



Universidade de Vigo

Trabajo Fin de Máster

---

# *Bootstrap* para series temporales en modelos de producción excedentaria

---

Julissa de la Rosa

Máster en Técnicas Estadísticas

Curso 2022-2023



# Propuesta de Trabajo Fin de Máster

<b>Título en galego:</b> Remostraxe para series temporais en modelos de produción excedentaria
<b>Título en español:</b> Remuestreo para series temporales en modelos de producción excedentaria
<b>English title:</b> Bootstrap for time series in surplus production models
<b>Modalidad:</b> Modalidad B
<b>Autor/a:</b> Julissa de la Rosa, Universidad de la Coruña
<b>Director/a:</b> Javier Roca Pardiñas, Universidad de Vigo
<b>Tutor/a:</b> Santiago Cerviño, Instituto Español de Oceanografía (IEO-CSIC); Marta Cousido Rocha, Instituto Español de Oceanografía (IEO-CSIC); M. Grazia Pennino, Instituto Español de Oceanografía (IEO-CSIC)
<p><b>Breve resumen del trabajo:</b></p> <p>La comprensión de cómo se evalúan las poblaciones de peces y cómo se recomienda sobre su explotación sostenible se ha convertido en una componente esencial de la gestión de los recursos pesqueros. Existe una gran variabilidad de métodos de evaluación en función de su complejidad y de su necesidad de datos. Una clase de métodos de evaluación ampliamente empleada es la basada en el concepto de producción excedentaria: producción de la población de peces tras extraer los que mueren tanto por mortalidad natural como por pesca. Dentro de los modelos de producción excedentaria, nuestro foco se encuentra en los modelos de producción excedentaria con biomasa conocida (denominados KBPM), los cuales se presentan como una alternativa a los clásicos modelos de producción excedentaria, MacCall (2002). Mientras los modelos clásicos estiman la producción excedentaria empleando una serie temporal de capturas y un índice de biomasa (o abundancia) de la población los KBPM emplean una serie de capturas y una serie de biomasa estimada previamente a través de otro modelo de evaluación. A pesar de la necesidad del ajuste previo de otro modelo de evaluación los KBPM tienen un gran número de aplicaciones, por ejemplo, en la evaluación de varias especies conjuntamente o en la inclusión de efectos ambientales en la evaluación, Bundy et al. (2012) y Mueter y Megrey (2006). Dentro del contexto de los modelos KBPM, nuestro objetivo es definir intervalos de confianza para las diferentes cantidades estimadas por el modelo permitiéndonos así hacer inferencia sobre las cantidades teóricas correspondientes. Para llevar a cabo esta tarea comenzaremos con la revisión de las técnicas Bootstrap para datos dependientes (series temporales) disponibles en la literatura y adaptables a nuestro contexto de trabajo. Tras la selección de los procesos <i>bootstrap</i> a aplicar, procederemos a su adaptación para la estimación de intervalos de confianza en los modelos KBPM. Una vez dispongamos de la metodología, comprobaremos su correcto funcionamiento en un entorno de simulación. Tras su validación, esta se implementará en un paquete de R llamado <b>knobi</b> que está actualmente siendo desarrollado por nuestro equipo y ya permite la estimación de modelos KBPM entre otras funciones.</p>
<b>Recomendaciones:</b> Interés por la validación de las metodologías en entornos simulados. Interés en aprender a crear paquetes de R. Dominio de R
<b>Otras observaciones:</b> Sí



Don Javier Roca Pardiñas, Catedrático de la Universidad de Vigo, don Santiago Cerviño, Científico Titular de Instituto Español de Oceanografía (IEO-CSIC), doña Marta Cousido Rocha, grupo profesional M3 de Instituto Español de Oceanografía (IEO-CSIC), y doña M. Grazia Pennino, Científica Titular de Instituto Español de Oceanografía (IEO-CSIC), informan que el Trabajo Fin de Máster titulado

*Bootstrap para series temporales en modelos de producción excedentaria*

fue realizado bajo su dirección por doña Julissa de la Rosa para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En A Coruña, a 05 de junio de 2023.

El director:  
Don Javier Roca Pardiñas

El tutor:  
Don Santiago Cerviño

La tutora:  
Doña Marta Cousido Rocha

*Julissa de la Rosa*

La autora:  
Doña Julissa de la Rosa

La tutora:  
Doña M. Grazia Pennino

---

**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/la autor/a 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

En primer lugar, quiero agradecer al equipo del Instituto Español de Oceanografía por brindarme la oportunidad de ser partícipe de este importante proyecto. A mis tutores/as Marta Cousido Rocha, Santiago Cerviño y Grazia Pennino por compartir sus conocimientos y su disponibilidad durante todo este proyecto. De igual forma quiero agradecer a Anxo Paz y David Nachón por su guía y acompañamiento durante la redacción de esta memoria. Por otro lado, quiero hacer especial mención a mi tutora Marta Cousido Rocha por su ayuda, paciencia y dedicación en este estudio, por compartir sus conocimientos y sus rigurosas correcciones. También quiero darle las gracias al director de este proyecto, Javier Roca Pardiñas por sus consejos y orientación.

Por último, pero no menos importante, quiero darle las gracias a mi madre Juana de la Rosa quien a pesar de la distancia siempre estuvo a mi lado para motivarme en los días más difíciles durante mis horas de estudio. A mis amigas y amigos por confiar y creer en mis expectativas.





# Índice general

<b>Resumen</b>	<b>XI</b>
<b>Prefacio</b>	<b>XIII</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Instituto Español de Oceanografía . . . . .	1
1.2. Planteamiento del problema . . . . .	1
1.3. Objetivo . . . . .	2
1.4. Método de resolución . . . . .	2
<b>2. Materiales y métodos</b>	<b>5</b>
2.1. Modelos de producción excedentaria (SPMs) . . . . .	5
2.1.1. Modelos de producción excedentaria con biomasa conocida (KBPMs) y knobi . . . . .	7
2.2. Intervalos de confianza (IC) . . . . .	9
2.2.1. <i>Bootstrap</i> . . . . .	9
2.2.2. Simulación Montecarlo . . . . .	13
<b>3. Resultados</b>	<b>15</b>
3.1. Definición de escenarios . . . . .	15
3.2. Escenario 1 . . . . .	18
3.3. Escenario 2 . . . . .	21
3.4. Escenario 3 . . . . .	24
<b>4. Conclusiones</b>	<b>27</b>
<b>Índice de figuras</b>	<b>29</b>
<b>Índice de tablas</b>	<b>31</b>
<b>Bibliografía</b>	<b>33</b>



# Resumen

## Resumen en español

Según la FAO (2018), una gran proporción (90 %) de las poblaciones de peces marinos del mundo se encuentran totalmente explotadas o sobreexplotadas, requiriendo medidas de gestión enfocadas a reducir la presión pesquera y conseguir una explotación al Máximo Rendimiento Sostenible (MSY; *Maximum Sustainable Yield*). Para ello, se han propuesto modelos matemáticos, conocidos como modelos de evaluación de *stocks*, para analizar y comprender el impacto de la pesca y los factores ambientales en las poblaciones de peces (*stocks*). En el caso de poblaciones con datos limitados, una clase de métodos de evaluación ampliamente utilizados se basa en el concepto de producción excedentaria, definiéndose esta como la suma de reclutamiento (peces jóvenes que entran en el área explotada) y el crecimiento de los individuos que ya forman parte de la población menos los que mueren de forma natural. Nuestra investigación se centra en los modelos de producción excedentaria con biomasa conocida (KBPMs; *Known Biomass Production Models*), introducidos por MacCall (2002). Estos son una alternativa a los modelos clásicos de producción excedentaria, que utilizan como datos de entrada una serie temporal de capturas y una serie de biomasa derivada de un modelo de evaluación más complejo previamente ajustado. Nuestro objetivo es cuantificar la incertidumbre en las estimaciones derivadas de los KBPMs proponiendo intervalos de confianza (IC) basados en procedimientos *bootstrap* para datos dependientes evaluando su desempeño en escenarios de simulación. La metodología propuesta proporciona IC's para cantidades cruciales en la gestión pesquera como la biomasa o la mortalidad por pesca asociada al MSY.

## English abstract

According to the FAO (2018), a large proportion (90 %) of the world's marine fish stocks are fully exploited or overexploited, requiring management measures focused on reducing fishing pressure and achieving exploitation at Maximum Sustainable Yield (MSY). To do this, mathematical models, known as stock assessment models, have been proposed to analyze and understand the impact of fishing and environmental factors on fish populations (*stocks*). In the case of data-limited stocks, one class of assessment methods widely used is based on the concept of surplus production, which is defined as the sum of recruitment (young fish entering the exploited area) and the growth of individuals already in the population minus those dying naturally. Our research focuses on the known biomass surplus production models (KBPMs), introduced by MacCall (2002). These models are an alternative to the classical surplus production models, which use as input data a catch time series and series of estimated biomass derived from a previously fitted more complex stock assessment model. Our objective is to quantify the uncertainty in the estimates of the parameters and quantities of interest derived from the KBPMs by proposing confidence intervals (CI) based on bootstrap procedures for dependent data, evaluating their performance in simulation scenarios. The proposed methodology provides CI's for crucial quantities in fisheries management, such as the biomass or the fishing mortality associated with MSY.



# Prefacio

El presente proyecto se ha realizado en colaboración con el Instituto Español de Oceanografía (IEO-CSIC) a través del convenio de colaboración con las tres universidades gallegas para la realización de las prácticas profesionales de los alumnos y alumnas del Máster en Técnicas Estadísticas. El IEO-CSIC, es un organismo público de investigación (OPI), dependiente del Ministerio de Ciencia e Innovación, dedicado a la investigación en ciencias del mar especialmente en lo relacionado con el conocimiento científico de los océanos, la sostenibilidad de los recursos pesqueros y el medio ambiente marino.

Para ello, una de las líneas de investigación dentro del IEO-CSIC se centra en el desarrollo de nuevos modelos matemáticos que permitan obtener información acerca del estado de un *stock* (población de peces); estos modelos se conocen como modelos de evaluación.

Dentro de esta línea, la propuesta para el presente trabajo se centra en construir intervalos de confianza para diferentes cantidades estimadas dentro de un modelo de evaluación denominado modelos de producción excedentaria con biomasa conocida, permitiendo hacer inferencia sobre las cantidades teóricas correspondientes.

Los intervalos de confianza sobre cantidades de interés nos permiten conocer el grado de incertidumbre asociado a dicha estimación, ayudándonos así en la obtención de conclusiones y en la toma de decisiones.

Para la creación de dichos intervalos de confianza en este trabajo hemos decidido emplear técnicas de remuestreo (*bootstrap*), con la dificultad añadida de que los datos de entrada a nuestro modelo de evaluación son series temporales y, por lo tanto, necesitaremos un proceso de remuestreo para datos dependientes. El procedimiento *bootstrap* para la construcción de intervalos de confianza lo validaremos mediante un estudio de simulación.

Este trabajo de fin de máster consta de una introducción, en la cual se describe al IEO, el marco y relevancia del problema, así como el objetivo principal y métodos de resolución, seguida del capítulo constituido por los de metodología donde se detallan los métodos de evaluación que se empleará en esta investigación junto a las técnicas *bootstrap* propuestas para la creación de los intervalos de confianza, así como el procedimiento de simulación empleado para validar dicha propuesta, seguidamente tenemos un tercer capítulo en el cual se presentan los resultados de los distintos escenarios y, finalmente, el cuarto capítulo que se corresponde con las conclusiones generales del trabajo. Luego se presenta la bibliografía consultada.



# Capítulo 1

## Introducción

En este capítulo se describe al Instituto Español de Oceanografía (IEO), además se expone el marco y la relevancia del problema resuelto. Luego, se presenta el objetivo general y, finalmente, se describe la perspectiva adoptada para solventar el problema, así como su motivación.

### 1.1. Instituto Español de Oceanografía

El instituto Español de Oceanografía es el principal organismo público en España dedicado a la investigación científica marina, en particular el conocimiento científico de los océanos, la sostenibilidad de las poblaciones de peces y el medio marino. Uno de los objetivos del IEO es prestar asesoramiento científico y tecnológico a las autoridades en materia marítima, además representa a España en organismos internacionales relacionadas con la oceanografía y la pesca por medio de su investigación y su punto de vista científico.

Entre sus funciones básicas se encuentran:

- Investigación científica de las ciencias oceánicas y marinas e investigación de diferentes aspectos del océano.
- Asesoramiento a funcionarios gubernamentales sobre política pesquera y marina en general.
- Representación de España en organismos internacionales de pesca y ciencias del mar.
- Promover la cooperación en investigación oceánica a nivel regional, nacional e internacional.
- Formar investigadores marinos y difundir los conocimientos oceanográficos.

El presente estudio se enmarca en mejorar la calidad de la evaluación pesquera dado que en la práctica se estarán construyendo intervalos de confianza de cantidades muy importantes en la gestión como la biomasa o la mortalidad por pesca asociada al  $MSY$ , es decir,  $B_{MSY}$  y  $F_{MSY}$  respectivamente.

### 1.2. Planteamiento del problema

En la actualidad, la pesca produce alteraciones en la edad, el tamaño, el número total de peces de una población, así como en su distribución espacial. Estas poblaciones marinas explotadas o sobreexplotadas necesitan ser recuperadas y atendidas, y por este motivo la ciencia pesquera ha buscado, por medio de uso de descripciones matemáticas y estadísticas comprender la dinámica de las poblaciones explotadas.

El rendimiento de una pesquería es producto de la combinación de los procedimientos que afectan a una población de manera intrínseca, como el reclutamiento (peces jóvenes que por primera vez se incorporan a la fase explotable de una población de peces), la mortalidad natural y el crecimiento, junto con factores externos que afectan a la dinámica poblacional, donde destaca la actividad pesquera. Por tanto, el objetivo es decidir si un esfuerzo o un nivel de captura particular es sostenible para un futuro determinado, teniendo en cuenta que diferentes especies tienen dinámicas específicas.

Los científicos pesqueros se han dedicado a comprender la dinámica de las poblaciones de peces utilizando modelos cuantitativos que intentan reproducir el comportamiento de las poblaciones, teniendo en cuenta la enorme dificultad que supone reproducir la realidad subyacente.

Para comprender dicha dinámica existe una gran variabilidad de modelos de evaluación en función de su complejidad y de los datos necesarios para poder emplearlos. Por ejemplo, cuando se tiene poca información se utilizan modelos de evaluación conocidos como modelos de datos limitados, que emplean información como distribución de tallas o series de capturas (peces capturados en operaciones pesqueras en unidades de peso). Sin embargo, también se disponen de modelos más complejos conocidos como modelos de datos ricos que incorporan la estructura por edad o por talla del pez (es decir, la información de entrada y salida del modelo se proporciona estructura por la talla o edad del pez, por ejemplo, capturas por edad o talla, mortalidad pesquera por edad o talla, etc.).

Dentro de los modelos de datos limitados, este trabajo se centra en los modelos de producción excedentaria (SPMs; *Surplus Production Models*), que pueden consultarse en Cousido-Rocha et.al (2022), y en cuantificar la incertidumbre asociada a cantidades de interés estimadas por los mismos (en el capítulo 2 se introducen en detalles estos modelos). Más concretamente, el trabajo se centra en unos modelos de producción excedentaria alternativos que surgen a partir de los modelos SPMs clásicos: los modelos de producción excedentaria con biomasa conocida (KBPMs; *Known Biomass Production Models*), MacCall (2002).

### 1.3. Objetivo

En modelos de producción excedentaria con biomasa conocida, el objetivo final es definir intervalos de confianza para las diferentes cantidades estimadas por el modelo, permitiendo así hacer inferencia sobre las cantidades teóricas correspondientes, siendo este objetivo la contribución en esta línea de investigación.

### 1.4. Método de resolución

El equipo del Instituto Español de Oceanografía está desarrollando un paquete de R llamado **knobi** creado para llevar a cabo diferentes análisis empleando KBPMs y actualmente busca complementarlo añadiendo intervalos de confianza a las estimaciones de estos modelos.

Pero, ¿cómo se construirán los intervalos de confianza?

Al tener un proceso de población dinámico, donde las variables que utiliza el modelo, capturas y biomasa, presenten una correlación a lo largo de los años, hace que la asunción de independencia presente en muchos procedimientos de remuestreo no se verifique, por lo que necesitamos trabajar con metodologías que tengan en cuenta dicha dependencia en la construcción de los intervalos de confianza.



Para llevar a la práctica este propósito se realiza una revisión de técnicas *bootstrap* disponibles en la literatura y adaptables a nuestro contexto de trabajo.

En la literatura nos encontramos con gran variedad de enfoques para datos dependientes. Entre ellos están el *AR-Sieve Bootstrap* (ARSB), véase Grenander (1981) y German y Hwang (1982) y el *Moving Block Bootstrap* (MBB), propuesto por Kunsch (1989) y Liu y Singh (1992); sin embargo, estos procesos están diseñados para series estacionarios y en nuestro caso las series de biomasa y capturas suelen presentar tendencias incumpliendo así dicha asunción de estacionariedad. Una opción que contemplamos fue eliminar la tendencia previamente a la aplicación del *bootstrap* sin embargo, los procesos que valoramos para la substracción de dicha tendencia no arrojaban un buen funcionamiento por lo que decidimos valorar otras opciones. Finalmente decidimos proponer un procedimiento *bootstrap* empleando Modelos Adictivos Generalizados (GAM; *Generalized Additive Models*) propuesto por Hastie y Tibshirani (1990) combinados con modelos autorregresivos (AR) para series temporales de la metodología Box-Jenkins, nombrado así en memoria de los estadísticos George E. P. Box y Gwilym Jenkins. Los detalles de la propuesta se proporcionan en el capítulo 2.

Luego de proponer e implementar la metodología antes mencionada, se comprueba su correcto funcionamiento en un entorno de simulación.

Tras la propuesta de los procedimientos *bootstrap* y su validación mediante un procedimiento de simulación de Montecarlo, el resultado del trabajo servirá para que el equipo de investigación del IEO-CSIC lo implemente en el paquete de R **knobi**.



## Capítulo 2

# Materiales y métodos

Este capítulo está dividido en diferentes secciones, la Sección 2.1 se utiliza para introducir los modelos de producción excedentaria (SPMs) y en particular los modelos de producción excedentaria con biomasa conocida (KBPMs) que vamos a abordar en este trabajo, mientras que la Sección 2.2 presenta la metodología utilizada para el cumplimiento del objetivo del trabajo, es decir, describe el procedimiento *bootstrap* para la construcción de los intervalos de confianza, así como el procedimiento de Montecarlo empleado para su validación.

### 2.1. Modelos de producción excedentaria (SPMs)

Según Haddon (2011), los modelos de producción excedentaria son el método analítico de evaluación más simple, dado que combina los efectos generales de reclutamiento, crecimiento y mortalidad en una sola función de producción. En efecto, la producción excedentaria (SP; *Surplus Production*) se define como la suma del reclutamiento y el crecimiento de los individuos que ya forman parte de la población menos los que mueren de forma natural.

Pero, ¿de dónde proviene la formulación de los modelos de producción excedentaria?

Russell (1931) describió el cambio en la biomasa (abundancia del *stock* en unidades de peso) de una población de peces explotada en términos de equilibrio entre reclutamiento, crecimiento y mortalidad por pesca por medio de la siguiente expresión algebraica:

$$B_{t+1} = B_t + (A_t + G_t) - (C_t + M_t) \quad (2.1)$$

donde  $B_t$  es la biomasa de la población (*stock*) en el año  $t$ ,  $A_t$ , es el peso total de los individuos que se reclutan para la población en el año  $t$ ,  $G_t$  es el crecimiento total en biomasa de los individuos ya reclutados anteriormente,  $C_t$  es el peso total de todos los peces capturados y  $M_t$  es el peso total de todos los peces que mueren por causas naturales durante el año. Russell a partir de esta expresión aclaraba que la biomasa de la población tiene ganancias por reclutamiento y crecimiento y pérdidas por mortalidad natural y por pesca.

La formulación de los modelos de producción excedentaria, la cual se deriva de la formulación verbal de Russell (1931), y tiene la siguiente estructura general:

$$B_{t+1} = B_t + f(B_t) - C_t \quad (2.2)$$

con

$$\hat{I}_t = \frac{C_t}{E_t} = qB_t \quad (2.3)$$

donde  $B_{t+1}$  es la biomasa explotable al principio del año  $t + 1$ ,  $B_t$  es la biomasa explotable al principio del año  $t$ ,  $f(B_t)$  es la producción de biomasa en función de la biomasa al inicio del año  $t$ ,  $C_t$  es la biomasa capturada durante el año  $t$ ,  $\hat{I}_t$  es el índice de biomasa relativa para el año  $t$ , y  $q$  es el coeficiente de capturabilidad.

La ecuación (2.2) que modela la dinámica de la población se conecta con las observaciones (índice de biomasa relativa) mediante la relación determinada por la ecuación (2.3) entre las capturas y la biomasa del *stock* a través del coeficiente de capturabilidad  $q$ . Cabe mencionar que el índice de biomasa relativa es una medida relativa de la biomasa de una población y se puede derivar de campañas oceanográficas o bien de información de capturas por unidad de esfuerzo en la pesca comercial.

Explicado esto, la función que describe la producción de biomasa en cualquier año puede tomar representaciones que se derivan de la ecuación logística. Entre las más usadas se encuentran:

La ecuación de Schaefer (1954):

$$f(B_t) = rB_t \left(1 - \frac{B_t}{K}\right) \quad (2.4)$$

donde los parámetros del modelo son  $r$ , que representa la tasa de crecimiento de la población de biomasa, y  $K$ , la biomasa máxima que el sistema puede alcanzar o (capacidad de carga del sistema).

El modelo (2.2) también deja patente la correlación que nos esperábamos encontrar en la práctica en las series temporales de capturas y biomasa.

La curva propuesta por Schaefer da lugar a una curva simétrica que alcanza su máxima producción en  $\frac{K}{2}$ . Por este motivo Pella y Tomlinson (1969) introducen el parámetro de asimetría  $p$  que modifica la función logística de la siguiente manera:

$$f(B_t) = \frac{r}{p} B_t \left[1 - \left(\frac{B_t}{K}\right)^p\right] \quad (2.5)$$

donde  $r$  es el parámetro de tasa de crecimiento (derivado de la tasa intrínseca de incremento natural), y  $K$  es el tamaño de la población para que el crecimiento sea positivo.

Recalcar que la propiedad clave del modelo Pella-Tomlinson es permite una amplia gama de curvas de producción asimétricas.

Por otro lado, otro enfoque popular es la modificación de Fox (1970):

$$f(B_t) = Ln(K)rB_t \left[ 1 - \left( \frac{Ln(B_t)}{Ln(K)} \right) \right] \quad (2.6)$$

cuando  $p$  tiende a cero se reduce a la formulación de Pella y Tomlinson (1969).

Explicado esto, a partir de los modelos de producción excedentaria se ha mencionado que se derivaron los modelos de producción excedentaria con biomasa conocida, denominados KBPMs (MacCall, 2002), los cuales se presentan como una alternativa a los clásicos SPMS. Los KBPMs se abordarán en la siguiente sub-sección, y son el método de evaluación utilizado en este trabajo.

### 2.1.1. Modelos de producción excedentaria con biomasa conocida (KBPMs) y knobi

Los modelos de producción excedentaria con biomasa conocida son una alternativa a los clásicos modelos de producción. Mientras los modelos clásicos estiman la  $SP$  empleando una serie temporal de capturas y un índice de biomasa de la población, los KBPMs emplean una serie de capturas ( $C$ ) y una serie de biomasa ( $B$ ) o biomasa reproductora ( $SSB$ ; *Spawning Stock Biomass*) estimada previamente a través de otro modelo de evaluación. Es decir, los KBPMs no requieren estimar el coeficiente de capturabilidad presentado en la ecuación (2.3), dado que utilizan, como datos de entrada la biomasa estimada mediante otros modelos de evaluación de poblaciones. Esto los diferencia de los SPMS tradicionales y reduce la imprecisión.

¿Por qué imprecisión? Cousido-Rocha et. al (2022) destaca que “*los modelos de producción excedentaria requieren la estimación del coeficiente de capturabilidad para relacionar el índice de biomasa relativa (ver ecuación 2.3) con la serie de biomasa (estimada)*”. Cousido-Rocha sostiene que este parámetro suele ser difícil de estimar por consiguiente implanta incertidumbre en otros parámetros estimados por los modelos de producción excedentaria.

