

**EL PROBLEMA DE LAS CORRECCIONES MULTIESPECÍFICAS DE LAS CAPTURAS
ATUNERAS. APLICACIÓN DE UN MODELO LOG-LINEAL ITERATIVO
A LAS CAPTURAS VENEZOLANAS DE SUPERFICIE (1987-93)**

Gaertner¹, D., J. Marcano², H. