

Un modelo aditivo semiparamétrico para estimación de capturas: el caso de las pesquerías de Terranova*

Eva Ferreira y Fernando Tusell

Universidad del País Vasco

Abstract

Se presenta un modelo aditivo semiparamétrico para la estimación de capturas en aguas de Terranova. El modelo se limita a tener en cuenta la presencia de los barcos sobre los caladeros, sus características, y efectos temporales y espaciales. La evolución temporal de las capturas se modeliza de forma no paramétrica, así como el efecto del tamaño de los buques (medido por su eslora). Los resultados mejoran los que se obtienen con el sistema actual.

1 Introducción

El control de la pesca se realiza en distintos países de muy diversos modos. En la Unión Europea prevalecen sistemas basados en la información proporcionada por los pescadores, que deben someter a las autoridades declaraciones de desembarque, notas de venta, o documentos de similar efecto.

El sistema se presta al fraude. La experiencia muestra que es casi imposible de implementar con garantías. De aquí que en la actualidad la Comisión tenga en estudio la instalación en barcos de una cierta dimensión de “cajas azules”, transmisores que permitirían la rápida localización de las unidades desde un satélite, y el control del tiempo que pasan en los caladeros —e indirectamente, por tanto, de su esfuerzo—. Se contempla así un cierto desplazamiento desde el control físico de los desembarcos hacia el control del esfuerzo de pesca realizado.

Hay casos, incluso, en que la presencia sobre los caladeros es la única variable observable, al no existir posibilidad legal de requerir ninguna información a los patrones o armadores. En el caso de la pesquería de bacalao y otras especies en aguas de Terranova, el Gran Banco se extiende fuera de las 200 millas que abarca la Zona Económica Exclusiva de Canadá. Los barcos que enarbolan pabellón de estados signatarios de la Convención North Atlantic Fisheries Organization, están sujetos a ciertas obligaciones formales que se señalan más abajo; pero existen otros muchos que no están sujetos a ningún tipo de declaración o control. Por ello, un sistema de estimación de capturas basado en la única información de que se puede disponer —presencia en los caladeros y, en algunos casos, características de los buques— es la única aproximación directa a la estimación de capturas.

En lo que sigue se describe muy sucintamente la pesquería objeto de estudio, y el entramado legal asociado. Una historia muy legible, escrita desde el punto de vista canadiense, pero realizando un esfuerzo

*Este trabajo se inició cuando el segundo autor estaba al servicio de la Comisión Europea, DG-XIV-C. El segundo autor está agradecido a Luc Debeer por los datos, a numerosos colegas en el Department of Fisheries and Oceans (St. John's) y en la DG-XIV, y, muy especialmente, al Cap. J.C. Mesquita, que durante un año respondió a sus preguntas, compartiendo su conocimiento del Gran Banco y su pasión por el mar. Estamos también muy agradecidos al director de la revista y a dos evaluadores anónimos, por sus comentarios.

notable de objetividad, es [2]. Una descripción mucho más técnica de las complejas relaciones biológicas que sustenta el Gran Banco puede encontrarse en [3].

El Gran Banco está formado por la plataforma continental, al Sur y Este de la isla de Terranova. Son aguas poco profundas, y las condiciones morfológicas de los fondos, temperatura del agua, y corrientes hacen de él un caladero de legendaria riqueza. Casi todo el Gran Banco, como se ha dicho, queda en el interior de la Zona Económica Exclusiva de Canadá. Pequeñas porciones del Gran Banco (“the Tail of the Bank” y “the Nose of the Bank”) salen de la línea de las 200 millas que Canadá explota en exclusiva. Lo mismo ocurre con otro rico caladero, Flemish Cap, especie de promontorio submarino que forma una zona de poco fondo rodeada de aguas mucho más profundas. Estas áreas, al quedar en aguas internacionales, son de libre acceso a todos los barcos.

El conjunto del Gran Banco y su área circundante está dividido en zonas a efectos administrativos y estadísticos. Las que cubre el área de interés en lo que sigue son la 3M, 3N, 3L y 3O. Las cuotas se distribuyen por zonas.

Para regular el ejercicio de la pesca fuera de la Zona Económica Exclusiva canadiense, un grupo de países, entre los que se cuentan los más ligados históricamente a la pesquería (España y Portugal, luego integrados en la UE) firmaron un convenio concediéndose mutuamente derechos de inspección y control. Los buques de pesca abanderados en cualquiera de los estados signatarios (o “partes contratantes”) pueden ser inspeccionados por autoridades de cualquier otro estado signatario. Además, están sujetos a obligaciones formales como declarar su entrada en, y salida de, cada una de las zonas indicadas, y su posición. Esta comunicación (el “hail report”), en el caso de los barcos de la UE, se hace a Canadá y a Bruselas.

Los estados signatarios intercambian *hails* e información sobre la ubicación de sus buques. Paralelamente, en el seno de la North Atlantic Fisheries Organization, cuyos miembros se reúnen periódicamente, se debate el monto y reparto de cuotas entre los diversos países. Es de notar que un estado signatario puede “declarar una excepción” en el plazo de 60 días desde la toma de un acuerdo, pero es lo cierto que las decisiones acaban siendo relativamente vinculantes. Naturalmente, los buques pertenecientes a estados no signatarios no pueden ser objeto de ningún control ni sujetos a obligaciones formales cuando ejercen su actividad en aguas internacionales, fuera de la Zona Económica Exclusiva¹.