MacCall (2002) explica que la base de los KBPMs es la idea de que la producción anual en una población no explotada es igual a  $B_{t+1} - B_t$ , y partiendo de esta idea, el cálculo de la  $SP$  de una población explotada depende de la captura; esto lo veremos en la ecuación (2.8) tras explicar la ecuación (2.7).

Los modelos de evaluación consideran la  $B$  o la  $SSB$  al comienzo del año, mientras que la  $C$  y la  $SP$  se refieren a todo el año. Esto lleva a relacionar la  $SP$  de un año con la media de la biomasa reproductora anual  $\bar{B}_t$ :

$$\bar{B}_t = (B_t + B_{t+1})/2 \quad (2.7)$$

Por tanto, en una población explotada, la  $SP$  anual se calcula como:

$$SP_t = \bar{B}_{t+1} - \bar{B}_t + C_t \quad (2.8)$$

para  $t = 1, \dots, N$ , siendo  $N$  el último año de la serie temporal de captura,  $\bar{B}_t$  es la media de la biomasa reproductora al inicio del año  $t$  y  $C_t$  la captura durante el año  $t$ . En este caso, usamos la

notación  $B$  para la biomasa, pero también se puede usar la biomasa reproductora  $SSB$ .

Una vez obtenida la  $SP_t$ , y teniendo en cuenta las ecuaciones (2.4) y (2.5), respectivamente, podemos ajustar la curva de producción. Tomando por ejemplo la generalización propuesta por Pella-Tomlinson:

$$SP_t = \frac{r}{p} \bar{B}_t \left( 1 - \left( \frac{\bar{B}_t}{K} \right)^p \right) \quad (2.9)$$

para  $t = 1, \dots, N$ , donde  $SP_t$  es la producción excedente,  $\bar{B}_t$  es la biomasa promediada,  $r$  es el parámetro de la tasa de crecimiento de la población,  $K$  es el tamaño máximo de la población y  $p$  es el parámetro de forma introducido por Pella y Tomlinson (1969) para permitir la asimetría en la curva de producción excedente.

El ajuste de los KBPMs se hará empleando el paquete `knobi` por medio de la función `knobi_fit`; puede consultarse el paquete `knobi` en el siguiente enlace <https://github.com/MERVEX-group/knobi>.

La función `knobi_fit` permite el ajuste de los KBPMs de manera automática, simplemente hay que suministrarle las series de captura observada y las estimaciones de biomasa o SSB derivadas del ajuste de un modelo de evaluación de la población rico en datos, y además posee un argumento que permite el ajuste considerando la modificación de Schaefer (2.4) o la de Pella-Tomlinson (2.5). En este estudio, el ajuste de la  $SP$  se realiza empleando la generalización de Pella-Tomlinson (ver ecuación 2.9).

Una vez ajustamos el modelo de los KBPMs en `knobi`, las estimaciones de las que estamos interesados en definir los intervalos de confianza para hacer inferencia sobre las cantidades teóricas son los parámetros  $r$ ,  $p$ ,  $K$ , más los puntos de referencia asociados al máximo rendimiento sostenible, los cuales definiremos en la próxima subsección.

### Puntos de referencia $MSY$

Los KBPMs se han utilizado para estimar los puntos de referencia asociados al máximo rendimiento sostenible. Kilduff et al. (2009) expone que los puntos de referencia en la pesca “*son herramientas que se utilizan para medir el estado de una población de peces o stock en relación con los objetivos de gestión*”. En otras palabras, con los puntos de referencia se pueden derivar conclusiones acerca de los niveles de explotación de la población de peces.

Los puntos de referencias se pueden obtener de acuerdo a las siguientes definiciones:

$MSY$  es la captura anual más alta posible que puede sostenerse a lo largo del tiempo manteniendo el *stock* en el nivel que produce el máximo crecimiento. La representación matemática de  $MSY$ , obtenida a partir de las ecuaciones 2.5 y 2.2, es:

$$MSY = rK \left( \frac{1}{p+1} \right)^{1+\frac{1}{p}} \quad (2.10)$$

Kilduff et al. (2009) define  $F_{MSY}$  como “*la tasa de mortalidad por pesca que dará como resultado que la biomasa de la población produzca el mayor rendimiento máximo a lo largo del tiempo, o el peso*”

de la captura dentro de un año”. La representación matemática de  $F_{MSY}$  es:

$$F_{MSY} = \left(\frac{r}{p}\right) \left(1 - \frac{1}{p+1}\right) \quad (2.11)$$

El término  $B_{MSY}$  puede consultarse en Kilduff et al. (2009) como “la biomasa promedio a largo plazo que se lograría si se pescara a una tasa de mortalidad por pesca constante igual a  $F_{MSY}$ ”. La representación matemática de  $B_{MSY}$  es:

$$B_{MSY} = K \left(\frac{1}{p+1}\right)^{\frac{1}{p}} \quad (2.12)$$

Por último, también se tiene el punto de referencia biológico  $F_{crash}$ , que se define como el nivel de pesca que llevará al *stock* a la extinción. Su expresión es la siguiente:

$$F_{crash} = \frac{r}{p} \quad (2.13)$$

Una vez explicada estos términos (parámetros de interés), en la siguiente sección se abordará la metodología implementada para la construcción de los intervalos de confianza en las estimaciones de los KBPMs.

## 2.2. Intervalos de confianza (IC)

Como se ha mencionado en el prefacio, la estimación puntual de un parámetro puede no ser suficiente dado que siempre habrá un margen de error a menos que se muestree toda la población. Por tanto, la solución es construir intervalos de confianza en los que el valor poblacional que deseamos estimar se encuentra entre un valor mínimo y máximo con una cierta probabilidad.

En ese sentido, en este trabajo se busca definir intervalos de confianza para hacer inferencia sobre los parámetros  $r$ ,  $p$ ,  $K$  y los puntos de referencias  $B_{MSY}$ ,  $F_{MSY}$  y  $MSY$  derivados del modelo KBPM.

En las siguientes secciones describiremos el *bootstrap* propuesto para el cálculo de los intervalos de confianza, así como el proceso de simulación de Montecarlo que nos ha permitido la validación del mismo.

### 2.2.1. *Bootstrap*

Uno de los primeros trabajos sobre *bootstrap* fue publicado por Bradley Efron en *Bootstrap methods: another look at the jackknife* (1979) y es un método de remuestreo que se utiliza para aproximar la distribución en el muestreo de un estadístico. A lo largo de los años se han propuesto multitud de versiones de esta técnica de remuestreo, sin embargo, muchas de ellas asumen la independencia de los datos presentes en la muestra, por lo que no son válidos en un contexto de series temporales como el del presente trabajo. Singh (1981) reconoció que si se aplica el *bootstrap* diseñado para datos independientes a datos que son dependientes, se elimina toda la información de dependencia presente en

los mismos. No obstante, la literatura *bootstrap* también proporciona procesos diseñados para datos dependientes. Los enfoques más populares son *Moving Block Bootstrap* (MBB) y *AR-Sieve Bootstrap* (ARSB).

Por un lado el *Moving Block Bootstrap* fue propuesto por Kunsch (1989) y Liu y Singh (1992) y trata de modelar la dependencia en la serie de tiempo remuestreando bloques independientes de observaciones consecutivas, este procedimiento preserva la estructura de dependencia de la serie de tiempo dentro de cada bloque.

El *AR-Sieve Bootstrap* sigue la idea básica de ajustar modelos paramétricos a la serie temporal y luego remuestrear los residuos de los mismos, que ya cumplen la asunción de independencia, véase Grenander (1981) y German y Hwang (1982).

Ambos enfoques fueron explorados para la realización de esta investigación, pero nos encontramos con ciertos inconvenientes. En primer lugar, son métodos para series estacionarias y en nuestro caso las series tienen tendencia, lo que llevaba a tener valorar diferentes opciones para poder eliminarla. Además de esto, incluso eliminado previamente la tendencia las muestras *bootstrap*, no se conseguía replicar el comportamiento de los datos. Por este motivo se ha optado por realizar un procedimiento *bootstrap* basado en Modelos Adictivos Generalizados (GAM; *Generalized Additive Models*; Hastie y Tibshirani, 1990). El procedimiento consiste en ajustar un GAM que permita captar la tendencia existente en nuestras series temporales de biomasa y capturas, y una vez tenemos modelada la tendencia comprobar si los residuos de este modelo son independientes, de ser así procedemos con un remuestreo aleatorio simple sobre los residuos al cual le sumaremos la tendencia ajustada para conseguir las muestras *bootstrap*. En el caso de residuos dependientes, ajustamos un modelo AR (*autoregressive model*) de series temporales sobre los residuos del GAM y aplicamos el remuestreo aleatorio simple a los residuos del modelo AR. Las remuestras finalmente serán los valores ajustados de GAM más los valores ajustados de AR más la remuestra *bootstrap* de los residuos. Es importante destacar que la metodología Box-Jenkins es una técnica clásica que permite el estudio de las series temporales, la cual considera únicamente valores del pasado para explicar el presente y el futuro de la serie.

### Bootstrap en Modelos Adictivos Generalizados (GAM)

Los Modelos Adictivos Generalizados, conocidos como modelos GAM, introducidos por Hastie y Tibshirani en 1990, son una extensión del modelo de regresión lineal tradicional que incorpora la no linealidad y la flexibilidad de los modelos no paramétricos.

Los GAM extienden el modelo lineal generalizado (GLM; *Generalized Linear Models*) reemplazando la forma lineal:

$$g(\mu(\mathbf{X})) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

donde  $\mathbf{X} = (X_1, \dots, X_p)$  son covariables,  $\mu(X) = E(Y|X)$  es la esperanza condicional de la variable respuesta  $Y$ ,  $g$  es el *link* (explicada más adelante) y  $\beta_0, \beta_1, \dots, \beta_p$  son los parámetros desconocidos, por la siguiente formulación:

$$g(\mu(\mathbf{X})) = \beta_0 + f_1(X_1) + \dots + f_p(X_p) \quad (2.14)$$



donde  $f_j(X_j)$  ( $j = 1, 2, 3, \dots$ ) es el efecto parcial suave desconocido de  $(X_j)$  en el predictor. La función *link g* permite considerar variables respuesta que siguen una distribución distinta de la gaussiana. En nuestro caso la respuesta es gaussiana y consecuentemente el *link* es la función identidad.

Partiendo de los modelos GAM, podemos aplicar el procedimiento *bootstrap*. Para ello, en primer lugar, se ajusta el modelo GAM a las series de captura y biomasa (de forma independiente) con relación al tiempo. En este caso, la variable tiempo  $(T_t)$  es una secuencia creciente que va desde 1 hasta el tamaño de la serie temporal,  $t = 1, 2, \dots, N$ , siendo  $N$  el número máximo de años considerados. El tiempo se incluye mediante una función suave (*spline*). Un *spline* es una función  $s(\cdot)$  definida a trozos sobre un rango de interés, mayormente son utilizados en aplicaciones que requieren un suavizado de curvas. La función `gam` de la librería `mgcv` en R permite el ajuste del modelo aditivo generalizado incorporando el término de suavizado no paramétrico mediante  $s$  por *smoothing splines*.

La ecuación del ajuste GAM a las capturas se representa de la siguiente forma:

$$(E(C_t|T_t)) = \beta_0 + f_t(T_t) \quad (2.15)$$

donde  $f_t(T_t)$  ( $t = 1, 2, \dots, N$ ) es el efecto parcial suave de tiempo  $(T_t)$  en el predictor de capturas.  $E(C_t|T_t) = \mu(T_t)$  es la esperanza condicional de la variable respuesta capturas  $(C_t)$ .

Por otro lado, la ecuación correspondiente al GAM de la serie de biomasa se representa de la siguiente forma:

$$(E(B_t|T_t)) = \beta_0 + f_t(T_t) \quad (2.16)$$

donde  $f_t(T_t)$  ( $t = 1, 2, \dots, N$ ) es el efecto parcial suave desconocido de tiempo  $(T_t)$  en el predictor de biomasa.  $(E(B_t|T_t) = \mu(T_t))$  es la esperanza condicional de la variable de respuesta biomasa  $(B_t)$ .

