveles de descarte, como la subárea 27.7.c en el trimestre 2 o 27.7.j en el trimestre 3, las estimaciones de varianza obtenidas mediante métodos bootstrap, especialmente en el enfoque reescalado, son notablemente altas, reflejando una considerable dispersión en los datos. Por el contrario, en dominios con una única marea o con bajos niveles de descarte, como 27.7.b en el segundo trimestre, el diseño en dos etapas arroja varianzas cercanas a cero, lo que podría indicar una subestimación de la incertidumbre debido al reducido tamaño muestral.

### Evaluación del impacto de la variabilidad en la segunda etapa (lances)

En este apartado buscaremos determinar la contribución de los lances a la varianza total estimada bajo el enfoque de diseño en dos etapas, mediante el cálculo de su proporción respecto a la varianza total.

La proporción de la varianza explicada por los lances con respecto a la varianza total se calcula como:

$$\text{Proporción de Varianza (Lances)} = \frac{\sigma_{\text{Lances}}^2}{\sigma_{\text{Total}}^2}$$

donde:

- $\sigma_{\text{Lances}}^2$  representa la varianza explicada por los lances.
- $\sigma_{\text{Total}}^2$  representa la varianza total.

Esta proporción nos permite visualizar en que medida la segunda etapa (lances) contribuye a la variabilidad total. Un valor cercano a 1 indica que la mayor variabilidad se concentra en esta etapa, mientras que un valor bajo sugiere una influencia menor.

Al observar los gráficos de la Figura 3.1, se aprecia que en la mayoría de los casos la varianza asociada a los lances, es decir, la variabilidad dentro de cada marea, es muy baja o incluso nula. Aunque, en algunos casos se puede apreciar que representa el 100 %, la varianza total solo es representada por la varianza asociada a los lances. Estos resultados indican que el aporte de los lances a la varianza total es mínimo, y puede considerarse despreciable en términos prácticos, sin comprometer la calidad ni la precisión del análisis de descartes. En consecuencia, la mayor parte de la variabilidad se concentra en la primera etapa del diseño (mareas), lo que justifica enfocar el análisis principalmente en esta fase.

Por lo tanto, en adelante se representará la variabilidad considerando únicamente la primera etapa (mareas) del diseño en dos etapas.

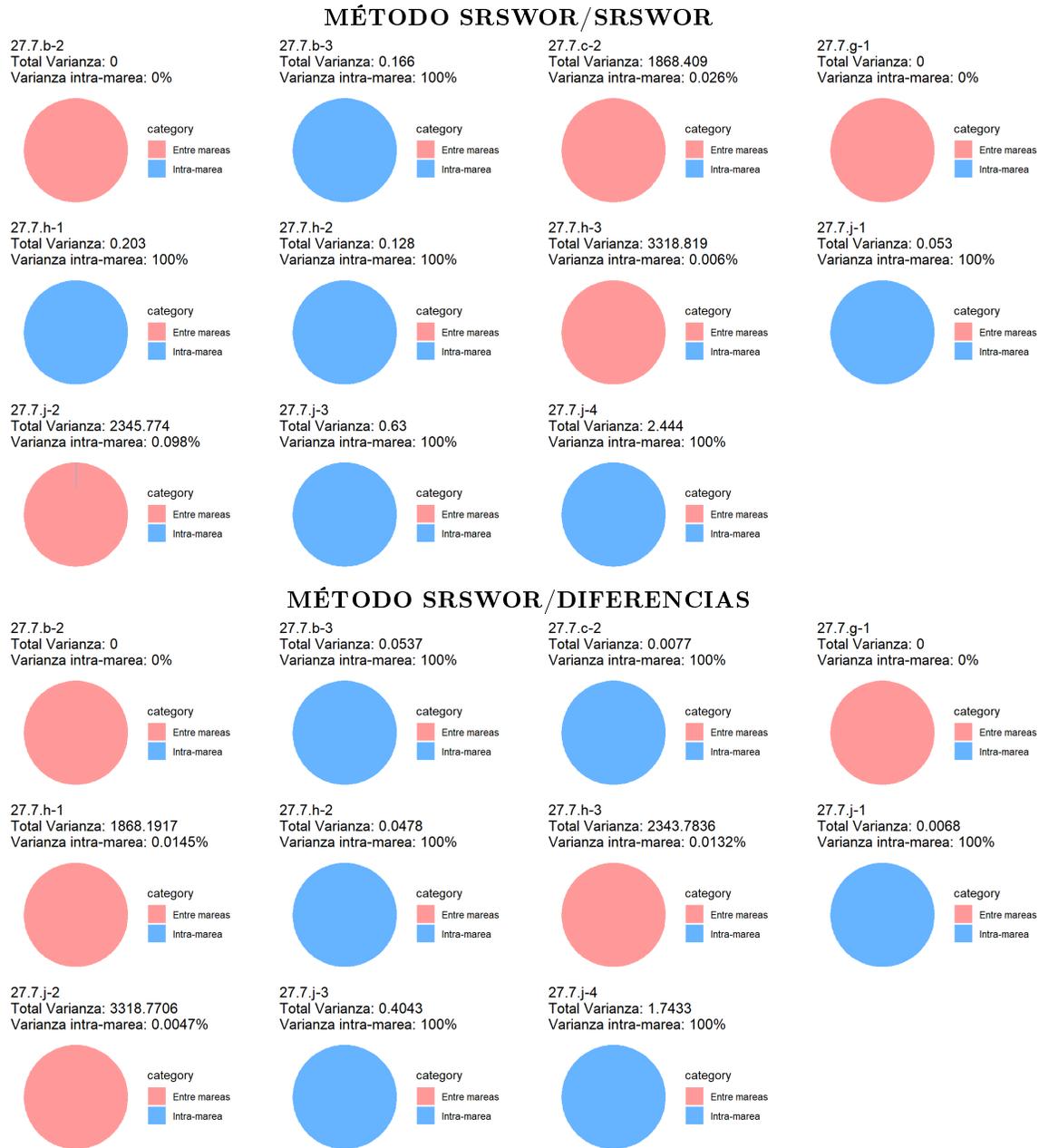


Figura 3.1: Porporcion de varianza por etapa de muestreo en cada domain ( trimestre, subáreas 7 y metier) para los métodos de diseño en 2 etapas

**Estimación y visualización de intervalos de confianza**

Una vez obtenidas las varianzas, el siguiente paso consiste en representarlas gráficamente junto con sus respectivos intervalos de confianza. Esta visualización permite evaluar de manera más clara la precisión de las estimaciones y comparar la estabilidad de los distintos métodos analizados.