Los buques abanderados en estados signatarios deben llevar un diario de a bordo en el que se anotan las capturas, por fecha, especie, tonelaje y zona de captura. Dicho diario refleja también las transformaciones (p.ej: fileteado de bacalao) que impliquen cambio de peso, los transbordos y toda eventualidad que registren las capturas a bordo, así como la actividad de pesca. Las anotaciones han de hacerse diariamente, o más frecuentemente si hay un cambio de zona.

Cualquier estado signatario que haya destacado inspectores en la zona puede inspeccionar a cualquier buque de su propio estado o de otro también signatario. En la actualidad, Canadá (con dos buques de inspección) y la UE (con uno) ejercen esta prerrogativa. Localizado un barco de pesca, los inspectores de cualquier parte contratante pueden decidir su abordaje, verificar el cumplimiento de la normativa (en cuanto a artes, mallas, etc.) y comparar el stock de pescado en la bodega con las anotaciones en el diario de a bordo. El buque de inspección puede también limitarse a tomar constancia de la presencia en el caladero del barco de que se trate, efectuando un avistamiento (*sighting*). Las inspecciones canadiense y europea procuran coordinarse para no inspeccionar con escaso intervalo el mismo barco, causando molestias innecesarias a los pescadores.

Adicionalmente, dos vuelos de reconocimiento diarios recorren el Gran Banco y Flemish Cap, localizando objetivos en el radar y descendiendo a identificarlos en vuelo rasante. Esto hace que incluso buques

¹Este párrafo se escribió en febrero de 1.995. Acontecimientos posteriores muestran que, si sigue siendo cierto *de iure*, cada vez es más difícil ignorar la tendencia de algunos países ribereños a atribuirse jurisdicción más allá de las 200 millas de su Zona Económica Exclusiva, en un movimiento similar al que propició en los años setenta el establecimiento de las actuales Zona Económica Exclusiva de 200 millas.

de países no signatarios de la Convención NAFO estén controlados en cuanto a su presencia en los caladeros. Como media, la vigilancia aérea garantiza el avistamiento de todos y cada uno de los barcos en la zona —salvo condiciones de tiempo extremadamente malas por tiempo prolongado— al menos cada cinco días. A ello se unen los *hails*, y los avistamientos de los buques de inspección. En conjunto, puede decirse que la presencia sobre los caladeros es objeto de control minucioso y exhaustivo.

Como resultado de cada abordamiento, los inspectores que lo realizan retienen una copia del diario de a bordo con la información señalada más arriba desde la última inspección efectuada. La información derivada de las respectivas inspecciones es procesada en Bruselas y St. John's (Newfoundland) para estimar las capturas. Estas estimaciones son más tarde objeto de comparación y examen, si se producen discrepancias.

Hasta la fecha se ha seguido un sistema de estimación muy simple: por cada mes y población de peces, se calcula una captura media diaria, que luego se multiplica por el número de días de esfuerzo, computados estos últimos mediante la recopilación de avistamientos y *hails*. El sistema descrito no carece de problemas, que describimos a continuación.

El cálculo mensual de tasas de captura por día ofrece una resolución temporal adecuada². Sin embargo, hay meses en los que puede haber muy pocas observaciones —por ejemplo, porque el mal tiempo continuado no ha permitido apenas abordamientos—. El elevar promedios diarios de capturas basados en muy pocas observaciones hace correr el riesgo de obtener estimaciones muy erráticas.

Por otra parte, los barcos que participan en la pesquería tienen capacidad de pesca muy diversa, derivada tanto de los diferentes tipos de pesca practicados, como del arte empleada y la propia dimensión del navío —un barco mayor puede arrastrar un arte de mayor apertura, y en el mismo número de horas de arrastre obtener más pescado—. Lo anterior es particularmente importante si se tiene en cuenta que no cabe hacer un diseño experimental y proceder a inspeccionar barcos de diferentes características en proporciones prefijadas. La posibilidad de abordar un buque concreto depende de que sea encontrado, de que el tiempo lo permita, y de si las autoridades canadienses lo han hecho en el pasado inmediato o no. En la práctica, el barco de inspección de la Unión Europea patrulla por zonas escogidas, pero no tiene opción a seleccionar el tipo de barcos que aborda de acuerdo con un diseño experimental. Puede así ocurrir, y de hecho ocurre, que durante meses concretos en los que se producen pocos abordajes, determinados tipos de barcos —arrastreros, volanteros, de pequeña o gran dimensión— tengan un peso en la muestra completamente diferente del correspondiente en la población. La falta de corrección de este efecto origina tasas estimadas de capturas diarias que deben ser modificadas mediante criterios no estadísticos.

2 Un modelo semiparamétrico para las capturas

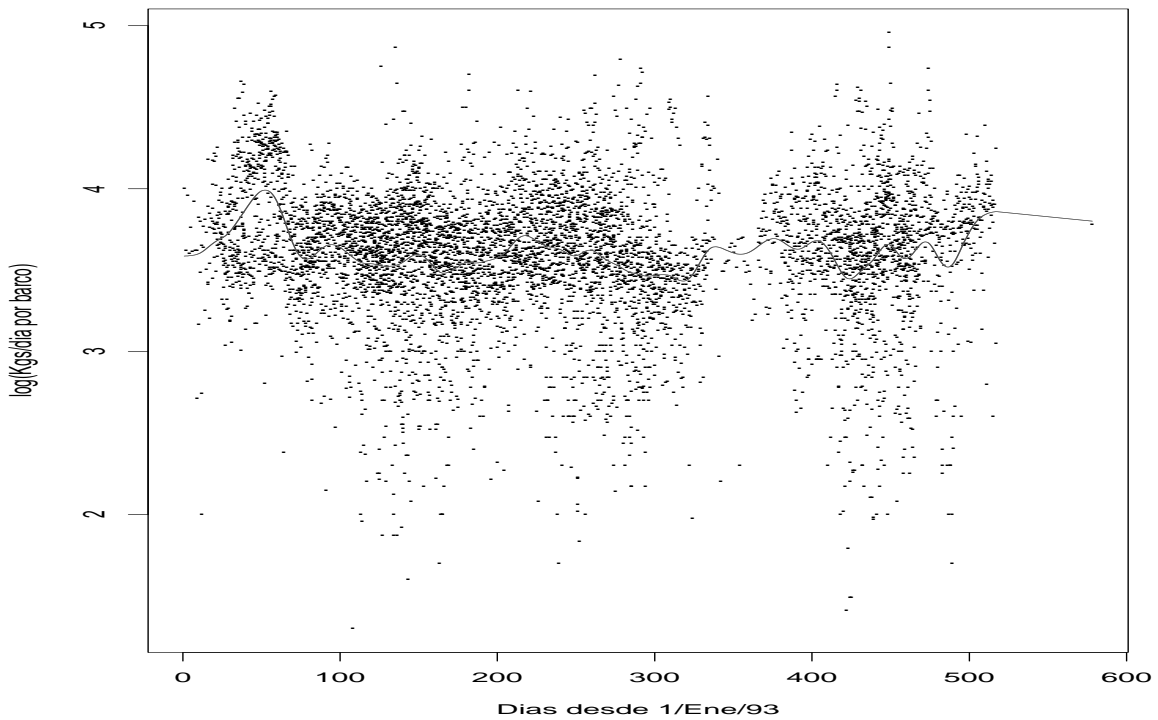
2.1 Descripción de los datos

Los datos disponibles proceden de una muestra de diarios de a bordo (los recolectados por la Inspección de la Unión Europea en las campañas realizadas por el “Ernest Haeckel” durante 1993 y “Kommandor Amalie” en la primera mitad de 1994). Se extienden por tanto sobre un periodo de unos 500 días. Suman 17189 observaciones conteniendo: identificación del barco, tipo de barco y arte empleada, malla, zona de pesca, especie y cantidad capturada y fecha.

Los promedios de capturas diarias a estimar han de ser multiplicados por días de presencia en los caladeros, procedentes de los *hails* y avistamientos. Importa, por ello, calcularlos de modo tan coherente como sea posible con los criterios para determinar si un barco está o no sobre los caladeros a efectos de *hail*

²Las capturas parecen presentar variaciones estacionales, derivadas del hecho de que en ciertas épocas el pescado se aleja de la costa y se adentra en las zonas del Gran Banco donde puede ser pescado por buques no canadienses, que no tienen acceso a la Zona Económica Exclusiva de 200 millas. Además están factores como el hielo, que impiden la pesca en ciertas zonas y periodos.

Figure 1: Capturas por día y barco (todas las especies sumadas) en las zonas NAFO 3M, 3N, 3L, y 3O.



o avistamiento. Por ejemplo, un diario de a bordo puede reflejar de modo inequívoco que en tres días consecutivos determinados no se pescó por mal tiempo (“a la capa”, “mau tempo”, o menciones similares). Si el cuarto día se registra pesca, asignamos ésta a cuatro días, pues durante los cuatro el barco ha permanecido en los caladeros y ha sido contado como presente; y ello independientemente de que las capturas se hayan realizado en un sólo día de pesca efectiva.

Al margen de eventualidades como la considerada, esto es importante en el caso de las parejas, que frecuentemente sólo registran pesca días alternos —la pesca de los días sin anotar en un barco de la pareja es recogida habitualmente en el diario de a bordo del compañero—. Lo ideal sería registrar la pareja como entidad, pero esto no es práctico en todos los casos: por ello se ajustan los datos de modo que se reflejen los días sobre los caladeros de cada unidad.

El Cuadro 1 proporciona una descripción somera de los datos empleados. El Cuadro 2 proporciona una explicación de las variables no clarificadas a lo largo del texto.

Como muestra de la fluctuación temporal de las capturas, el Gráfico 1 exhibe totales diarios por barco (sumadas todas las especies); la línea sólida corresponde a un suavizador no paramétrico, que se explica en la siguiente sección. Son aparentes en el gráfico las elevadas capturas hacia Febrero de 1993 (días 30 a 50 aproximadamente) y la relativa escasez de datos en Diciembre 1993-Enero 1994, periodo en que se produjo el relevo del barco de inspección de la Unión Europea en la zona.

Table 1: Resumen de datos empleados.

| Variable | Niveles | Número | Mínimo | Primer cuartil | Mediana | Media | Tercer cuartil | Máximo |
|-----------------|----------------|---------------|---------------|-----------------------|----------------|--------------|-----------------------|---------------|
| Nación | DNK | 26 | | | | | | |
| | ESP | 12560 | | | | | | |
| | PRT | 4603 | | | | | | |
| Zona | 3L | 5186 | | | | | | |
| | 3M | 4352 | | | | | | |
| | 3N | 6600 | | | | | | |
| | 3O | 1051 | | | | | | |
| Especie | COD | 931 | | | | | | |
| | GHL | 5593 | | | | | | |
| | RED | 975 | | | | | | |
| | RNG | 3883 | | | | | | |
| | SKA | 3038 | | | | | | |
| | Resto | 2769 | | | | | | |
| Kgs | | | 1 | 148 | 576 | 2223 | 3150 | 90600 |
| Día | | | 1 | 134 | 228 | 243 | 340 | 517 |
| Pareja | Solo | 15585 | | | | | | |
| | Par | 1604 | | | | | | |
| Tipo | GO | 1195 | | | | | | |
| | SN | 126 | | | | | | |
| | TO | 278 | | | | | | |
| | TS | 132 | | | | | | |
| | TT | 10996 | | | | | | |
| | TTF | 4462 | | | | | | |
| Malla | | | 110.0 | 120.0 | 125.0 | 126.4 | 130.0 | 150 |
| Eslora | | | 29.0 | 42.7 | 52.8 | 54.1 | 64.8 | 84.9 |
| TRB | | | 242.3 | 388.2 | 742.6 | 832.5 | 996.0 | 2382.0 |
| Kwh | | | 478.0 | 735.0 | 1164.0 | 1223.0 | 1470.0 | 2648.0 |

Nota: Para las variables cualitativas, en “Número” se recoge el número de observaciones en cada nivel.

Table 2: Significado de las variables consideradas.

| Variable | Nivel | Significado |
|---|--|--|
| Arte | GN GNS OT OTB PTB | Red de enmalle a la deriva Red de enmalle anclada “Otter trawl” : arrastre “Otter trawl bottom” : arrastre de fondo Arrastre de fondo por parejas |
| Tipo | GO SN TO TS TT TTF | Volantero Cerquero (“seiner”) Arrastrero no especificado Arrastrero por el costado (“side trawler”) Arrastrero por la popa Arrastrero por la popa congelador |
| Pareja | Par Solo | Pescando en pareja Pescando sólo |
| Especie | GHL (62.3%) RED (12.5%) COD (8.7%) SKA (9.0%) RNG (3.8%) Resto (3.7%) | Fletán negro (<i>Hippoglossus reinhardtius</i>) Gallineta (<i>Sebastes spp.</i>) Bacalao (<i>Gadus morhua</i>) Raya (<i>Raja spp.</i>) Granadero (<i>Macrurus rupestris</i>) Locha, Perritos del norte, Falso fletán Camarón boreal, Platija, Mendo, Limanda Fletán, Rape, Escorpión, Atún y Caballas |
| País | DNK ESP PTR | Dinamarca España Portugal |
| Zona Kgs Día Malla Eslora TRB Kwh | 3L,3M,3N,3O | Zona NAFO Capturas diarias en Kg. peso vivo Días transcurridos a partir del 1/1/93 Malla en mm. Eslora total en metros Toneladas de registro bruto Potencia del motor principal en kilowatios |

Nota: Los porcentajes lo son sobre la captura total inspeccionada.

2.2 Estimadores tipo *spline*

Los estimadores *spline* han sido los elegidos para estimar los efectos no paramétricos. De forma breve describimos su obtención. Dadas dos variables relacionadas por

$$y_i = f(x_i) + \epsilon_i \quad (1)$$

donde $f(\cdot)$ es una función “suave”, los errores ϵ_i son independientes e idénticamente distribuidos, e $i = 1, \dots, n$, el *spline* cúbico $\hat{f}(\cdot)$ estimador de $f(\cdot)$ se determina minimizando la expresión

$$MSE + \lambda \int (s'')^2 = \frac{1}{n} \sum_{i=1}^n (y_i - s(x_i))^2 + \lambda \int (s'')^2 \quad (2)$$

en un espacio de funciones $s(\cdot)$ “suaves”. El primer término MSE denota el error cuadrático medio y λ es un parámetro de suavizado.

De este modo, una elección adecuada de λ proporciona un estimador que, a la vez de ajustarse a los datos (ajuste medido por el MSE), se somete a restricciones de suavidad (medida por $\int (s'')^2$). El empleo del término $\lambda \int (s'')^2$ permite, tomando diferentes valores de λ , seleccionar la suavidad que se desea para el estimador y, en consecuencia, escoger el número equivalente de grados de libertad. El valor de λ puede ser fijado por el analista en un valor juzgado conveniente; o, alternativamente, puede determinarse de modo automático, mediante alguno de los métodos disponibles. El empleo de un método automático es casi imperativo cuando se emplea un algoritmo iterativo de búsqueda, que ha de hacer un suavizado en cada iteración. Es usual el método de *validación cruzada*. Una descripción detallada puede encontrarse en [4].

Esencialmente, se trata de un ingenioso procedimiento de reutilización de la muestra, en que cada observación se utiliza por turno para estimar y para validar. En el caso que nos ocupa, si la muestra es de tamaño N se deja fuera de la misma una observación diferente cada vez, que se intenta ajustar mediante las restantes. Se selecciona el λ que proporciona mejor ajuste global.

Como ejemplo en el caso que nos ocupa, la línea sólida del Gráfico 1 muestra un *spline* cúbico con 36 grados de libertad. Puede verse que realiza un ajuste de la evolución temporal de las capturas que sería difícil lograr con ayuda de una especificación paramétrica.

2.3 La modelización

Un examen somero de los datos y la consulta con personas con experiencia en la pesquería proporcionó la siguiente información: es razonable suponer una variación estacional en las capturas —clara por otra parte en el Gráfico 1—, y también en función de las especies perseguidas, el tipo de arte y, quizá, la dimensión del navío. También parece razonable esperar diferentes rendimientos, en igualdad de todo lo demás, para barcos pescando solos y barcos pescando en pareja³.

Con la anterior información hemos ensayado buen número de modelos, en todos los cuales el regresando es el logaritmo de las capturas diarias por barco, sea sumadas todas las especies, sea desglosando por especies de acuerdo al detalle en el Cuadro 2. El empleo de una escala logarítmica para el regresando surge de modo bastante natural. Parece sensato que variables como el tipo de arte pesca empleado tengan un impacto *multiplicativo* en la captura diaria media: un arrastrero, a igualdad de zona, características del buque, etc., podría pescar, digamos, un 20% más que un volantero, pero *no* dos toneladas diarias más. Las capturas son función de la abundancia relativa, y es poco realista pensar en un arte rindiendo una cantidad fija más que otra, de modo independiente de la abundancia de pescado.

³Supuestamente, el sistema de arrastre por parejas permite una mayor rentabilidad —derivada quizá de un mayor número diario de horas de arrastre—. Los resultados obtenidos más abajo parecen diferir de lo esperado en este punto.

La especificación escogida con el regresando en logaritmos⁴ es por ello mucho más atractiva, pese a los problemas que introduce en la estimación de las varianzas de las predicciones, problemas que examinaremos en la Sección 3.

Los modelos considerados son de la forma

$$Y = \beta_0 + \sum_{j=1}^{J_1} \beta_{1j} X_{1j} + \dots + \sum_{l=1}^{J_I} \beta_{Il} X_{Il} + f_1(Z_1) + \dots + f_p(Z_p) + \epsilon, \quad (3)$$

siendo X_1, \dots, X_I variables cualitativas (zona, arte, especie, etc.), cada una desglosada en J_1, \dots, J_I variables 0-1 respectivamente, Z_1, \dots, Z_p son variables continuas, y $f_1(\cdot), \dots, f_p(\cdot)$ son *splines* cúbicos. La variable Y es, como se ha dicho, el logaritmo de la captura diaria por barco y especie. Los coeficientes β_{ij} están sujetos a las restricciones habituales en análisis de varianza (o, alternativamente, se prescinde de uno de ellos en cada bloque, para eliminar la dependencia lineal con la columna constante que de otro modo aparecería: el modelo, por tanto, es identificable).

La especificación semiparamétrica permite una gran flexibilidad; la influencia de cada una de las variables Z_1, \dots, Z_p consideradas sobre las capturas no está constreñida a adoptar una forma funcional pre-determinada. Como se verá, ello es particularmente atractivo en el caso de la variable Día, que recoge la evolución temporal.

El método de ajuste empleado es el conocido como *backfitting* (véase, por ejemplo, Hastie y Tibshirani, 1991, o Green y Silverman, 1994; Chambers y Hastie, 1992, también facilitan detalles sobre el procedimiento y el programa empleado). Los parámetros correspondientes a términos lineales, β_{ij} en (3), no ofrecen dificultad. En la iteración k -ésima, el *spline* $f_s(\cdot)$ se estima suavizando los residuos parciales

$$\hat{\epsilon}_{(s)} = Y - \hat{\beta}_0 - \sum_{i=1}^I \sum_{j=1}^{J_i} \hat{\beta}_{ij} X_{ij} - \sum_{m \neq s} \hat{f}_m(Z_m). \quad (4)$$

En (4), $\hat{f}_m(Z_m)$ son los estimadores *spline* ajustados en la iteración anterior: inicialmente pueden ser cero, u otro valor conveniente. Bajo condiciones bastante generales el procedimiento anterior converge, aunque la convergencia puede ser lenta: véase por ejemplo Green y Silverman, (1994, p. 68).

La idea intuitiva que subyace al método de ajuste de *backfitting* puede describirse brevemente así: para estimar la influencia de una cierta variable Z_s sobre Y se elimina del regresando aquella parte atribuible a todos los demás regresores (paramétricos o no). Esto es lo que se hace en la ecuación (4). Lo que resta (los residuos parciales $\hat{\epsilon}_{(s)}$) se regresa entonces sobre la variable Z_s para dar una estimación $\hat{f}_s(Z_s)$.

La búsqueda de un modelo adecuado se ha hecho mediante un algoritmo *stepwise*. Con un conjunto de datos relativamente grande y bastantes variables candidatas a ser introducidas en el modelo de forma no paramétrica, es natural que la búsqueda haya de ser restringida para mantener en límites razonables el esfuerzo de cálculo.

En nuestro contexto, ello se ha hecho limitando los suavizados a intentar para los respectivos residuos parciales, obtenidos mediante (4) en cada iteración. Por ejemplo, para la variable Kwh (potencia del barco) nos hemos limitado a considerar como términos posibles $\gamma_{Kwh, s(Kwh, g1=1)}$, $s(Kwh, g1=2)$ y $s(Kwh, g1=3)$, en donde γ es un parámetro a estimar y $s(Kwh, g1=n)$ designa un *spline* suavizador con n grados de libertad. La primera opción es un término lineal en Kwh; en cierto sentido, el término más suave que cabe imaginar. Las sucesivas son *splines* con un número creciente de grados de libertad consumidos.

El algoritmo *stepwise* procede incluyendo términos de progresivamente mayor complejidad en tanto ello mejore el valor de un criterio de bondad de ajuste (en nuestro caso, el AIC de Akaike; véase por ej. Akaike,

⁴Logaritmos decimales, cuestión que un evaluador solicitaba se aclarara. No hay razón especial para ello: podrían serlo en cualquier base, sin otro efecto que introducir un factor de escala.

1972). La selección de términos opera en ambos sentidos, pudiéndose excluir un término previamente incluido si su contribución al ajuste se hiciera no significativa.