Lo que se busca con el ajuste GAM es obtener residuos independientes e idénticamente distribuidos (iid) para poder realizar el remuestreo aleatorio simple. Es por eso que una vez ajustado el modelo GAM, se analiza la correlación presente en sus residuos. Para ello, a través de la función `auto.arima` en R se calcula el orden  $p$  de un modelo autorregresivo (AR). En caso de obtener un  $p = 0$  se asume que los residuos son independientes, mientras que en otro caso se ajusta el  $AR(p)$  con el orden  $p$  obtenido.

Por lo tanto, si los residuos obtenidos a partir del modelo GAM son independientes se concluye que la función suave del tiempo capta toda la estructura de dependencia presente en las series temporales de captura o biomasa y se procede a realizar un muestreo aleatorio simple sobre los residuos, finalmente las muestras *bootstrap* se obtienen sumando la muestra *bootstrap* de los residuos y los valores ajustados del GAM. No obstante, cuando en los residuos se detecta autocorrelación, implica que existe un patrón sin modelar, es decir, el efecto suavizado del tiempo no consigue modelar toda la estructura de dependencia en las series temporales en cuestión. En ese caso los residuos ajustados por el modelo GAM poseen correlación y se procede a su ajuste mediante un modelo autorregresivo (AR) por medio de la metodología Box-Jenkins, en el cual la serie temporal  $\{\epsilon_j\}$  se representa como:

$$\epsilon_j = \phi_1\epsilon_{j-1} + \phi_2\epsilon_{j-2} + \dots + \phi_p\epsilon_{j-p} + a_j = \sum_{i=1}^p \phi_i\epsilon_{j-i} + a_j \quad (2.17)$$

donde  $\phi_1, \dots, \phi_p$  ( $\phi_p \neq 0$ ) son constantes,  $a_t$  es ruido blanco (una secuencia de variables aleatorias incorreladas con media nula y varianza finita  $\sigma_a^2$ ).

Para identificar el orden  $p$ , se utiliza la función `auto.arima` del paquete `forecast` en R, esta función devuelve el mejor modelo  $AR(p)$  de acuerdo con el AIC (Criterio de información de Akaike) de entre todos los modelos ajustados a los residuos del GAM.

Una vez introducidos los residuos del GAM en la función `auto.arima` y obtenido el modelo que da lugar al menor AIC, se realiza el ajuste del mismo mediante la función `Arima`. En este caso se utiliza la función `Arima` para realizar el ajuste del  $AR(p)$  obtenido por el criterio AIC de la función `auto.arima`.

Una vez ajustado el modelo  $AR$  se obtienen sus residuos que serán independientes y sobre los cuales podemos llevar a cabo el muestreo aleatorio simple. Finalmente las remuestras serán la suma de los valores ajustados por el modelo GAM más los valores ajustados por el  $AR$  más las remuestras de los residuos de este último.

### Resumen del *Bootstrap* final

En este caso vamos a suponer el remuestreo *bootstrap* para la serie de capturas (y análogo a la biomasa) de acuerdo a lo explicado anteriormente.

1. Consideramos el ajuste inicial GAM de la siguiente forma (ver ecuación 2.15):

$$(E(C_t|T_t)) = \beta_0 + f_t(T_t)$$

2. Calcular los residuos:

$$\epsilon_t = C_t - (\hat{\beta}_0 + \hat{f}_t(T_t))$$

3. En caso de no obtener residuos iid se procede a realizar el proceso autorregresivo (AR) sobre los residuos del GAM, en este caso:  $\epsilon_t$ :

$$\epsilon_t = \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \dots + \phi_p\epsilon_{t-p} + a_t$$

4. Estimar los coeficientes del proceso autorregresivo:

$$\hat{\phi}_1, \dots, \hat{\phi}_p$$

5. Calcular los residuos para los índices que sean posibles:

$$\hat{a}_t = \epsilon_t - \hat{\phi}_1\epsilon_{t-1} - \hat{\phi}_2\epsilon_{t-2} + \dots - \hat{\phi}_p\epsilon_{t-p}$$

para  $t = p+1, p+2, \dots, N$

6. Finalmente, se define las remuestras *bootstrap*:

$$C_t^* = (\hat{\beta}_0 + \hat{f}_t(T_t)) + (\hat{\phi}_1\epsilon_{t-1} + \hat{\phi}_2\epsilon_{t-2} + \dots + \hat{\phi}_p\epsilon_{t-p}) + \hat{a}_t^*$$

para  $t = 1, \dots, N$ , donde  $\hat{a}_t^*$  se corresponde con las muestras aleatorias simples calculadas a partir de los residuos del modelo AR. En caso de obtener residuos iid con el ajuste GAM, las remuestras *bootstrap* se definen de la siguiente forma:

$$C_t^* = (\hat{\beta}_0 + \hat{f}_t(T_t)) + \epsilon_t^*$$

para  $t = 1, \dots, N$  y donde  $\epsilon_t^*$  se corresponde con las muestras aleatorias simples calculadas a partir de los residuos del modelo GAM.

### 2.2.2. Simulación Montecarlo

En este estudio, la simulación Montecarlo consiste en considerar como dinámica real de la población un modelo SPMs (ecuación 2.2), y con dicha dinámica y valores de los parámetros de este modelo crear la biomasa y capturas de esta población simulada. Con respecto, a los niveles de explotación por pesca estos se definirán como veremos más adelante empleando los puntos de referencia  $B_{MSY}$ ,  $MSY$ ,  $F_{MSY}$  y  $F_{crash}$  estimados empleando las ecuaciones (2.12), (2.10), (2.11) y (2.13), respectivamente.

El primer paso es el cálculo de los puntos de referencia. Una vez tenemos los puntos de referencia relacionados con la mortalidad por pesca, i.e.  $F_{MSY}$  y  $F_{crash}$ , los empleamos para definir la trayectoria de explotación pesquera de nuestra población simulada (la cual denominamos  $F_t$ ; *Fishing mortality*). Esta trayectoria se explica más adelante en el capítulo 3, específicamente en la Sección (3.1), correspondiente a la definición de los escenarios de simulación. Cabe mencionar que esta trayectoria se crea buscando que permita que la biomasa disminuya desde cierto punto hasta alcanzar valores muy bajos para después permitiendo así obtener valores de capturas para un rango amplio de biomasa, desde un estado virgen, sin explotación, hasta unos valores próximos al colapso. Cubrir bien el rango de posibles valores de biomasa es crucial para un buen ajuste de los modelos de producción. Es importante destacar que en la construcción de esta trayectoria  $F_t$  se introducen diferentes niveles de incertidumbre en los datos simulados (esto se explica en detalle en el próximo capítulo).

Una vez definida la mortalidad por pesca, las capturas se definen de la siguiente manera:

$$C_t = F_t B_t \quad (2.18)$$

donde  $F_t$  es el vector de diferentes mortalidades por pesca y  $B_t$  es la biomasa. En este estudio se ha decidido iniciar la biomasa en el instante 0 en  $K$ .

Empezamos en el tiempo  $t = 0$ , y las capturas quedarían representadas de la siguiente forma:

$$C_0 = F_0 B_0 \quad (2.19)$$

luego, se calcula la biomasa en  $t = 1$ ,  $B_1$  empleando la ecuación general de un modelo SPM:

$$B_1 = B_0 + \frac{r}{p} B_0 \left( 1 - \left( \frac{B_0}{K} \right)^p \right) - C_0 \quad (2.20)$$

que se corresponde con la ecuación (2.5) explicada en la sección anterior.

Luego las capturas en el  $t = 1$ ,  $C_1$ , son definidas como:

$$C_1 = F_1 B_1, \quad (2.21)$$

y se procede al cálculo de la biomasa en  $t = 2$ ,  $B_2$ , representada así:

$$B_2 = B_1 + \frac{r}{p} B_1 \left( 1 - \left( \frac{B_1}{K} \right)^p \right) - C_1 \quad (2.22)$$

Y se sigue la secuencia hasta obtener  $B_t$  con  $t = 1, \dots, N$  siendo  $N$  el tamaño de la serie temporal (número máximo de años).

Una vez tenemos simuladas las capturas y la biomasa, el vector  $B_t$  coincide en el tiempo con las  $C_t$  y la  $F_t$ . Es muy habitual en los SPMs que, mientras que las capturas o la mortalidad por pesca se reportan a mediados de año, la biomasa se estima a principio de año. Por este motivo, los modelos KBPM implementados en **knobi** tienen en cuenta esta consideración y requieren como datos de entrada la biomasa a inicio de año y las capturas a lo largo del año. En nuestro caso, los valores de biomasa y capturas simulados se corresponden a mediados de año, esto implica que, a partir de los valores de biomasa a mitad de año debemos obtener la serie temporal a principio del mismo. Para ello, ajustamos un modelo GAM con respuesta la biomasa y variable explicativa los años mediante una función suave ó *spline*. Una vez ajustado este modelo, se utiliza para predecir la biomasa a inicio de año.

## Capítulo 3

# Resultados

En este capítulo se presentan los resultados de los intervalos de confianza obtenidos para las cantidades de interés en los KBPMs según distintos escenarios de incertidumbre en la simulación de las trayectorias de capturas y biomasa.

Antes de presentar los resultados, se describen los distintos escenarios creados en base a la metodología introducida en la Sub-sección 2.2.2 del capítulo anterior. El código desarrollado para abordar este trabajo se puede consultar en el siguiente enlace de GitHub: [https://github.com/julissadelarosa/Bootstrap\\_para\\_series\\_temporales](https://github.com/julissadelarosa/Bootstrap_para_series_temporales)

### 3.1. Definición de escenarios

En primer lugar los valores de los parámetros  $r$ ,  $p$  y  $K$  presentes en la curva de producción proceden del ajuste de SPMs en datos simulados mediante la dinámica por edades de una población, es decir, de una simulación más compleja donde el SPM no es empleado para simular los datos (ver Tabla 3.1).

Parámetro	Valor
$r$	0.4644
$p$	1.2
$K$	110.8215
$B_0$	110.8215

Tabla 3.1: Valores de los parámetros SPM para la construcción de la curva de producción.

Una vez tenemos los valores “teóricos” de los parámetros  $r$ ,  $p$  y  $K$  se pueden obtener los puntos de referencia teóricos  $B_{MSY}$ ,  $F_{MSY}$ ,  $F_{crash}$  y  $MSY$  según las ecuaciones (2.12), (2.11), (2.13) y (2.10), respectivamente, las cuales se han descrito en detalle en el capítulo anterior, específicamente en la

## Sub-sección 2.1.1.

Siguiendo las ecuaciones mencionadas, se obtienen los siguientes valores para los puntos de referencia mostrados en la Tabla 3.2.

Puntos de referencia	Valor
$B_{MSY}$	57.4476
$F_{MSY}$	0.2111
$MSY$	12.1269
$F_{crash}$	0.3870

Tabla 3.2: Valores de los puntos de referencia.

Es importante mencionar que de los puntos de referencia presentados en la Tabla 3.2, solamente estamos interesados en obtener los intervalos de confianza para los tres primeros  $B_{MSY}$ ,  $F_{MSY}$  y  $MSY$ , además de para los parámetros mostrados en la (Tabla 3.1), es decir,  $r$ ,  $p$  y  $K$ . Sin embargo, el punto de referencia  $F_{crash}$  es necesario para crear la trayectoria de mortalidad por pesca  $F_t$  que junto a la biomasa virgen ( $B_0$ ) se requiere para determinar la serie de capturas.

Cabe mencionar que la trayectoria de capturas y biomasa tendrá 100 años, es decir,  $N = 100$  siendo  $N$  el tamaño de la serie, y se crearan 200 réplicas de Montecarlo en cada escenario de simulación. Tras establecer esto, se crea la  $F_t$ , donde  $t = 1, \dots, N$ .

Con respecto a la trayectoria  $F_t, t = 1, \dots, N$  esta se define como se indica a continuación:

En primer lugar, el valor de mortalidad por pesca para cada año se obtiene generando valores de una distribución normal, para ello el primer paso es definir la media de dicha distribución en cada año. Iniciamos con una media 0 en  $t = 1$ , luego las medias para los siguientes 10 años proceden de la interpolación <sup>1</sup> realizada desde el valor  $0.25 * F_{MSY}$  hasta el  $0.5 * F_{MSY}$  con un total de 10 valores (cada uno se corresponde a un año). En los siguientes 19 años, la media de la normal es igual a  $0.5 * F_{MSY}$  y las medias para los siguientes 20 años vienen de la interpolación de los valores desde  $0.5 * F_{MSY}$  hasta  $0.9 * F_{crash}$ . Seguido de esto, la media para los siguientes 20 años es  $0.9 * F_{crash}$ , y los siguientes 20 años vienen de interpolar los datos desde  $0.9 * F_{crash}$  hasta  $0.5 * F_{MSY}$ . Por último los últimos 10 años tienen media  $F_{MSY}$  (véase Figura 3.1).

Una vez obtenemos la media de la distribución normal que se usará para generar los valores de mortalidad por pesca, la desviación típica de dicha distribución se deriva del coeficiente de variación  $CV$ , consideramos tres escenarios acordes a los diferentes niveles de variación, con un  $CV$  igual a 0.05, a 0.1 y a 0.15.

A partir de la trayectoria  $F_t$  cuyos valores anuales se han generado aleatoriamente de una distribución normal con las medias detalladas previamente se procede a la simulación de las capturas y la

<sup>1</sup>Los datos interpolados se crean con la función `approx` de R

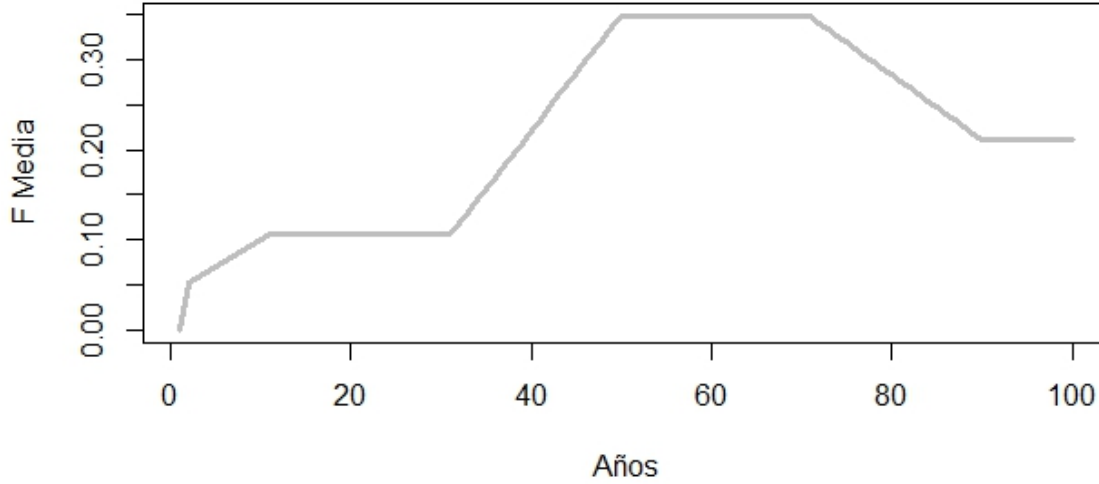


Figura 3.1: Media (para cada año) de la distribución normal que se emplea para generar los valores de mortalidad por pesca.

biomasa tal y como se ha explicado en la sección anterior (ver ecuación 2.18) para las capturas y la ecuación 2.20 para la biomasa. Cabe mencionar que como hemos indicado previamente la biomasa se simula a mediados de año y dado que los KBPMs trabajan con la biomasa a principio de año necesitamos ajustar un modelo GAM con variable respuesta la biomasa y predictor una función suave de los años para obtener predicciones biomasa a principio de año.

Una vez tenemos las 200 réplicas Montecarlo, a partir de cada una de estas 200 réplicas se realiza el *bootstrap*, es decir, para cada una de las réplicas de Montecarlo se crean 200 remuestras *bootstrap* para un total de 40.000 remuestras *bootstrap*. Es importante destacar que hacer el ajuste de los modelos KBPM para un número tan alto de remuestras *bootstrap* ha supuesto un reto a nivel de computación y ha requerido el uso de herramientas no estándares (empleo de servidores de computación).

Sobre las réplicas de Montecarlo y las remuestras *bootstrap* de capturas y biomasa, se ajusta la ecuación (2.9) a través del paquete `knobi` en R con la función `knobi_fit` y se obtienen las estimaciones de  $r$ ,  $p$ ,  $K$ ,  $B_{MSY}$ ,  $F_{MSY}$  y  $MSY$  para cada una de las remuestras *bootstrap* y réplicas de Montecarlo. Una vez se tienen las estimaciones del KBPM ajustado en cada muestra, se resume la información calculando medidas de tendencia central y dispersión, en concreto, media, mediana y desviación estándar. Igualmente se obtienen los percentiles del 0.025, y 0.975, es decir el intervalo de confianza deseado, mediante la función *quantile* la cual es una función base de R que proporciona cuantiles de acuerdo a la probabilidad asignada.

A continuación, se mostraran los resultados de cómo funcionaron los procedimientos en los distintos escenarios de incertidumbre (CVs). Hay que considerar que los resultados se sustentan en 200 réplicas de Montecarlo y 40.000 remuestras *bootstrap* para cada escenario.

### 3.2. Escenario 1

En este escenario se mostrarán los resultados de la simulación Montecarlo y del procedimiento *bootstrap* utilizando un  $CV = 0.05$ .

En la Figura 3.2 se pueden encontrar las réplicas por Montecarlo y las remuestras *bootstrap* para las series de capturas y biomasa.

En este caso, las remuestras *bootstrap* de la biomasa (gráfico derecha inferior) replican correctamente el comportamiento de las trayectorias de Montecarlo (gráfico derecha superior); algo similar a lo que ocurre en las remuestras *bootstrap* de capturas (gráfico izquierda inferior). A nivel gráfico, el remuestreo realizado parece tener un buen comportamiento ajusta tanto para la biomasa como para la serie de capturas a lo largo de los 100 años.

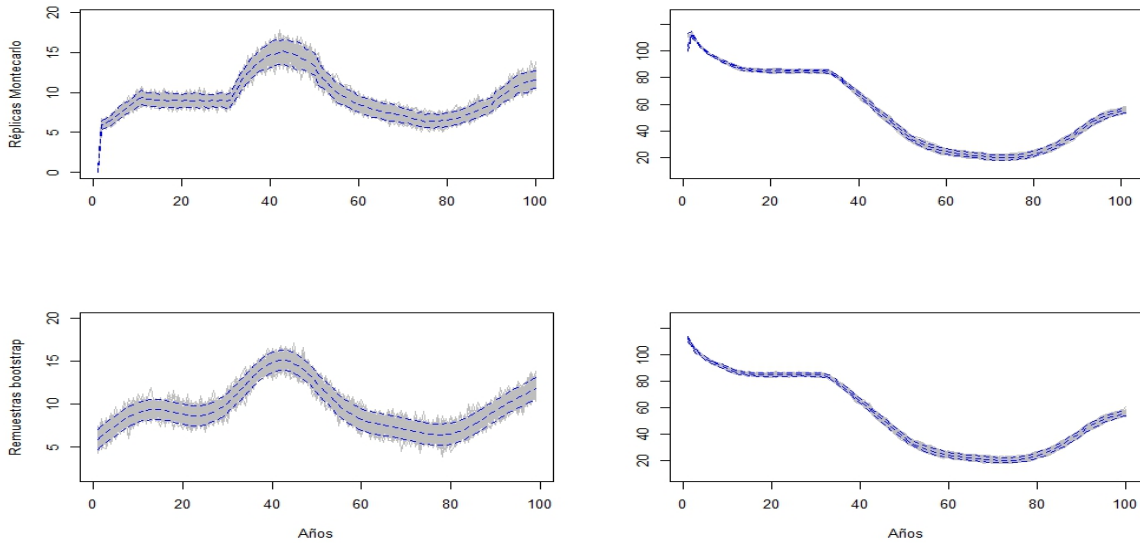


Figura 3.2: Réplicas Montecarlo (gráfico izquierda y derecha superior) y remuestras *bootstrap* (gráfico izquierda y derecha inferior), captura (izquierda) y biomasa (derecha) utilizando un coeficiente de variación igual a 0.05. Estas series, hasta el año  $N = 100$ , se corresponden con 200 réplicas Montecarlo, mientras que el *bootstrap* se corresponden con un total de 40.000 remuestras. Las líneas discontinuas azules son los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %. Además, la línea discontinua azul central representa la mediana.

Por otro lado, los resultados de los KBPMs aplicados a las réplicas de ambos procedimientos se encuentran en la (Tabla 3.3). En esta tabla, se abordan dos medidas de tendencia central: media y mediana, más una medida de dispersión: desviación estándar, que cuantifica la variabilidad de los datos.

En la Tabla 3.3 se puede notar que en general las medidas de tendencia central del Montecarlo y remuestras *bootstrap* son cercanas al valor real del parámetro, con una desviación típica asumible, por ejemplo, en todos los casos del Montecarlo es menos del 5 % del valor real del parámetro. Lo mismo ocurre en el procedimiento *bootstrap*, dado que en todos los casos la desviación es menos del 5 % del valor real, excluyendo el parámetro  $p$ , cuya desviación es del 7 % en relación al valor real.

En términos de media, el *bootstrap* está funcionando tan bien como el Montecarlo e incluso en algu-



nos casos mejor, por ejemplo, la media del parámetro  $K$  en el *bootstrap* es ligeramente más cercana al valor real que el valor medio de Montecarlo. En concreto, se ha comprobado que en todos los casos del procedimiento *bootstrap* no hay más del 2 % de diferencia entre el valor real y lo estimado en su media, mientras que en la simulación Montecarlo hay parámetros como el  $p$  que tiene al menos un 5 % de diferencia entre lo real y lo estimado. Lo mismo ocurre en términos de mediana, donde el procedimiento *bootstrap* en algunos casos también proporciona valores más cercanos a la realidad, por ejemplo, en los parámetros  $r$  y  $p$ . Para otros parámetros ambos procedimientos muestran un funcionamiento similar. Por ejemplo, para los puntos de referencia  $B_{MSY}$ ,  $MSY$  y  $F_{MSY}$  ambas medias y las medianas son similares entre sí y similares al valor real.

En términos de desviación típica el procedimiento *bootstrap* proporciona valores más altos que el Montecarlo. No obstante, a pesar de que es mayor que las obtenidas por Montecarlo, la diferencia no es elevada. Por ejemplo, la desviación típica del parámetro  $p$  (parámetro cuya desviación es del 7 % del valor real en el procedimiento *bootstrap*) no duplica a la obtenida por Montecarlo. Además, este comportamiento puede estar influenciado porque estamos trabajando con 40.000 remuestras y solo con 200 réplicas de Montecarlo, lo que implica que tenemos más variabilidad en *bootstrap*. Es decir, si en la simulación Montecarlo estuviéramos trabajando con 40.000 réplicas, podríamos obtener unas desviaciones similares, e igual su media y mediana serían más cercanas al valor real del parámetro. En ese sentido, la decisión de llevar a cabo las 200 réplicas Montecarlo y 40.000 remuestras *bootstrap* (200 por cada réplica de Montecarlo), se debe a que asumir como datos observados una sola de las réplicas de Montecarlo puede llevar a resultados sesgados sobre el funcionamiento del procedimiento *bootstrap*. Por este motivo, se tienen en cuenta cada una de las réplicas de Montecarlo a la hora de hacer el remuestreo.

En conclusión, la técnica *bootstrap* utilizando un  $CV=0.05$ , posee estimaciones de los parámetros y puntos de referencia KBPMs muy similares a las derivadas del proceso de Montecarlo y también a los valores reales de los mismos.

Cantidades Teóricas		Simulación Montecarlo			Remuestras <i>Bootstrap</i>		
Parámetro	Real	Media	Mediana	Desviación	Media	Mediana	Desviación
$r$	0.4644	0.4770	0.4819	0.0132	0.4659	0.4666	0.0191
$p$	1.2	1.2629	1.284	0.0560	1.2145	1.2163	0.0848
$K$	110.8215	110.2356	109.9	0.7688	110.8395	110.7	1.0828
$B_{MSY}$	57.4476	57.7306	57.78	0.1292	57.5731	57.58	0.4108
$F_{MSY}$	0.2111	0.2107	0.211	0.0006	0.2104	0.2104	0.0015
$MSY$	12.1269	12.1688	12.19	0.0639	12.1124	12.11	0.1001

Tabla 3.3: Medidas de tendencia central y dispersión de los resultados por simulación Montecarlo y remuestras *bootstrap* con el escenario de incertidumbre ( $CV=0.05$ ).

La Tabla 3.4 proporciona los intervalos de confianza del 95 %, i.e. con una probabilidad del 95 %

de contener al verdadero valor del parámetro. Estos IC nos permiten hacer inferencia sobre nuestras cantidades de interés en los modelos KBPM, es decir,  $r$ ,  $p$ ,  $K$ ,  $F_{MSY}$ ,  $B_{MSY}$  y  $MSY$ .

Cantidades Teóricas		Simulación Montecarlo		Remuestras <i>Bootstrap</i>	
Parámetro	Real	2.5 %	97.5 %	2.5 %	97.5 %
$r$	0.4644	0.4551	0.4946	0.4284	0.5020
$p$	1.2	1.1704	1.3387	1.0475	1.3756
$K$	110.8215	109.2462	111.5128	108.9801	113.1319
$B_{MSY}$	57.4476	57.5085	57.9233	56.7621	58.3626
$F_{MSY}$	0.2111	0.2096	0.2116	0.2074	0.2133
$MSY$	12.1269	12.0586	12.2542	11.9133	12.3018

Tabla 3.4: Intervalos de confianza de las estimaciones en KBPMs por simulación Montecarlo y remuestreo *bootstrap*,  $CV = 0.05$ . Se ajustan a los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %.

En la Tabla 3.4 se observa que todos los intervalos contruidos con el procedimiento *bootstrap* para los diferentes parámetros contienen el valor real. No obstante, no todos los intervalos contruidos con el método de Montecarlo contienen el valor real. En concreto, el intervalo de Montecarlo no contiene al valor real del punto de referencia  $B_{MSY}$ . El *bootstrap* obtiene una estimación más adecuada del punto de referencia  $B_{MSY}$ . En la Tabla 3.3 podemos observar como en la simulación de Montecarlo obtenemos valores ligeramente por encima del valor real de  $B_{MSY}$ . El mismo comportamiento lo observamos en la Tabla 3.4, donde el extremo inferior del IC es superior al valor real. Hay que tener en cuenta que el buen funcionamiento de los IC tanto para Montecarlo como para *bootstrap* no sólo dependen del correcto funcionamiento de estos procedimientos, si no también del correcto funcionamiento del ajuste del KBPM. En concreto, existen parámetros como  $p$  que son difíciles de estimar aún en escenarios de baja incertidumbre, necesitando para una correcta estimación disponer de valores de biomasa cubriendo el rango de biomosas posibles para este *stock*, i.e. desde el tamaño máximo poblacional hasta una biomasa próxima al colapso, ya que de esta manera tendremos observaciones suficientes para definir correctamente el máximo de la curva de producción. En el caso del *bootstrap*, el aumento de réplicas con respecto a Montecarlo o la variabilidad introducida en las trayectorias por la propia naturaleza del proceso nos permiten tener un IC que contiene el valor real del parámetro, así como una media y medianas más próximas a este.

En la Tabla 3.4 también podemos observar que los intervalos *bootstrap* son ligeramente más amplios que los de Montecarlo para todos los parámetros. Este hecho va acorde a lo explicado anteriormente. A pesar de tener una amplitud mayor, cabe mencionar que dicha amplitud no es mayor de un 15 % del valor del parámetro excepto para  $p$  que casi alcanza el 30 %. Este resultado en el parámetro  $p$  va acorde a lo discutido anteriormente respecto a su desviación típica.

A modo de resumen se puede decir que el remuestreo *bootstrap* logró el objetivo deseado, pro-

porcionando intervalos de confianza adecuados para los valores reales. En este escenario con poca variabilidad, hemos observado además que los parámetros reales que definen a la dinámica poblacional se encuentran muy cercanos a la mediana y la media de sus remuestras *bootstrap*.

### 3.3. Escenario 2

En este escenario se mostrarán los resultados de la simulación Montecarlo y remuestras *bootstrap* utilizando un  $CV = 0.1$  analizando así el funcionamiento de los procedimientos ante el incremento de la incertidumbre.

En la Figura 3.3 se pueden observar las réplicas por Montecarlo y remuestras *bootstrap* para las series de capturas y biomasa.

Cuando se observa la Figura (3.3), las réplicas Montecarlo de capturas y biomasa poseen mayor variabilidad en sus datos en comparación con las obtenidas en la Figura (3.2) del escenario anterior, como era de esperar. En el caso de las remuestras *bootstrap* ocurre lo mismo, el procedimiento *bootstrap* parece adaptarse correctamente y muestra un patrón similar al reflejado por Montecarlo.

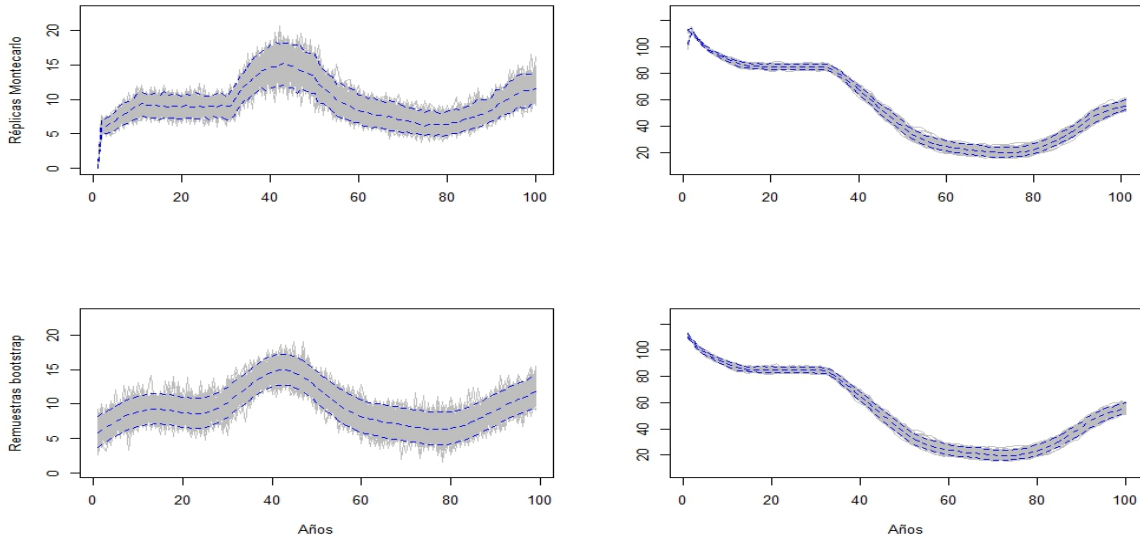


Figura 3.3: Réplicas Montecarlo (gráfico izquierda y derecha superior) y remuestras *bootstrap* (gráfico izquierda y derecha inferior), captura (izquierda) y biomasa (derecha) utilizando coeficiente de variación igual a 0.1. Estas series, hasta el año  $N = 100$ , se corresponden con 200 réplicas Montecarlo. Mientras que el *bootstrap* se corresponden con un total de 40.000 remuestras. Las líneas discontinuas azules son los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %. Además, la línea discontinua azul central representa la mediana.

La Tabla (3.5) muestra las estadísticas descriptivas de las estimaciones de los parámetros de interés de ambos procedimientos. El aumento de la incertidumbre implica que la media y la mediana de las réplicas de Montecarlo como *bootstrap* no sean tan cercanas al valor real como en el escenario previo. Sin embargo, aún así se puede concluir que ambos procesos siguen proporcionando estimaciones cercanas a

la realidad simulada. Por ejemplo, en el escenario anterior para todas las medias no hubo diferencia de más del 2 % entre el valor real y lo estimado por KBPM cuando utiliza las remuestras *bootstrap*, y en este escenario tampoco hay diferencias de más del 2 %, con la excepción del parámetro  $r$  (que posee una diferencia del 4 %) y del parámetro  $p$  (con una diferencia del 7 % entre el valor real y lo estimado). Esto hace indicar que el *bootstrap* consigue un buen funcionamiento incluso al aumentar la incertidumbre. La simulación Montecarlo en este escenario no muestra grandes diferencias con el escenario anterior en relación a los valores medios, dado que para todos los parámetros no hay diferencia mayor del 3 % entre el valor real y lo estimado por KBPM. Incluso parece que en este escenario el valor de la media del parámetro  $p$  se encuentra mejor estimada, ya que en el anterior escenario la diferencia de este parámetro con el valor real fue de 5 %. Estos resultados que describen a la media, también se sustentan con la mediana, ya que los porcentajes de diferencia en ambas medidas de tendencia central se ha comprobado que son iguales en ambos casos.

Cantidades Teóricas		Simulación Montecarlo			Remuestras Bootstrap		
Parámetro	Real	Media	Mediana	Desviación	Media	Mediana	Desviación
$r$	0.4644	0.4585	0.4569	0.0077	0.4421	0.4420	0.0258
$p$	1.2	1.1855	1.179	0.0328	1.1136	1.1130	0.1186
$K$	110.8215	111.3440	111.4	0.4367	112.3713	112.3	1.5179
$B_{MSY}$	57.4476	57.5734	57.56	0.0849	57.3345	57.34	0.7467
$F_{MSY}$	0.2111	0.2098	0.2097	0.0003	0.2091	0.2092	0.0026
$MSY$	12.1269	12.0783	12.07	0.0383	11.9899	11.99	0.1479

Tabla 3.5: Medidas de tendencia central y dispersión de los resultados por simulación Montecarlo y remuestras *bootstrap* con el escenario de incertidumbre ( $CV=0.1$ ).

A pesar que la simulación Montecarlo en este escenario posee mejor estimación que el procedimiento *bootstrap* en algunos parámetros de las medidas de tendencia central, hay casos que la estimación KBPM por medio de las remuestras *bootstrap* es mejor, como ocurre por ejemplo con el punto de referencia  $B_{MSY}$  (Tabla 3.5). De igual modo, aunque en este escenario haya menos homogeneidad en los datos, el *bootstrap* parece captar bien la dinámica simulada.

En términos de desviación típica con el procedimiento *bootstrap*, como es de esperar, se obtienen desviaciones más grandes que las obtenidas en el escenario anterior. Por ejemplo, el parámetro  $p$  obtuvo una desviación del 7 % del valor real (la mayor desviación), mientras que en este escenario posee una desviación del 9 % del valor real del parámetro. Esto implica que la incertidumbre no afectó de manera tan relevante, debido que se observa una diferencia suave.

Los intervalos con un nivel de confianza del 95 % obtenidos en este escenario se puede observar en la Tabla 3.6. En este escenario no todos los intervalos obtenidos por las estimaciones de KBPMs utilizando las réplicas Montecarlo contienen el verdadero valor del parámetro. Este es el caso de los

Cantidades Teóricas		Simulación Montecarlo		Remuestras Bootstrap	
Parámetro	Real	2.5 %	97.5 %	2.5 %	97.5 %
$r$	0.4644	0.4527	0.4864	0.3914	0.4929
$p$	1.2	1.1613	1.3077	0.8825	1.3467
$K$	110.8215	109.7086	111.6732	109.6643	115.6201
$B_{MSY}$	57.4476	57.4869	57.8732	55.8369	58.7658
$F_{MSY}$	0.2111	0.2093	0.2109	0.2040	0.2142
$MSY$	12.1269	12.0420	12.2019	11.6924	12.2743

Tabla 3.6: Intervalos de confianza de las estimaciones en KBPMs por simulación Montecarlo y remuestreo *bootstrap*,  $CV = 0.1$ . Se ajustan a los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %.

puntos de referencia  $B_{MSY}$  y  $F_{MSY}$ . El punto de referencia  $F_{MSY}$  posee estimaciones de tendencia central inferiores al verdadero parámetro (Tabla 3.5), y de igual forma se puede observar que el punto de referencia  $F_{MSY}$  posee sesgo a la derecha de los intervalos de confianza construidos por simulación Montecarlo, dado que el parámetro real se posicionó a la derecha del valor superior del intervalo. En ese sentido, a medida que aumenta la incertidumbre, las estimaciones realizadas por el modelo KBPM son más difíciles de estimar para el proceso de Montecarlo. Sin embargo, debido a la gran cantidad de remuestras *bootstrap*, cuando se realizan las estimaciones del modelo KBPM, todos los intervalos contienen el verdadero valor del parámetro en este último procedimiento.

No obstante, hay que destacar que, debido a la incertidumbre introducida con el coeficiente de variación, todos los intervalos por remuestras *bootstrap* se han ampliado ligeramente en relación al escenario anterior. En este caso la amplitud es mayor de un 20 % para el parámetro  $r$ , y para el parámetro  $p$  casi alcanza el 40 %. Los demás parámetros poseen una amplitud relativamente pequeña, es decir es de menos del 5 % del valor del parámetro. Ambos parámetros ( $r$  y  $p$ ) son los que poseen mayor variabilidad en la estimación realizada por KBPM en las remuestras *bootstrap*, ya que ambos poseen una desviación de entre el 5 % y el 10 % en relación al valor real del parámetro. No obstante, queda demostrado que, pese a introducir una mayor incertidumbre por medio del coeficiente de variación a los datos, el procedimiento del remuestreo *bootstrap* hace un buen ajuste aún sin conocer la dinámica real y permite que el intervalo contenga al valor teórico.

La amplitud de los intervalos en la simulación Montecarlo son más pequeños que las obtenidas por el procedimiento *bootstrap* dado que no sobrepasan el 7 % del valor real aunque en algunos casos puede llegar al 12 % como ocurre con el parámetro  $p$ . Sin embargo, en algunos casos es tan pequeña la amplitud del intervalo que no permite contener el verdadero valor del parámetro. Por ejemplo, la amplitud del intervalo de  $B_{MSY}$  y  $F_{MSY}$  es de menor al 1 %, lo que implica que, a pesar de que el procedimiento de Montecarlo simula los datos de forma correcta, se le hace difícil al modelo KBPM estimar de forma más precisa.

En resumen, el remuestreo *bootstrap* logró su objetivo, pero con intervalos más grandes, lo que indica que la precisión es menor. En las gráficas de este escenario se pudo observar que al aumentar la variabilidad en relación al escenario 1, los intervalos también son más amplios y menos exactos que en el dicho escenario.

### 3.4. Escenario 3

En este escenario se mostrarán los resultados de la simulación Montecarlo y remuestras *bootstrap* utilizando un  $CV = 0.15$  en sus datos.

En primer lugar, las réplicas Montecarlo y las remuestras *bootstrap* se muestran en la Figura 3.4, donde podemos apreciar el aumento de variabilidad en relación a los escenarios anteriores.

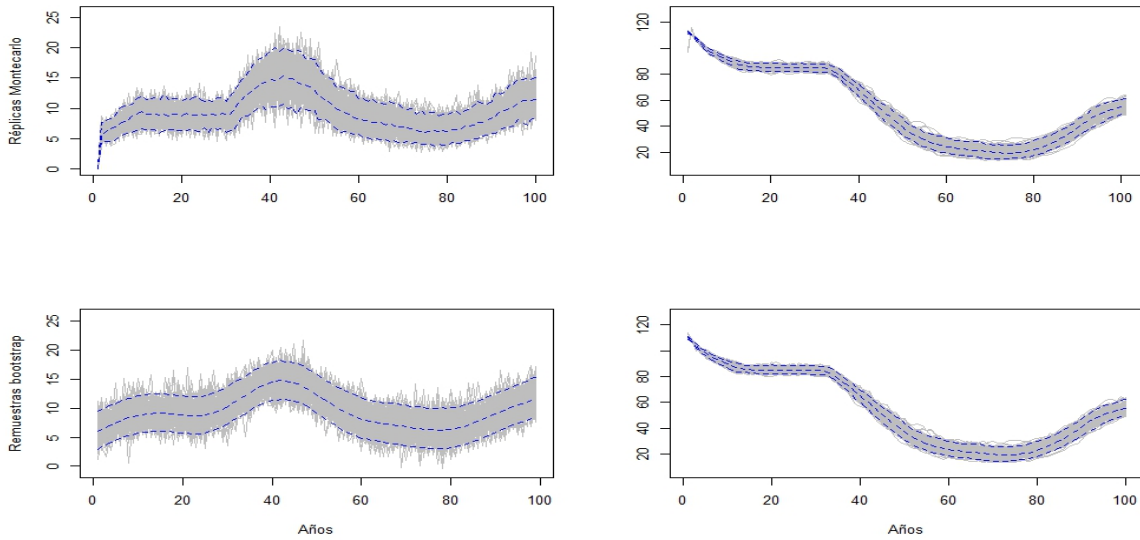


Figura 3.4: Réplicas Montecarlo (gráfico izquierda y derecha superior) y remuestras *bootstrap* (gráfico izquierda y derecha inferior), captura (izquierda) y biomasa (derecha) utilizando coeficiente de variación igual a 0.15. Estas series, hasta el año  $N = 100$ , se corresponden con 200 réplicas Montecarlo, mientras que el *bootstrap* se corresponden con un total de 40.000 remuestras. Las líneas discontinuas azules son los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %. Además, la línea discontinua azul central representa la mediana.

En la Figura 3.4 se observa en mayor medida como aumenta la variabilidad en las réplicas Montecarlo de capturas y biomasa en comparación con los escenarios previos y consecuentemente las remuestras *bootstrap*, lo que indicaría que los intervalos de confianza por ambos procedimientos deberían ser mucho más amplios.

En la Tabla 3.7 se presentan las medidas de tendencia central y dispersión de las estimaciones de los KBPM obtenidas por simulación Montecarlo y *bootstrap*.

En los resultados mostrados en la Tabla 3.7, la media y la mediana por simulación Montecarlo se pueden considerar próximos al valor del parámetro real.

Cantidades Teóricas		Simulación Montecarlo			Remuestras Bootstrap		
Parámetro	Real	Media	Mediana	Desviación	Media	Mediana	Desviación
$r$	0.4644	0.4575	0.4570	0.0057	0.4320	0.4322	0.0379
$p$	1.2	1.1830	1.181	0.0242	1.0716	1.0716	0.1750
$K$	110.8215	111.4338	111.5	0.3117	113.0958	112.9	2.3684
$B_{MSY}$	57.4476	57.5963	57.59	0.0758	57.2027	57.22	1.1305
$F_{MSY}$	0.2111	0.2096	0.2096	0.0004	0.2085	0.2086	0.0039
$MSY$	12.1269	12.0706	12.07	0.0341	11.9261	11.93	0.2210

Tabla 3.7: Medidas de tendencia central y dispersión de los resultados por simulación Montecarlo y remuestras *bootstrap* con el escenario de incertidumbre ( $CV=0.15$ ).

Concretamente, en este escenario las medidas de tendencia central como mucho muestran un 2 % de diferencia entre lo real y lo estimado. Este patrón parece repetirse en los tres escenarios, lo que implica que las medidas de tendencia central no se vean tan afectadas por las estimaciones KBPM en sus medias y medianas. Los resultados obtenidos a partir de las réplicas *bootstrap* son ligeramente menos precisas que los obtenidos por Montecarlo. Aunque por ejemplo, para el valor medio y mediana obtenidos de las réplicas *bootstrap* como mucho hay un 2 % de diferencia entre lo real y lo estimado, similar a los resultados obtenidos para Montecarlo. No obstante, las estimaciones por el KBPM por medio de las remuestras *bootstrap* también se obtienen diferencia del 7 % para el parámetro  $r$  y del 10 % para el parámetro  $p$ , de alguna manera estos parámetros son los que se ven más influenciados en las medidas de tendencia central por el procedimiento *bootstrap* y su consecuente estimación en **knobi** de todos los escenarios, donde mientras más incertidumbre mayor diferencia entre el valor real, como cabía esperar.

Por otro lado, en términos de desviación estándar con el procedimiento *bootstrap* se obtiene un mayor dispersión de lo que obtenemos en Montecarlo y también más dispersión que los escenarios anteriores. Concretamente, el procedimiento *bootstrap* en todos los casos de este escenario, la desviación típica es de menos del 3 % del valor real del parámetro, con excepción del parámetro  $r$  que posee una desviación del 8 % del valor real y el parámetro  $p$  que en este caso su desviación es del casi 15 % del valor real. Este último parámetro  $p$ , en este caso duplica la desviación del escenario 1, que le corresponde un 7 % del valor real y de un 9 % para el escenario 2. Al parecer, el parámetro de asimetría  $p$  es el que está teniendo más dificultad de estimar por los modelos KBPM. Seguido del parámetro  $r$  que en el primer escenario su desviación es de menos del 3 %, en el segundo escenario es del 5 % y en este último escenario que es del 8 % del valor real. Por otro lado, a pesar de la incertidumbre añadida en este escenario, con simulación Montecarlo en todos los casos la desviación típica es de menos del 3 % del valor real del parámetro. Lo que implica que las estimaciones en KBPM utilizando las réplicas Montecarlo consigue tener estimaciones similares al valor real tomando como referencia la dispersión.

La Tabla 3.8 posee los intervalos con un nivel de confianza del 95 %. En este caso la simulación Mon-

tecarlo logra contener menos parámetros reales dentro de sus intervalos, como es el caso del parámetro  $K$ ,  $B_{MSY}$  y  $F_{MSY}$ . Por ejemplo, para el parámetro  $K$  y el punto de referencia  $B_{MSY}$ , según la Tabla 3.7, su valor real está posicionado a la izquierda de su media y mediana, lo que implica que el valor real presenta un sesgo a la izquierda del valor inferior del intervalo de confianza. Lo mismo ocurre con  $F_{MSY}$  pero en este caso el valor real de este punto de referencia tiene sesgo a la derecha con respecto a las estimaciones KBPMs de tendencia central e intervalos de confianza. A pesar que las estimaciones KBPMs por Montecarlo con la media, mediana y desviación típica es más precisa debido que no tiene tanta diferencia con el valor real y en comparación con el procedimiento *bootstrap*, hace que al tener mayor variabilidad y no tantas réplicas como remuestras *bootstrap*, hace que las estimaciones KBPM sean menos precisas y por ende menos acertadas a la hora de estimación para fines de construcción de intervalos de confianza. Dado que con el procedimiento *bootstrap* aún teniendo la misma incertidumbre asociada y sin conocer la dinámica real la estimación KBPM consigue contener los verdaderos valores de los parámetros en los intervalos de confianza.

Esta variabilidad mencionada y la cantidad de remuestras *bootstrap* produce que la amplitud de sus intervalos sea más grande que los escenarios anteriores. En concreto, para todos los parámetros la amplitud no es mayor del 10 %, sin embargo, para el parámetro  $r$  la amplitud es del 32 % del verdadero valor del parámetro, la cual en el escenario anterior fue de 20 %, mientras que el parámetro  $p$  en este escenario posee una amplitud del 57 % la cual en el escenario anterior fue del 40 % del valor del parámetro.

Cantidades Teóricas		Simulación Montecarlo		Remuestras Bootstrap	
Parámetro	Real	2.5 %	97.5 %	2.5 %	97.5 %
$r$	0.4644	0.4570	0.4644	0.3568	0.5055
$p$	1.2	1.1537	1.2129	0.7262	1.4140
$K$	110.8215	111.1187	111.8193	109.1499	118.4302
$B_{MSY}$	57.4476	57.4889	57.7339	54.9011	59.3438
$F_{MSY}$	0.2111	0.2089	0.2104	0.2007	0.2161
$MSY$	12.1269	12.0242	12.1292	11.4744	12.3424

Tabla 3.8: Intervalos de confianza de las estimaciones en KBPMs por simulación Montecarlo y remuestreo *bootstrap*,  $Cv = 0.15$ . Se ajustan a los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %.

A modo de cierre se puede argumentar que con incertidumbre asociada a un CV igual a 0.15, por remuestras *bootstrap* se obtuvo intervalos más grandes que en los escenarios anteriores, pero en casos como el  $r$  y el  $p$  la incertidumbre asociada produce que la amplitud del intervalo sea más o menos el 50 % del valor del parámetro, específicamente del 32 % ( $r$ ) y 57 % ( $p$ ). Esto implica que el intervalo contiene el verdadero valor del parámetro, pero menos exactitud a la hora de identificar el valor real.



## Capítulo 4

# Conclusiones

En definitiva, en el presente proyecto se expuso el objetivo del trabajo tras una introducción de la evaluación de recursos pesqueros. A continuación, se explicaron en detalle los KBPMs en los que se centra el objetivo del trabajo, una vez abordados estos modelos hemos propuesto un procedimiento *bootstrap* para series temporales el cual hemos evaluado su funcionamiento mediante simulación. El procedimiento *bootstrap* nos ha permitido crear intervalos de confianza para las cantidades de interés estimadas en los KBPM cumpliendo así el objetivo y proporcionando al IEO-CSIC una metodología para implementar en el paquete **knobi** complementado la metodología ya implementada.

Entre las metodologías *bootstrap* para datos dependientes exploradas estuvieron el *Ar-Sieve Bootstrap* y el *Moving Block Bootstrap* en los cuales no se obtuvieron los resultados deseados dado a que son métodos para series estacionarias y en nuestro caso las series tienen tendencia, lo que implicaba valorar distintas opciones para eliminar la tendencia, que incluso eliminando previamente esta tendencia, las muestras *bootstrap* no conseguían replicar el comportamiento de los datos. Es por este motivo que se ha optado en utilizar un procedimiento *bootstrap* basado en Modelos Adictivos Generalizados, que consigue captar la tendencia existente en nuestras series de biomasa y captura, y a partir de dicho ajuste si se consiguen residuos independientes procedemos con un remuestreo aleatorio simple sobre los residuos al cual le sumaremos la tendencia ajustada para conseguir las muestras *bootstrap*, en caso de que los residuos fueran dependientes, se ajusta un modelo *autoregressive* (AR) de series temporales sobre los residuos del GAM, las remuestras finalmente serían los valores ajustados del GAM más los valores ajustados por el proceso AR más la remuestra aleatoria simple *bootstrap* de los residuos. Además de esto, presentamos también un procedimiento (simulaciones Montecarlo) para la verificación de la metodología *bootstrap* implementada, cumpliendo así el objetivo de este trabajo.

Una vez se obtuvo esta base metodológica se crearon tres escenarios para la verificación de la misma. En este trabajo la tarea estuvo en hacer una simulación Montecarlo, es decir, tratar de modelar la dinámica de los modelos SPMs con los parámetros adecuados y obtener un modelado exacto donde se conocía la población, para así poder hacer inferencia de las cantidades teóricas. La diferencia de cada escenario estuvo relacionada con la incertidumbre (CVs) introducida en los datos y así observar cómo funcionaban los procedimientos. Una vez se obtuvieron las réplicas, se hizo el procedimiento *bootstrap* sobre estas réplicas para luego obtener las estimaciones con las remuestras *bootstrap*.

Una vez se aplicó esta base metodológica en el estudio, se obtuvieron las siguientes conclusiones:

Esta investigación nos ha permitido proponer un nuevo *bootstrap* que también se podrá utilizar en otros contextos de series temporales que no sean necesariamente nuestro contexto de trabajo.

Se ha podido comprobar el comportamiento de este procedimiento “*bootstrap*” en un proceso de

Montecarlo específicamente diseñado para los modelos KBPM. También se ha podido obtener conclusiones de cómo funciona el *bootstrap* en diferentes escenarios de incertidumbre, por lo tanto, cuando sea aplicado en la práctica por el IEO tendrá una idea de los diferentes resultados que puede obtener por el procedimiento *bootstrap* en los posibles escenarios. En ese sentido, estimar el modelo KBPM con una incertidumbre alta su funcionamiento será regular, por ejemplo, en el escenario 3, observamos que la amplitud de los intervalos en parámetros como  $r$  y  $p$  la amplitud alcanza el 37 % y 57 %, respectivamente del valor real del parámetro, mientras que en el escenario 1 la amplitud del intervalo del parámetro  $r$  no es mayor del 15 % y en  $p$  no es mayor del 30 % del valor real del parámetro. Es importante destacar que esta amplitud sólo se ve influenciada en estos dos parámetros ( $r$  y  $p$ ), dado que en el resto de parámetros la amplitud del intervalo de confianza no sobrepasa el 10 % del valor real en este escenario 3 y en los de menos incertidumbre el resto de parámetros no sobrepasan el 5 % del valor real (escenario 2) y 3 % del valor real (escenario 1).

Con esta investigación, también se puede valorar cómo de bien funcionan los modelos KBPM en los escenarios al mismo tiempo que valoramos el *bootstrap*. Indirectamente en este trabajo se está evaluando cómo funcionan los KBPM dado que podemos obtener información de que tan bien estiman por ambos procesos, por ejemplo, cuando se tienen menos réplicas Montecarlo o más remuestras *bootstrap*. Que, en este caso, al parecer los KBPM estimaran mejor los puntos de referencia y sus respectivos intervalos de confianza cuanto más remuestras *bootstrap* posean, dado que por simulación Montecarlo no se logró estimar bien los intervalos de confianza de los puntos de referencia  $B_{MSY}$  y  $F_{MSY}$ , asociándolo a las pocas réplicas que se tenían.

No obstante, una vez contruidos los intervalos de confianza para las cantidades estimadas por los KBPMs en los distintos escenarios se puede concluir que el estudio nos ha brindado una perspectiva general de que el procedimiento *bootstrap* ha funcionado correctamente y se ha adaptado a cada incertidumbre asociada. Por ejemplo, en momentos donde hay mucha ausencia de incertidumbre el procedimiento *bootstrap* es muy parecido a lo que se obtiene con Montecarlo, incluso en algunos momentos la estimación de las medidas de tendencia central y los intervalos de confianza son mejores que los obtenidos por Montecarlo. Es decir, hacer la evaluación pesquera en los KBPMs por medio de **knobi** introduciendo las remuestras *bootstrap*, hace que las estimaciones arrojadas por los KBPMs sean en algunos momentos mejores que las estimaciones que hace cuando le introduces la dinámica real realizada por las réplicas Montecarlo. Lo que valida el procedimiento *bootstrap* ante baja incertidumbre.

En cualquier caso, creemos conveniente que el IEO podría aplicar esta metodología para la construcción de intervalos de confianza para datos reales. En cierto modo el estudio permite al IEO entender el comportamiento de ambos procedimientos (*bootstrap* y ajuste KBPM) en diferentes escenarios, el cual sabemos que se ve afectado por la incertidumbre pero gracias a este estudio sabemos hasta que punto afecta dicha incertidumbre. Este comportamiento es muy importante tener dominio de su conocimiento dado que se implementará en **knobi** y se podrá usar en la práctica para obtener intervalos de confianza de cantidades muy importantes en la gestión como  $B_{MSY}$  o  $F_{MSY}$ . Habiendo completado estas tareas confío en que se han logrado los objetivos del proyecto.

En otro orden, vale la pena señalar que este estudio también brindó conclusiones generales sobre la metodología utilizada y es que, no cualquier metodología *bootstrap* para datos dependientes puede adaptarse al contexto del trabajo, por ejemplo, específicamente la *Ar-Sieve* y *Moving Block Bootstrap* no acabaron de replicar correctamente la serie temporal en nuestra investigación. Sin embargo, la metodología "GAM+AR", una metodología compleja pero con accesibilidad a herramientas para aplicación automatizada tuvo una correcta adaptación en este contexto.

# Índice de figuras

3.1. Media (para cada año) de la distribución normal que se emplea para generar los valores de mortalidad por pesca. . . . .	17
3.2. Réplicas Montecarlo (gráfico izquierda y derecha superior) y remuestras <i>bootstrap</i> (gráfico izquierda y derecha inferior), captura (izquierda) y biomasa (derecha) utilizando un coeficiente de variación igual a 0.05. Estas series, hasta el año $N = 100$ , se corresponden con 200 réplicas Montecarlo, mientras que el <i>bootstrap</i> se corresponden con un total de 40.000 remuestras. Las líneas discontinuas azules son los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %. Además, la línea discontinua azul central representa la mediana. . . . .	18
3.3. Réplicas Montecarlo (gráfico izquierda y derecha superior) y remuestras <i>bootstrap</i> (gráfico izquierda y derecha inferior), captura (izquierda) y biomasa (derecha) utilizando coeficiente de variación igual a 0.1. Estas series, hasta el año $N = 100$ , se corresponden con 200 réplicas Montecarlo. Mientras que el <i>bootstrap</i> se corresponden con un total de 40.000 remuestras. Las líneas discontinuas azules son los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %. Además, la línea discontinua azul central representa la mediana. . . . .	21
3.4. Réplicas Montecarlo (gráfico izquierda y derecha superior) y remuestras <i>bootstrap</i> (gráfico izquierda y derecha inferior), captura (izquierda) y biomasa (derecha) utilizando coeficiente de variación igual a 0.15. Estas series, hasta el año $N = 100$ , se corresponden con 200 réplicas Montecarlo, mientras que el <i>bootstrap</i> se corresponden con un total de 40.000 remuestras. Las líneas discontinuas azules son los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %. Además, la línea discontinua azul central representa la mediana. . . . .	24



# Índice de tablas

3.1. Valores de los parámetros SPM para la construcción de la curva de producción. . . . .	15
3.2. Valores de los puntos de referencia. . . . .	16
3.3. Medidas de tendencia central y dispersión de los resultados por simulación Montecarlo y remuestras <i>bootstrap</i> con el escenario de incertidumbre ( $CV=0.05$ ). . . . .	19
3.4. Intervalos de confianza de las estimaciones en KBPMs por simulación Montecarlo y remuestreo <i>bootstrap</i> , $CV = 0.05$ . Se ajustan a los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %. . . . .	20
3.5. Medidas de tendencia central y dispersión de los resultados por simulación Montecarlo y remuestras <i>bootstrap</i> con el escenario de incertidumbre ( $CV=0.1$ ). . . . .	22
3.6. Intervalos de confianza de las estimaciones en KBPMs por simulación Montecarlo y remuestreo <i>bootstrap</i> , $CV = 0.1$ . Se ajustan a los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %. . . . .	23
3.7. Medidas de tendencia central y dispersión de los resultados por simulación Montecarlo y remuestras <i>bootstrap</i> con el escenario de incertidumbre ( $CV=0.15$ ). . . . .	25
3.8. Intervalos de confianza de las estimaciones en KBPMs por simulación Montecarlo y remuestreo <i>bootstrap</i> , $Cv = 0.15$ . Se ajustan a los percentiles 0.025 y 0.975, es decir, IC con un nivel de confianza del 95 %. . . . .	26



# Bibliografía

- [1] Bundy A, Bohaboy E, Hjermann D, Mueter F, Fu C, Link J (2012) Common patterns, common drivers: comparative analysis of aggregate surplus production across ecosystems. *Marine Ecology Progress Series*, 459:203–218
- [2] Cousido-Rocha M, Pennino M, Izquierdo F, Paz A, Lojo D, Tifoura A, Zanni M, Cerviño S (2022) Surplus Production Models: a practical review of recent approaches. *Fish Biology and Fisheries* 32:1085–1102.
- [3] Efron B (1979) Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7(1):1-26
- [4] FAO (2018) The state of world fisheries and aquaculture. Meeting the sustainable development goals. Food and Agriculture Organization of the United Nations
- [5] Fox W (1970) An exponential surplus-yield model for optimizing exploited fish populations. *Transactions of the American Fish Society* 99:80–88.
- [6] German S, Hwang C (1982) Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* 10:401–414.
- [7] Grenander U (1981) *Abstract Inference*. New York: Wiley.
- [8] Haddon M (2011) *Modelling and Quantitative Methods in Fisheries* 2nd Edition. CRC Press.
- [9] Hastie T, Tibshirani R (1990) *Generalized Additive Models*. Monographs on Statistics and Applied Probability 43. CHAPMAN & HALL.
- [10] Kilduff P, Carmichael J, Latour R (2009) *Guide to Fisheries Science and Stock Assessments*. Atlantic States Marine Fisheries Commission.
- [11] Kunsch H (1989) The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Statist* 17:1217–1241.
- [12] Liu R, Singh K (1992) Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap* (R. LePage and L Billard, Eds.), pp. 225-248. New York: Wiley.
- [13] MacCall A (2002) Use of Known-Biomass Production Models to Determine Productivity of West Coast Groundfish Stocks. *North American Journal of Fisheries Management*, 22:272–279.
- [14] Mueter F, Megrey B (2006) Using multi-species surplus production models to estimate ecosystem-level maximum sustainable yields. *Fisheries Research*, 81:189–201.
- [15] Schaefer M (1954) Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Bulletin, Inter-American Tropical Tuna Commission* 1:25–56
- [16] Singh K (1981) On the Asymptotic Accuracy of Efron’s Bootstrap. *Ann. Statist.* 9:1187-1195

- [17] Pella J, Tomlinson P (1969) A generalized stock-production model. Bulletin of the Inter-American Tropical Tuna Commission, 13:421-58.
- [18] Russell E (1931) Some theoretical considerations on the “overfishing” problem. Journal du Conseil International pour l’Exploration de la Mer 6:3–20.