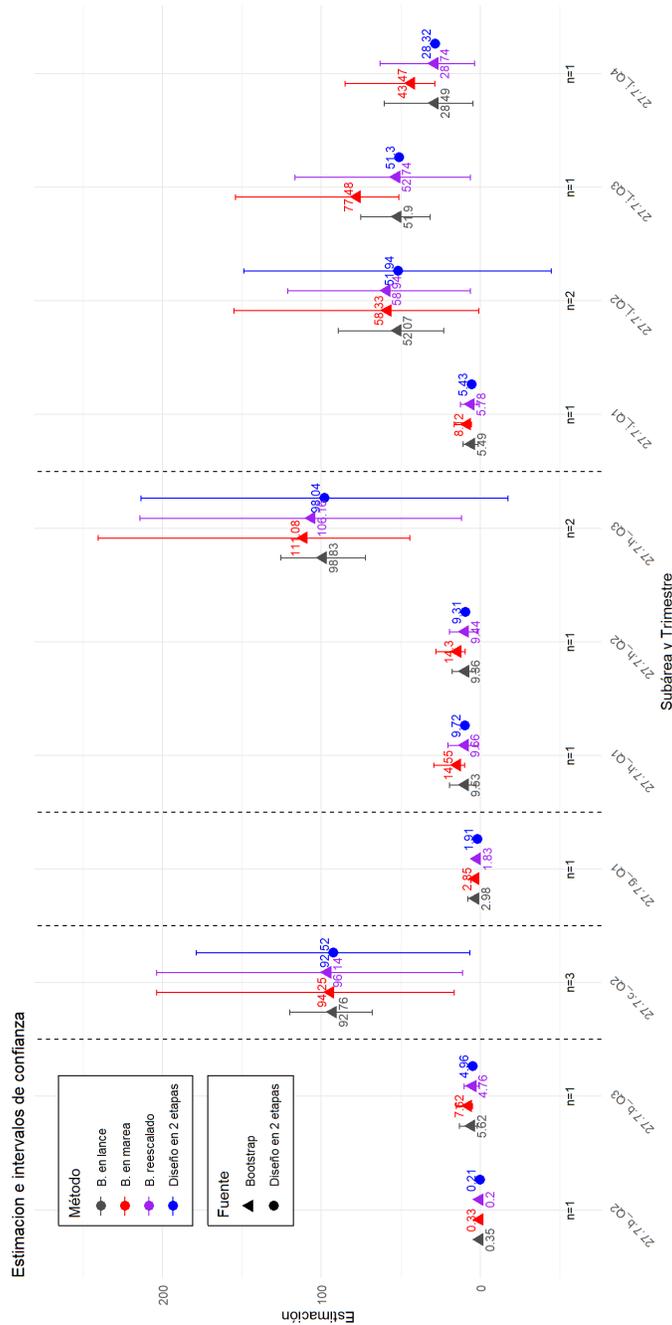


Figura 3.2: Estimador total de descarte e intervalos de confianza por trimestre, subáreas 7 y metier para los métodos de diseño de muestro (marea y lance) y de Bootstrap. Consultar Tabla 5.1

Los intervalos de confianza representados en la Figura 3.2, revelan diferencias importantes en la forma en que cada método representa la incertidumbre asociada, particularmente en condiciones de bajo tamaño muestral (entre 1 y 3 mareas). El diseño en dos etapas genera en varios casos intervalos que incluyen valores negativos, lo que sugiere limitaciones en su aplicación con muestras pequeñas y sugiere una subestimación de la varianza. En cambio, el método bootstrap en marea logra intervalos más amplios y realistas, reflejando adecuadamente la variabilidad entre mareas. Por su parte, el método bootstrap por lance capta de forma más precisa la variabilidad intra-marea, pero tiende a generar intervalos algo más estrechos, dado que no incorpora la variabilidad entre mareas. Finalmente, el método bootstrap reescalado, que ajusta la varianza teniendo en cuenta el diseño original, produce los intervalos más amplios y conservadores. Este enfoque ofrece una representación más prudente y coherente de la incertidumbre total. En síntesis, los métodos bootstrap, especialmente en mareas y el reescalado, resultan más adecuados para contextos de escaso tamaño muestral, al ofrecer estimaciones más estables y estadísticamente consistentes que el diseño clásico en dos etapas.

### 3.2.2. Caso 2: En función lances $N = \frac{M_0}{M}$

Como se detalló en el caso anterior, ahora el análisis se centrará exclusivamente en la variabilidad observada entre mareas, despreciando la contribución de la varianza intra mareas.

DOMINIO		N° MAREAS (N° LANCES)	VARIANZA			
TRIMESTRE	AREA		DISEÑO EN 2 ETAPAS	BOOTSTRAP		
				LANCES	MAREAS	REESCALADO
1	27.7.g	1 (1)	0	3.2866	2.3564	4.0450
1	27.7.h	1 (10)	0	25.9417	61.2865	113.6772
1	27.7.j	1 (6)	0	7.3532	19.1181	40.9825
2	27.7.b	1 (1)	0	0.0370	0.0338	0.0515
2	27.7.c	3 (71)	2399.3041	235.0674	2759.6245	3429.2635
2	27.7.h	1 (7)	0	18.0857	59.4198	103.3119
2	27.7.j	2 (17)	2998.2343	391.2391	2778.0356	3938.8572
3	27.7.b	1 (2)	0	12.1474	18.4198	28.0642
3	27.7.h	2 (52)	4151.3797	250.6227	3591.0659	5290.8503
3	27.7.j	1 (29)	0	167.3420	1820.8110	3722.5647
4	27.7.j	1 (6)	0	258.9485	486.3537	1107.0672

Cuadro 3.4: Varianza por trimestre, subáreas 7 y metier según los métodos de diseño de muestro y bootstrap.

Al igual que en el caso anterior, la varianza asociada a la estimación del descarte total en este escenario evidencia una marcada dependencia del tamaño muestral y del método estadístico utilizado. En dominios con una sola marea observada, el estimador de varianza bajo un diseño en dos etapas resulta inaplicable, ya que no es posible estimar la variabilidad entre mareas a partir de una sola observación. En estos casos, los métodos bootstrap permiten obtener estimaciones de varianza, aunque con valores considerablemente mayores, especialmente cuando se incorpora un reescalado para reflejar la variabilidad del diseño.

Estimación y visualización de intervalos de confianza

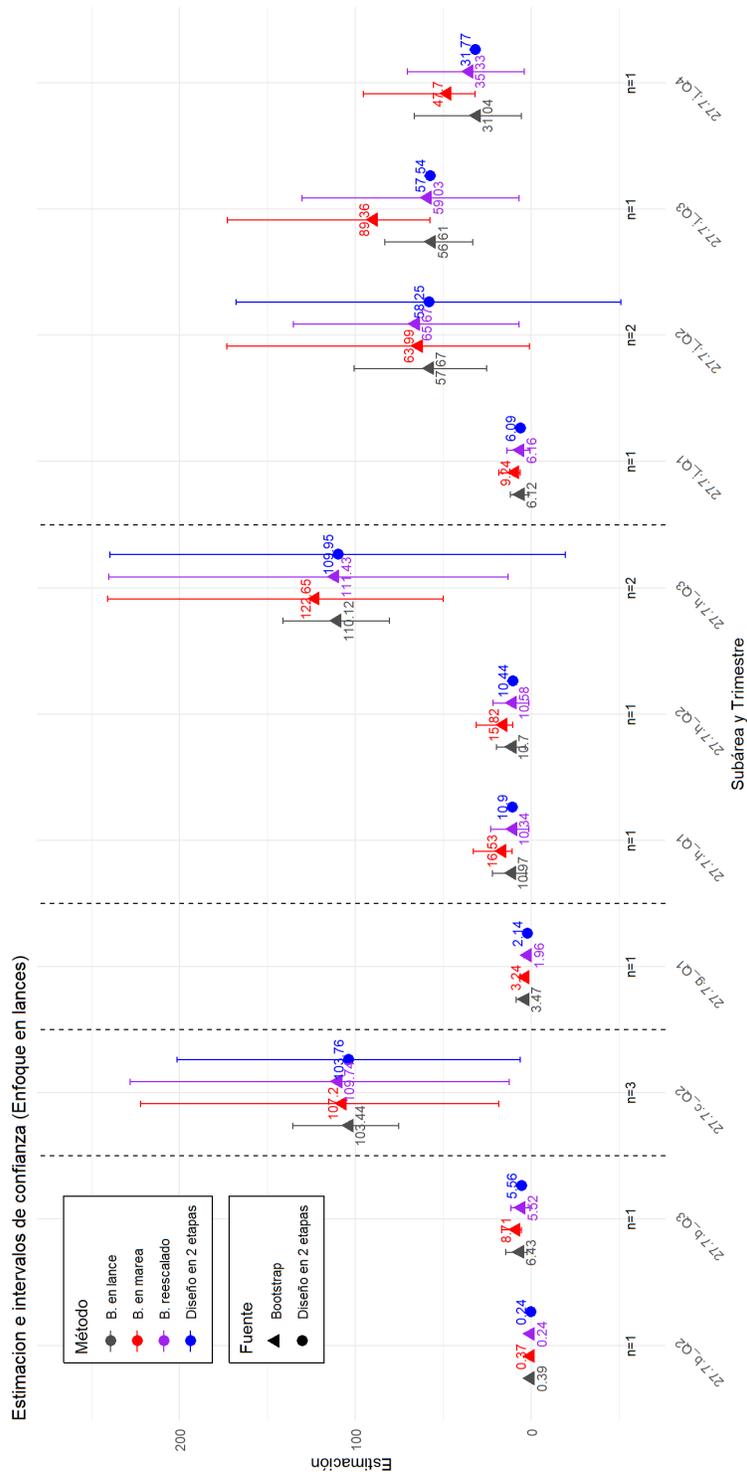


Figura 3.3: Estimador e intervalos de confianza por trimestre, subáreas 7 y metier para los métodos estudiados (Enfoque en lances). Consultar Tabla 5.2

Los intervalos de confianza de la Figura 3.3 obtenidos para el estimador del descarte total muestran una tendencia coherente con los resultados presentados en la Figura 3.2, con estimaciones similares en magnitud y rango. El diseño de muestreo en dos etapas presenta las mismas limitaciones observadas anteriormente: en varios casos los intervalos incluyen valores negativos, lo cual no es coherente con la naturaleza del descarte, cuya magnitud no puede ser inferior a cero. En contraste, los métodos bootstrap, especialmente en sus versiones por marea y reescalada, ofrecen intervalos de confianza más realistas y consistentes, con límites inferiores positivos y rangos plausibles. Esta concordancia entre métodos y resultados refuerza la robustez del estimador aplicado y sugiere que las técnicas bootstrap son más adecuadas para cuantificar la incertidumbre asociada al descarte total en tamaños muestrales pequeños.

### 3.3. Comparación de metodologías

En términos metodológicos, los diferentes enfoques utilizados para estimar los intervalos de confianza del descarte total muestran contrastes significativos. Los métodos clásicos, como SRSWOR/SRSWOR y SRSWOR/SYSS, parten de supuestos estadísticos más rígidos, como la normalidad, que en contextos con alta variabilidad o tamaños muestrales limitados, pueden dar lugar a intervalos poco realistas, incluso con valores negativos, lo cual no tiene sentido práctico en el caso del descarte. Por el contrario, los métodos basados en remuestreo, como el bootstrap, ofrecen una alternativa más flexible al no depender de estos supuestos paramétricos. En particular, el bootstrap aplicado a nivel de marea y su versión reescalada permiten obtener intervalos más coherentes con los datos observados, siempre positivos y ajustados a la distribución empírica del fenómeno. En general, estos métodos resultan más robustos y adecuados para estimar la incertidumbre en muestras pequeñas o distribuciones no normales.



## Capítulo 4

# Discusión

### 4.1. Comparación con el método oficial actual del IEO-CSIC

Actualmente, el Instituto Español de Oceanografía (IEO-CSIC) estima los descartes a partir de los pesos medios por marea, calculados de forma agregada para cada trimestre y por dominio. Este método no tiene en cuenta ni los lances individuales ni las divisiones geográficas dentro del área de estudio. Una vez obtenida esta estimación trimestral global, se reparte entre las distintas divisiones en función del esfuerzo pesquero ejercido por la flota en cada una.

Este enfoque, aunque operativo y funcional para obtener estimaciones generales, presenta limitaciones importantes. Al trabajar con valores promedio para áreas muy amplias, se pierden posibles variaciones espaciales significativas, lo que puede ocultar diferencias entre divisiones. Lo cual puede comprometer la calidad de las evaluaciones y decisiones de gestión basadas en estos datos.

En esta tesis se propone una metodología alternativa basada en el uso de la base de datos RDBES, desarrollada a nivel europeo para estandarizar los datos y estimaciones en el ámbito de la pesca. A partir de los datos disponibles en RDBES, se aplica un diseño muestral en dos etapas, en el que se consideran las mareas como unidades primarias de muestreo y los lances como unidades secundarias. Sobre esta estructura, se implementan distintos métodos para estimar la varianza:

- Bootstrap en marea
- Bootstrap en lance
- Bootstrap reescalado
- Varianza clásica bajo diseño muestral en dos etapas

Este enfoque permite obtener estimaciones de descarte desagregadas por división geográfica, respetando la estructura jerárquica del muestreo y mejorando así la resolución espacial del análisis. En resumen, el objetivo de este trabajo no es solo mejorar la precisión en las estimaciones de descarte, sino también ofrecer una herramienta estadística más completa y adaptada a los estándares actuales, como los que exige el uso de bases de datos como RDBES.

### 4.2. Implicaciones para la implementación del RDBES

La implementación de RDBES representa una oportunidad para modernizar y estandarizar la gestión de datos pesqueros. Su estructura, basada en distintos niveles jerárquicos, requiere una adaptación

tanto en la forma de recoger los datos como en los métodos de análisis que se utilizan.

En esta tesis se ha trabajado con datos ya estructurados según el modelo RDBES, lo que ha permitido aplicar metodologías estadísticas más detalladas y realistas, como el diseño muestral en dos etapas y diversos enfoques para estimar la varianza. Estos métodos, además de ajustarse a la lógica del sistema, ofrecen resultados más precisos y permiten cuantificar la incertidumbre de forma clara. Adoptar RDBES no es solo una cuestión técnica, sino también una transformación en la forma de entender y trabajar con los datos, ya que permite desagregarlos por zonas y periodos, detectar patrones más específicos y seguir el recorrido del dato desde su recogida hasta la estimación final.

### 4.3. Limitaciones del estudio

A pesar de los avances metodológicos presentados en este trabajo, existen algunas limitaciones que deben tenerse en cuenta al interpretar los resultados.

En varios dominios analizados, el número de mareas disponibles fue muy reducido (en algunos casos limitado a una sola marea). Esto imposibilita la aplicación de ciertos métodos de estimación de varianza bajo diseño muestral y puede afectar la estabilidad de las estimaciones obtenidas generadas mediante bootstrap.

Además, aunque RDBES ofrece una estructura muy detallada y flexible para la recolección de datos pesqueros, en la práctica, muchos de sus campos opcionales aún no están siendo utilizados de forma sistemática. Esta situación limita el potencial del sistema para capturar información clave, como esfuerzo a nivel de lance, detalles adicionales del muestreo a bordo, etc.

La ausencia de esta información reduce la posibilidad de aprovechar completamente estas variables reduciendo la capacidad analítica del sistema y reduciendo el uso de modelos más complejos o desagregados. Por tanto, una implementación más completa y homogénea de RDBES, que incluya tanto los campos obligatorios como los opcionales más relevantes, permitiría enriquecer los análisis y obtener estimaciones más precisas y contextualizadas.

### 4.4. Recomendaciones para el futuro

A partir de los resultados obtenidos y de las limitaciones identificadas en este estudio, se proponen las siguientes líneas de mejora y trabajo futuro:

- Como línea futura de desarrollo metodológico, se recomienda considerar la implementación de modelos bayesianos, que permiten representar de forma más flexible estructuras jerárquicas y generar distribuciones completas de incertidumbre. Este enfoque sería especialmente útil en dominios con pocos datos o alta variabilidad, y se adapta bien a la estructura multinivel de RDBES.
- Sería útil realizar simulaciones con datos sintéticos para ver cómo se comportan los distintos métodos de estimación de varianza bajo condiciones específicas. Esto ayudaría la comparación objetiva del rendimiento en distintos escenarios y a adaptar mejor las metodologías a las características específicas de cada pesquería.
- Muchos de los campos disponibles en el estándar RDBES aún no se están utilizando de forma sistemática, lo que limita su potencial. Sería recomendable avanzar hacia una implementación más completa.
- Para mejorar la robustez de las estimaciones, sería conveniente ampliar el tamaño muestral en dominios, especialmente de la varianza, sería conveniente asegurar un número mínimo de mareas

observadas por dominio. Esto permitiría aplicar con mayor confianza los métodos basados en diseño muestral y mejorar la estabilidad de los enfoques bootstrap.

- Tuve la oportunidad de presentar estos resultados ante el grupo de expertos del STECF, los cuales fueron muy bien recibidos. De hecho, manifestaron interés en que compartiera los resultados y procedimientos aplicados, ya que podrían servir de base para desarrollar recomendaciones metodológicas que ayuden a los distintos países a mejorar la estimación de sus descartes y las varianzas asociadas, de esta forma contribuir a mejorar la calidad y comparabilidad de los datos a nivel europeo.



## Capítulo 5

# Conclusiones

1. Este trabajo ha demostrado el potencial de mejorar las estimaciones de descartes pesqueros mediante el uso de metodologías más detalladas y adaptadas a las nuevas estructuras de datos, como RDBES. A diferencia del enfoque actualmente empleado por el IEO-CSIC, basado en promedios trimestrales agregados y distribuciones según esfuerzo, el método propuesto permite una estimación más precisa y desagregada, respetando la jerarquía muestral real (marea-lance). Es importante resaltar que el sistema RDBES no solo actúa como un repositorio de datos, sino que también proporciona una estructura sólida para la aplicación de metodologías estadísticas avanzadas. No obstante, su implementación enfrenta retos en la cobertura de campos opcionales y en el tamaño muestral en algunos dominios.
2. Los dos métodos clásicos (SRSWOR/SRSWOR y SRSWOR/SYSS) para estimar la varianza en muestreo en dos etapas, ya sea usando el enfoque en marea o en lance, arrojan resultados prácticamente idénticos en las zonas analizadas. Esta consistencia indica que, aunque parten de enfoques diferentes dentro del mismo sistema de muestreo, ambos métodos son metodológicamente válidos, y que la elección entre uno u otro puede depender principalmente de factores prácticos, como la disponibilidad y organización de los datos.
3. En el enfoque de los métodos clásicos, los resultados obtenidos indican que, en muchos casos, desprestigiar la variabilidad intra-marea en la estimación de la varianza no altera de forma significativa los resultados. Esto sugiere que, cuando la variabilidad entre lances dentro de una misma marea es reducida en comparación con la variabilidad entre mareas, puede considerarse una simplificación válida omitir esa componente en el análisis.
4. Este estudio se concluye que, el método clásico bajo diseño en dos etapas es eficiente y apropiado cuando se dispone de un tamaño muestral adecuado y una estructura clara del muestreo, ya que utiliza explícitamente la jerarquía entre mareas y lances para calcular la varianza. Por su parte, los métodos bootstrap ofrecen mayor flexibilidad para adaptarse a situaciones con muestras pequeñas o distribuciones de datos complejas, al generar múltiples réplicas del muestreo original para estimar la variabilidad. Dentro de los métodos bootstrap, el bootstrap reescalado es especialmente adecuado porque combina el remuestreo en mareas y lances, ajustando las estimaciones para reflejar correctamente la estructura jerárquica del diseño muestral. Esta característica le confiere mayor precisión para captar tanto la variabilidad dentro de las mareas como entre ellas, mejorando la estabilidad de las estimaciones, especialmente en dominios con muestras reducidas.

5. Finalmente, se recomienda mantener una evaluación continua y comparativa de ambos métodos en los distintos estratos de muestreo. Esta práctica permitirá determinar de manera más precisa cuál es el enfoque más apropiado para cada situación específica, fortaleciendo así la calidad, fiabilidad y exactitud de las estimaciones obtenidas. Asimismo se sugiere realizar comparativas entre los diferentes años que cuenten con su implementación en RDBES.

# Bibliografía

- Acosta-Yepes, J. et al. (2019). *Análisis de la actividad pesquera de la flota española de aguas ibéricas atlánticas y su uso en la gestión de stocks*. Madrid: Sede Central IEO.
- Arnab, Raghunath (2017). *Survey Sampling Theory and Applications*. 1.<sup>a</sup> ed. Academic Press. ISBN: 0128118482; 9780128118481.
- Borges, Lisa (2009). *The evolution of a discard policy in Europe*. ICES CM 2009/M:02. Accedido en 2025. International Council for the Exploration of the Sea (ICES). URL: <https://www.ices.dk/sites/pub/CM%20Documents/CM-2009/M/M0209.pdf>.
- Cochran, William G. (1977). *Sampling Techniques*. 3.<sup>a</sup> ed. New York: John Wiley & Sons.
- Efron, Bradley y R.J. Tibshirani (1993). *An Introduction to the Bootstrap*. 1.<sup>a</sup> ed. Chapman y Hall/CRC.
- FAO (2019). *A third assessment of global marine fisheries discards*. Technical Paper 633. Rome. URL: <https://www.fao.org/3/ca2905en/ca2905en.pdf>.
- (2022a). *El estado mundial de la pesca y la acuicultura 2022: Hacia la transformación azul*. Roma: FAO. DOI: 10.4060/cc0461es. URL: <https://doi.org/10.4060/cc0461es>.
- (2022b). *Stock assessment for fishery management*. Rome: FAO. URL: <https://openknowledge.fao.org/server/api/core/bitstreams/b7b8da38-3a41-47ea-ae58-5b78985f7490/content>.
- ICES (2024). *Documentation of the Regional Database and Estimation System Data Model*. Accedido en 2025. URL: <https://github.com/ices-tools-dev/RDBES>.
- (2025). *Regional Database and Estimation System (RDBES)*. Accedido en 2025. Copenhagen, Denmark. URL: <https://rdbes.ices.dk/>.
- Lohr, Sharon L. (2009). *Sampling: Design and Analysis*. 2.<sup>a</sup> ed. Boston, MA: Brooks/Cole.
- Preston, John (2009). “Rescaled bootstrap for stratified multistage sampling”. En: *Survey Methodology* 35.2, págs. 227-234.
- Rao, J. N. K. y C. F. J. Wu (1988). “Resampling inference with complex survey data”. En: *Journal of the American Statistical Association* 83.401, págs. 231-241.
- STECF (2023). *FDI methodology*. Inf. téc. Accedido en 2025. European Commission, Joint Research Centre. URL: [https://stecf.jrc.ec.europa.eu/documents/d/stecf/stecf\\_23-05](https://stecf.jrc.ec.europa.eu/documents/d/stecf/stecf_23-05).

- Tamsett, D. et al. (1999). “Onboard sampling for measuring discards in commercial fishing based on multilevel modelling of measurements in the Irish Sea from NW England and N Wales”. En: *Fisheries Research* 42.1-2, págs. 117-126. DOI: [10.1016/S0165-7836\(99\)00030-2](https://doi.org/10.1016/S0165-7836(99)00030-2).
- Thompson, Steven K. (2012). *Sampling*. 3rd. Hoboken, NJ: Wiley.
- UE (2008). *Reglamento (CE) nº 199/2008 del Consejo*. Diario Oficial de la Unión Europea, L 60, 3.3.2008, pp. 1-14. Disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32008R0199>.
- Vigneau, Joël (2006). *Raising procedures for discards: Sampling theory*. ICES CM 2006/K:16. Retrieved from <https://archimer.ifremer.fr/doc/00000/2699/>.

# Anexo

A continuación, se presentan los algoritmos diseñados específicamente para resolver las problemáticas abordadas en este estudio y los resultados de los intervalos de confianza analizados y mostrados visualmente en la sección 3.

Listing 5.1: Código R para la implementación de los métodos clásico del diseño en etapas etapas

```
discard_estimate <- function(Samp, domain, alpha) {

  FT_total <- Samp %>%
  group_by(across(all_of(domain)), FTid) %>%
  summarise(
    y_i = sum(TotalWeightLive),
    m_i = m_gral[1],
    M_i = M_gral[1],
    n_gral = n_gral[1],
    N_gral = N_gral[1],
    m_row = n(),
    y_bar_i = y_i / m_i,
    y_hat_i = M_i * y_bar_i,
    FO_var_i = ifelse(m_row > 1, sum((TotalWeightLive - y_bar_i)^2) / (m_i - 1), 0),
    var_SRSWOR = ifelse(m_row > 0, M_i^2 * (1 - m_i / M_i) * FO_var_i / m_i, 0),
    FO_var_diff_i = ifelse(m_row > 1,
      sum((TotalWeightLive[1:(length(FOid) - 1)] -
        TotalWeightLive[2:length(FOid)])^2) /
      (2 * (m_i - 1)), 0),
    var_SYSS = ifelse(m_row > 0, M_i^2 * (1 - m_i / M_i) * FO_var_diff_i / m_i, 0),
    .groups = 'drop')

  Samp_var <- FT_total %>%
  group_by(across(all_of(domain))) %>%
  summarise(
    df = sum(m_i) - 1,
    N = N_gral[1],
    n = n_gral[1],
    n_row = n(),
    y_bar = sum(y_hat_i) / n,
    y_hat = N * y_bar,
    FT_var = ifelse(n_row > 1, sum((y_hat_i - y_bar)^2) / (n - 1), 0),
    FT_var_aux = ifelse(n_row > 1, sum((1 - n / N) * (y_hat_i - y_bar)^2) / (n - 1), 0),
    FO_SRSWOR = sum(var_SRSWOR),
    FO_SYSS = sum(var_SYSS),
    .groups = 'drop') %>%
  mutate(
    var_between = N ^ 2 * FT_var / n,
    var_between_aux = N ^ 2 * FT_var_aux / n,
```

```

var_within_SRSWOR = N * FO_SRSWOR / n,

var_total_SRSWOR = var_between + var_within_SRSWOR,
var_total_aux_SRSWOR = var_between_aux + var_within_SRSWOR,

UC_SRSWOR = y_hat + qt(1 - alpha / 2, df) * sqrt(var_total_SRSWOR),
LC_SRSWOR = y_hat - qt(1 - alpha / 2, df) * sqrt(var_total_SRSWOR),

UC_aux_SRSWOR = y_hat + qt(1 - alpha / 2, df) * sqrt(var_total_aux_SRSWOR),
LC_aux_SRSWOR = y_hat - qt(1 - alpha / 2, df) * sqrt(var_total_aux_SRSWOR),

var_within_SYSS = N * FO_SYSS / n,

var_total_SYSS = var_between + var_within_SYSS,
var_total_aux_SYSS = var_between_aux + var_within_SYSS,

UC_SYSS = y_hat + qt(1 - alpha / 2, df) * sqrt(var_total_SYSS),
LC_SYSS = y_hat - qt(1 - alpha / 2, df) * sqrt(var_total_SYSS),

UC_aux_SYSS = y_hat + qt(1 - alpha / 2, df) * sqrt(var_total_aux_SYSS),
LC_aux_SYSS = y_hat - qt(1 - alpha / 2, df) * sqrt(var_total_aux_SYSS),
)

return(Samp_var)
}

```

Listing 5.2: Código R para la implementación del métodos bootstrap en lances

```

discard_estimate_boot_haul <- function(datos, domain, SP, alpha) {

  B <- 1000
  upms <- unique(datos$FTid)

  result <- system.time({

    bootstrap_iteration <- function(b) {
      datos_b <- do.call(rbind, lapply(upms, function(id) {
        usm_orig <- datos[datos$FTid == id, ]
        usm_remuestreadas <- usm_orig[sample(nrow(usm_orig), replace = TRUE), ]
        return(usm_remuestreadas)
      })))

    datos_b <- datos_b %>%
      filter(SAspeciesCode == SP,
             SScatchFraction == "Dis")

    FT_total <- datos_b %>%
      group_by(across(all_of(domain)), FTid) %>%
      summarise(
        y_i = sum(TotalWeightLive),
        m_i = m_gral[1],
        M_i = M_gral[1],
        n_gral = n_gral[1],
        N_gral = N_gral[1],
        y_bar_i = y_i / m_i,
        y_hat_i = M_i * y_bar_i,

```

```

    .groups = 'drop')

Samp_var <- FT_total %>%
  group_by(across(all_of(domain))) %>%
  summarise(
    df = sum(m_i) - 1,
    N = N_gral[1],
    n = n_gral[1],
    y_bar = sum(y_hat_i) / n,
    y_hat = N * y_bar,
    .groups = 'drop')

  return(Samp_var)
}

cl <- makeCluster(detectCores() - 1)
clusterExport(cl, varlist = c("datos", "upms", "domain", "SP"), envir = environment())
clusterEvalQ(cl, {
  library(stats)
  library(dplyr)
})

y_hat_boot <- parLapply(cl, 1:B, bootstrap_iteration)

var_result <- bind_rows(y_hat_boot, .id = "iter") %>%
  group_by(across(all_of(domain))) %>%
  summarise(
    mean_boot_haul = mean(y_hat),
    var_y_hat_haul = var(y_hat),
    LC_boot_haul = quantile(y_hat, 1 - alpha / 2),
    UC_boot_haul = quantile(y_hat, alpha / 2),
    .groups = "drop"
  )

stopCluster(cl)
})

print(result)

return(var_result)
}

```

Listing 5.3: Código R para la implementación del métodos bootstrap en mareas

```

discard_estimate_boot_trip <- function(datos, domain, SP, alpha) {

  B <- 1000
  upms <- unique(datos$FTid)

  result <- system.time({

    bootstrap_iteration <- function(b) {
      upms_resampled <- sample(upms, replace = TRUE)
      datos_b <- do.call(rbind, lapply(seq_along(upms_resampled), function(i) {
        id <- upms_resampled[i]

```

```

    usm_orig <- datos[datos$FTid == id, ]
    usm_orig$UPMid <- paste0(id, "_", i)
    return(usm_orig)
  })

  datos_b <- datos_b %>%
    filter(SASpeciesCode == SP,
           SScatchFraction == "Dis")

  FT_total <- datos_b %>%
    group_by(across(all_of(domain)), UPMid) %>%
    summarise(
      y_i = sum(TotalWeightLive),
      m_i = m_gral[1],
      M_i = M_gral[1],
      n_gral = n_gral[1],
      N_gral = N_gral[1],
      y_bar_i = y_i / m_i,
      y_hat_i = M_i * y_bar_i,
      .groups = 'drop')

  Samp_var <- FT_total %>%
    group_by(across(all_of(domain))) %>%
    summarise(
      df = sum(m_i) - 1,
      N = N_gral[1],
      n = n_gral[1],
      y_bar = sum(y_hat_i) / n,
      y_hat = N * y_bar,
      .groups = 'drop')

  return(Samp_var)
}

cl <- makeCluster(detectCores() - 1)
clusterExport(cl, varlist = c("datos", "upms", "domain", "SP"), envir = environment())
clusterEvalQ(cl, {
  library(stats)
  library(dplyr)
})

y_hat_boot <- parLapply(cl, 1:B, bootstrap_iteration)

var_result <- bind_rows(y_hat_boot, .id = "iter") %>%
  group_by(across(all_of(domain))) %>%
  summarise(
    mean_boot_trip = mean(y_hat),
    var_y_hat_trip = var(y_hat),
    LC_boot_trip = quantile(y_hat, 1 - alpha / 2),
    UC_boot_trip = quantile(y_hat, alpha / 2),
    .groups = "drop"
  )

stopCluster(cl)
})

```

```

print(result)

return(var_result)
}

```

Listing 5.4: Código R para la implementación del métodos bootstrap reescalado

```

discard_estimate_boot_rescaled <- function(datos, domain, SP, alpha) {

  B <- 1000
  datos$ID <- 1:nrow(datos)
  upms <- unique(datos$FTid)

  result <- system.time({

    bootstrap_iteration <- function(b) {

      upms_resampled <- sample(upms, pmax(1, floor(length(upms) / 2)), replace = FALSE)

      datos_weight <- datos %>%
        mutate( delta_trip = if_else(FTid %in% upms_resampled, 1, 0) )

      datos_b <- do.call(rbind, lapply(upms, function(i) {
        usm_orig <- datos[datos$FTid == i, ]
        usm_resampled <- usm_orig[sample(nrow(usm_orig) / 2, replace = FALSE), ] %>%
          mutate(delta_haul = 1)
        return(usm_resampled)
      })))

      datos_weight <- datos_weight %>% left_join(
        datos_b %>%
          ungroup() %>%
          select(ID, delta_haul),
        by = "ID") %>%
        mutate(
          delta_trip = coalesce(delta_trip, 0),
          delta_haul = coalesce(delta_haul, 0),
          weight_trip = N_gral / n_gral,
          n_r = pmax(1, floor(n_gral / 2)),
          prob_n = n_gral / N_gral,
          lambda_1 = sqrt(n_r * (1 - prob_n) / (n_gral - n_r)),
          weight_r_trip = weight_trip * (1 - lambda_1 + lambda_1 * (n_gral / n_r) * delta_trip
          ),
          weight_haul = M_gral / m_gral, # peso original de lance
          m_r = pmax(1, floor(m_gral / 2)),
          prob_m = m_gral / M_gral,
          lambda_2 = sqrt(m_r * prob_n * (1 - prob_m) / (m_gral - m_r)),
          weight_r_haul = weight_haul * weight_trip * (1 - lambda_1 + lambda_1 * (n_gral / n_r
          ) *
            delta_trip - lambda_2 * sqrt(n_r / n_gral) * delta_trip + lambda_2 *
            sqrt(n_gral / n_r) * delta_trip * (m_gral / m_r) * delta_haul) /
            weight_r_trip,
        )

    }

  )

  # nuestra subpoblacion

```

```

datos_weight <- datos_weight %>%
  filter(SAspeciesCode == SP,
         SScatchFraction == "Dis")

# calculo del estimador bootstrap:

FT_total <- datos_weight %>%
  group_by(across(all_of(domain)), FTid) %>%
  summarise(
    y_hat_i = sum(weight_r_haul * TotalWeightLive),
    weight_r_trip = weight_r_trip[1],
    .groups = 'drop')

Samp_var <- FT_total %>%
  group_by(across(all_of(domain))) %>%
  summarise(
    y_hat = sum(weight_r_trip * y_hat_i),
    .groups = 'drop')

  return(Samp_var)
}

cl <- makeCluster(detectCores() - 1)
clusterExport(cl, varlist = c("datos", "upms", "domain", "SP"), envir = environment())
clusterEvalQ(cl, {
  library(stats)
  library(dplyr)
})

y_hat_boot <- parLapply(cl, 1:B, bootstrap_iteration)

var_result <- bind_rows(y_hat_boot, .id = "iter") %>%
  group_by(across(all_of(domain))) %>%
  summarise(
    mean_boot_rescaled = mean(y_hat),
    var_y_hat_rescaled = var(y_hat),
    LC_boot_rescaled = quantile(y_hat, 1 - alpha / 2),
    UC_boot_rescaled = quantile(y_hat, alpha / 2),
    .groups = "drop"
  )

stopCluster(cl)
})

print(result)

return(var_result)
}

```

DOMINIO	N° MAREAS (N° LANCES)		INTERVALOS DE CONFIANZA											
	TRIMESTRE	AREA	DISEÑO EN 2 ETAPAS				LANCES				BOOTSTRAP			
			UC	LC	UC	LC	UC	LC	UC	LC	UC	LC	UC	LC
1	27.7.g	1 (1)	1.91	1.91	7.63	1.91	1.91	5.72	1.91	1.91	3.85	0.23		
1	27.7.h	1 (10)	9.72	9.72	19.42	2.66	2.66	29.17	9.72	9.72	20.39	1.17		
1	27.7.j	1 (6)	5.43	5.43	10.66	1.35	1.35	16.29	5.43	5.43	12.24	0.65		
2	27.7.b	1 (1)	0.21	0.21	0.85	0.21	0.21	0.64	0.21	0.21	0.43	0.03		
2	27.7.c	3 (71)	178.62	6.42	120.04	68.03	68.03	197.90	16.48	16.48	203.63	11.11		
2	27.7.h	1 (7)	9.31	9.31	17.68	2.75	2.75	27.93	9.31	9.31	19.25	1.12		
2	27.7.j	2 (17)	148.71	-44.83	89.43	22.89	22.89	154.22	0.80	0.80	121.14	6.24		
3	27.7.b	1 (2)	4.96	4.96	12.89	2.00	2.00	14.88	4.96	4.96	10.03	0.60		
3	27.7.h	2 (52)	213.69	-17.61	125.49	72.12	72.12	240.47	44.39	44.39	214.32	11.77		
3	27.7.j	1 (29)	51.30	51.30	75.30	31.49	31.49	153.91	51.30	51.30	116.54	6.16		
4	27.7.j	1 (6)	28.32	28.32	60.29	4.56	4.56	84.97	28.32	28.32	62.88	3.40		

Cuadro 5.1: Intervalos de confianza por trimestre, subáreas 7 y metier según el método de diseño y el método Bootstrap.

DOMINIO	N° MAREAS (N° LANCES)	INTERVALOS DE CONFIANZA											
		DISEÑO EN 2 ETAPAS						BOOTSTRAP					
		LANCES		MAREAS		REESCALADO		LANCES		MAREAS		REESCALADO	
TRIMESTRE	AREA	UC	LC	UC	LC	UC	LC	UC	LC	UC	LC	UC	LC
1	27.7.g	1 (1)	2.14	2.14	8.55	2.14	2.14	6.41	2.14	4.33	0.25	4.33	0.25
1	27.7.h	1 (10)	10.90	10.90	22.11	2.87	2.87	32.71	10.90	22.89	1.29	22.89	1.29
1	27.7.j	1 (6)	6.09	6.09	11.89	1.51	1.51	18.27	6.09	13.69	0.72	13.69	0.72
2	27.7.b	1 (1)	0.24	0.24	0.95	0.24	0.24	0.72	0.24	0.48	0.03	0.48	0.03
2	27.7.c	3 (71)	201.34	6.18	135.39	75.17	75.17	221.94	18.48	227.98	12.29	227.98	12.29
2	27.7.h	1 (7)	10.44	10.44	19.68	3.21	3.21	31.32	10.44	21.63	1.24	21.63	1.24
2	27.7.j	2 (17)	167.71	-51.20	100.61	25.37	25.37	172.96	0.90	135.21	6.90	135.21	6.90
3	27.7.b	1 (2)	5.56	5.56	14.45	2.24	2.24	16.69	5.56	11.29	0.66	11.29	0.66
3	27.7.h	2 (52)	239.30	-19.41	141.12	80.36	80.36	240.64	49.79	240.05	13.03	240.05	13.03
3	27.7.j	1 (29)	57.54	57.54	83.14	33.17	33.17	172.61	57.54	130.25	6.82	130.25	6.82
4	27.7.j	1 (6)	31.77	31.77	66.27	5.47	5.47	95.30	31.77	70.37	3.76	70.37	3.76

Cuadro 5.2: Intervalos de confianza por trimestre, subáreas 7 y metier según el método de diseño y el método Bootstrap (Enfoque en lances).



D<sup>a</sup> Lucía Cañas Ferreiro, Titulada Superior de Actividades Técnicas y Profesionales del Instituto Español de Oceanografía (IEO) destinada en el Centro Oceanográfico de A Coruña, y participante como experta independiente en el grupo del *Scientific, Technical and Economic Committee for Fisheries (STECF) Fisheries Dependent Information (FDI) EWG 25-05*, que tuvo lugar entre el 19 y el 23 de mayo de 2025

INFORMA:

Que D Joel Marcelo Morales, estudiante del "Máster Interuniversitario en Técnicas Estadísticas" de la Universidad de Santiago de Compostela, presentó en el EWG 25-05 un caso de estudio de especial relevancia sobre diferentes metodologías para el cálculo del coeficiente de variación de las estimas de descarte, las cuales proceden del Programa de Observación a Bordo gestionado por el IEO para el seguimiento de la actividad pesquera de la flota que opera en la subárea 7 del *International Council for the Exploration of the Sea (ICES)*. El estudio fue acogido con interés y quedó pendiente el posible uso del script, que contiene el procedimiento para realizar los cálculos, por parte del grupo, tras la defensa de su trabajo fin de máster.

En A Coruña, a tres de junio de dos mil veinticinco.

CAÑAS  
FERREIRO  
LUCIA - DNI  
32836922Y

Firmado  
digitalmente por  
CAÑAS FERREIRO  
LUCIA - DNI  
32836922Y  
Fecha: 2025.06.03  
14:20:26 +02'00'