2.4 Elección de la especificación.

Cabe considerar varios tipos de modelos, todos ellos de la forma (3). En primer lugar, tendríamos modelos para el total de las capturas, lo que supone la supresión del desglose por especies. Un modelo así sería de utilidad si hay información adicional sobre la composición por especies: originada, por ejemplo, como resultado de campañas científicas que se efectúan periódicamente, en que se realizan arrastres cuidadosamente controlados y se obtienen datos sobre composición por especies, y tamaños/edades. El modelo daría así sólo una estimación de la pesca total, y la asignación por especies se haría de acuerdo con la información adicional aludida.

En segundo lugar, podríamos modelizar las capturas teniendo en cuenta la especie. Ello supone una especificación más detallada, pero la información reflejada en los diarios de a bordo sobre la composición por especie no es tan fiable como la referente a la cuantía total de la pesca, de más fácil comprobación.

Cabría, finalmente, considerar modelos separados para cada especie.

Todas las alternativas han sido ensayadas, pero en interés de la brevedad nos limitamos a mostrar un modelo con desglose por especies (segunda de las posibilidades mencionadas).

El modelo seleccionado ha sido

$$\begin{aligned} \log(\text{Kgs diarios barco}) = & \beta_0 + \sum_j \beta_{1j} \text{Zona}_j + \sum_k \beta_{2k} \text{Arte}_k + \sum_l \beta_{2l} \text{Especie}_l \\ & + s(\text{Dia}, \text{gl}=24) + s(\text{Eslora}, \text{gl}=3) + \epsilon, \end{aligned} \quad (5)$$

en donde la variable dependiente es el logaritmo de las capturas diarias por especie y barco. Hay tantas observaciones por barco como días de pesca efectiva del mismo.

Las variables explicativas son Zona, Arte, y Especie, variables 0-1 recogiendo el efecto de los distintos niveles de dichos efectos, listados en el Cuadro 2, y las variables consideradas como continuas, Día y Eslora.

El modelo (5) utiliza un total de 40 grados de libertad equivalentes. Logra un $R^2 = 0.5226$, lo que es aceptable teniendo en cuenta la naturaleza del problema; las capturas de un día a otro son muy variables, incluso para un mismo barco, especie, arte y zona, y no es extraño que un modelo que considera sólo estas variables proporcione un ajuste modesto.

Un estadillo con los resultados del ajuste se ofrece en el Cuadro 3. Se presentan los resultados de la estimación por mínimos cuadrados ordinarios y mínimos cuadrados generalizados, cuya motivación se verá en lo que sigue; los resultados son muy similares, resultado esperable con una muestra tan grande. En ambos casos se empleó *backfitting* dado el carácter semiparamétrico del modelo.

Obsérvese que Zona, Arte y Especie, con 4, 5 y 6 niveles (véase Cuadro 2), dan lugar a 3, 4 y 5 parámetros estimados, respectivamente: el nivel que falta en cada bloque está así incluido en el término constante.

2.5 Comentarios al modelo seleccionado

El ajuste para las capturas diarias no es espectacular, como el moderado R^2 pone de manifiesto. La función que relaciona la variable endógena con la eslora de los buques consume tres grados de libertad y muestra un cambio de pendiente para una eslora de, aproximadamente, 43 metros. Las capturas de buques muy pequeños parecen depender menos de la eslora que a partir de dicho umbral, en que la pendiente se incrementa.

Table 3: Parámetros estimados y estadísticos de ajuste del modelo (5)

| Coeficientes | MCO | | | MCG | | |
|------------------------|---------|----------|----------|---------|----------|----------|
| | Valor | Estad. t | Pr(> t) | Valor | Estad. t | Pr(> t) |
| (Constante) | 2.5002 | 77.3907 | 0.0000 | 2.4982 | 76.3267 | 0.0000 |
| Especie: GHL | 0.0930 | 4.2799 | 0.0000 | 0.0917 | 4.3279 | 0.0000 |
| Especie: RED | -0.1309 | -5.1009 | 0.0000 | -0.1697 | -6.7049 | 0.0000 |
| Especie: RNG | -1.0772 | -47.6989 | 0.0000 | -1.0968 | -49.6718 | 0.0000 |
| Especie: Resto | -1.1607 | -53.2326 | 0.0000 | -1.2088 | -56.3747 | 0.0000 |
| Especie: SKA | -0.8243 | -36.6660 | 0.0000 | -0.8299 | -37.8393 | 0.0000 |
| s(Dia, gl = 24) | -0.0003 | -10.5261 | 0.0000 | -0.0003 | -9.5270 | 0.0000 |
| Arte: GNS | 0.3445 | 9.9890 | 0.0000 | 0.3484 | 9.8428 | 0.0000 |
| Arte: OT | 0.7719 | 7.8423 | 0.0000 | 0.7289 | 8.4747 | 0.0000 |
| Arte: OTB | 0.6943 | 32.6481 | 0.0000 | 0.6817 | 30.2948 | 0.0000 |
| Arte: PTB | 0.5305 | 21.4939 | 0.0000 | 0.5279 | 21.0004 | 0.0000 |
| Zona: 3M | -0.0247 | -1.9227 | 0.0545 | -0.0445 | -3.6011 | 0.0003 |
| Zona: 3N | 0.0837 | 7.9558 | 0.0000 | 0.0945 | 9.2466 | 0.0000 |
| Zona: 3O | 0.0272 | 1.3698 | 0.1708 | 0.0027 | 0.1448 | 0.8849 |
| s(Eslora, gl = 3) | 0.0048 | 16.7068 | 0.0000 | 0.0052 | 18.1451 | 0.0000 |
| Error cuadrático medio | 0.2816 | | | 1.057 | | |
| Grados de libertad | 17174 | | | 17174 | | |
| R^2 | 0.5226 | | | 0.5432 | | |

Los parámetros estimados para los diferentes tipos de Arte muestran el menor rendimiento de los volaneros en relación con los arrastreros, lo que era de esperar. Muestran también un menor rendimiento de los barcos operando en pareja, circunstancia menos previsible. Es posible que esta variable esté recogiendo efectos no bien modelados de la dimensión de los barcos (los que faenan por parejas son habitualmente menores).

Especialmente interesante es la función que recoge la evolución de las capturas atribuible al paso del tiempo (variable Día). El perfil de la función, bastante similar al recogido en la Figura 1, muestra un efecto “tiempo” importante (no se olvide que la escala es logarítmica y se extiende sobre varios órdenes de magnitud para acomodar las observaciones extremas). Es difícil imaginar un ajuste comparable mediante un modelo paramétrico convencional.

El examen de los gráficos de residuos parciales no sugiere problemas importantes, salvo acaso motivados por heterocedasticidad. El Gráfico 2 muestra un *spline* suavizador de las varianzas muestrales de los residuos día a día, con suavizado escogido por validación cruzada. Parece claro que hay una fluctuación apreciable, con un periodo de unos 100 a 120 días, que no ha sido posible atribuir a factores estacionales ni a otras causas conocidas.

Sea cual fuere el origen de dicha pauta temporal en la varianza residual, parece procedente hacer uso de la misma para reestimar el modelo por mínimos cuadrados generalizados. Los resultados son los que aparecen en las tres últimas columnas del Cuadro 3. En el proceso iterativo de *backfitting*, la parte paramétrica se ajusta del modo habitual mediante,

$$\hat{\beta} = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} \hat{\nu}, \quad (6)$$

en que $\hat{\Omega} = \text{diag}(\hat{\sigma}_k^2)$, $\hat{\sigma}_k^2$ es el valor suavizado de la varianza en el día k -ésimo, X es la matriz de regresores paramétricos, y

$$\hat{\nu} = Y - \sum_{m=1}^p \hat{f}_m(Z_m). \quad (7)$$

En cuando a la parte no paramétrica, se ajustan *splines* ponderados, soluciones del problema

$$\min \left\{ \sum_{k=1}^N w_k (\hat{\epsilon}_{(s)k} - f_s(Z_{sk}))^2 + \lambda_s \int [f_s''(z)]^2 dz \right\} \quad (8)$$

donde $\hat{\epsilon}_{(s)k}$ es la k -ésima observación del residuo parcial $\hat{\epsilon}_{(s)}$ en (4), $w_k = \hat{\sigma}_k^{-2}$ y λ_s el parámetro de suavizado.

3 Utilización del modelo

A partir de los *hail* se puede obtener para cada barco y día sobre los caladeros un vector x_{*i} de valores de las variables explicativas. Sustituyendo x_{*i} en la expresión

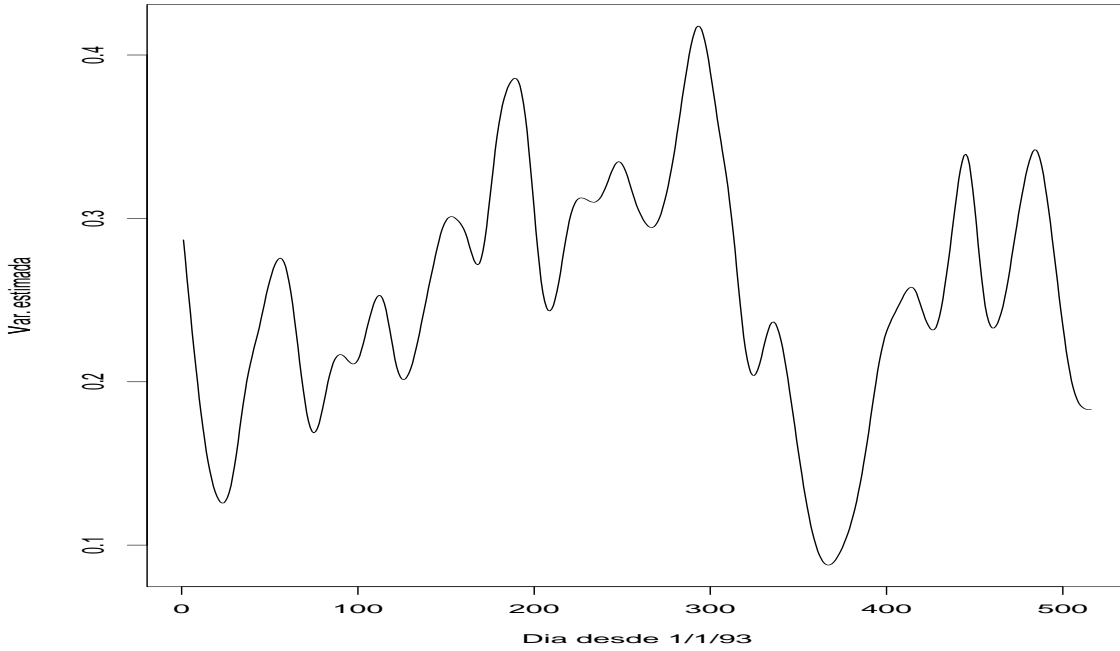
$$\begin{aligned} \hat{y}_{*i} = & \hat{\beta}_0 + \sum_j \hat{\beta}_{1j} \text{Zona}_j + \sum_k \hat{\beta}_{2k} \text{Arte}_k + \sum_l \hat{\beta}_{2l} \text{Especie}_l \\ & + s(\text{Dia}, \text{gl}=24) + s(\text{Eslora}, \text{gl}=3), \end{aligned} \quad (9)$$

obtenemos un valor estimado del logaritmo de las capturas asociado a x_{*i} .

Si prescindimos de la aparente heterocedasticidad en las perturbaciones, por analogía con lo que ocurre en un modelo de regresión ordinario, podemos escribir

$$\text{Var}(\hat{y}_{*i}) \approx \hat{\sigma}_\epsilon^2 (1 + x_{*i}' (X' X)^{-1} x_{*i}), \quad (10)$$

Figure 2: Varianza estimada de los residuos MCO, suavizada mediante un *spline* con 36 grados de libertad.



donde X es la matriz de diseño que engloba tanto las variables ordinarias como los términos “suaves”. Para tomar en consideración la heterocedasticidad encontrada, habríamos de reemplazar (10) por

$$\text{Var}(\hat{y}_{*i}) \approx \hat{\sigma}_{\epsilon_i}^2 (1 + x_{*i}'(X'X)^{-1}x_{*i}), \quad (11)$$

donde $\hat{\sigma}_{\epsilon_i}^2$ es la varianza residual estimada en el día i -ésimo. Tanto en (10) como en (11) estamos prescindiendo del hecho de que algunas columnas de X —las correspondientes a términos “suaves”— son a su vez ajustes, y consumen algunos grados de libertad.

Más importante es el hecho de que estamos interesados en estimar no el logaritmo de las capturas, y_{*i} , sino las capturas mismas, $C_i = 10^{y_{*i}}$, y sumas de las mismas (sobre especies, zonas, etc.). Podemos hacer $\hat{C}_i = 10^{\hat{y}_{*i}}$ con un ligero sesgo (desigualdad de Jensen), que cabe despreciar. Requerimos no obstante una aproximación, siquiera de cierta crudeza, para la varianza de \hat{C}_i y la de sus sumas.

Sea $m_i = E[y_{*i}]$ y $C_i = 10^{y_{*i}} = g(y_{*i})$. La varianza de C_i puede entonces aproximarse con facilidad como función de la de y_{*i} (véase Cox-Hinkley 1980, p. 105). Consideremos el desarrollo en serie

$$C_i = g(y_{*i}) = g(m_i) + g'(m_i)(y_{*i} - m_i) + \frac{1}{2}g''(y_{*i})(y_{*i} - m_i)^2 + \dots \quad (12)$$

Tomando esperanzas,

$$E[C_i] = E[g(y_{*i})] = g(m_i) + \frac{1}{2}g''(y_{*i})E(y_{*i} - m_i)^2 + \dots \quad (13)$$

De (12) y (13) obtenemos:

$$E(C_i - E[C_i])^2 = \text{Var}(C_i) = (g'(m_i))^2 \text{Var}(y_{*i}) + \dots \quad (14)$$

Pueden tomarse el primero o más términos de (14). Una aproximación basada en el primer término es:

$$\text{Var}(\hat{C}_i) \simeq (g'(\hat{y}_{*i}))^2 \text{Var}(\hat{y}_{*i}) = [10^{\hat{y}_{*i}} (\log_e 10)]^2 \hat{\sigma}_{\epsilon_i}^2 (1 + x_{*i}'(X'X)^{-1}x_{*i}) \quad (15)$$

4 Conclusiones

Este trabajo presenta un modelo para estimar capturas con la única base de los diarios de a bordo inspeccionados y la información sobre presencia en los caladeros y características de los buques. La consideración tanto del arte empleada como del tamaño del barco, reduce apreciablemente la varianza de los residuos. El ajuste no paramétrico de la estacionalidad proporciona un método para tomar en cuenta variaciones en las capturas medias a lo largo del tiempo, más flexible que la simple utilización de medias aritméticas simples por meses de calendario (sistema actualmente empleado). Por otra parte, permite un tratamiento automático adecuado de los meses con pocas observaciones —caso de los periodos Diciembre/Enero, en que el buque de inspección de la Unión Europea típicamente regresa a Ostende—. Permite, finalmente, tomar en cuenta la diferente capacidad de pesca de buques de distintas dimensiones.

El procedimiento puede fácilmente funcionar en tiempo real, dado que los *hails* se transmiten en el día y lo propio puede hacerse con los datos correspondientes a diarios de a bordo inspeccionados.

El procedimiento propuesto, en fin, no sustituye al indispensable control físico de los desembarcos, ni ofrece por sí mismo protección contra declaraciones incorrectas o falseadas en los diarios de a bordo; pero es un instrumento útil, capaz de producir de modo simple estimaciones en una situación en la que un diseño muestral convencional no es factible, y por tanto capaz de mostrar anomalías y de orientar el esfuerzo de inspección.

References

- [1] J.M. Chambers and T.J. Hastie. *Statistical Models in S*. Wadsworth & Brooks/Cole, Pacific Grove, Ca., 1992.
- [2] Pol Chantraine. *The Last Cod Fish*. Jespersen Press, St. John's, Nfld, 1993.
- [3] Manuel do Carmo Gomes. *Predictions under Uncertainty. Fish Assemblages and Food Webs on the Grand Banks of Newfoundland*. ISER, Memorial University of Newfoundland, St. John's, Nfld., 1993.
- [4] R.L. Eubank. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York, 1988.

Abstract

The paper describes an additive, semi-parametric model for catch estimation in the Newfoundland waters. The model only requires data on the presence/absence of the vessels in the fishing grounds, and their size, fishing method and gear. It also takes into account temporal and spatial factors. The evolution over time of the catch is modelled in a nonparametric fashion, as is the influence of vessel size (measured by its length). The results improve on the present system.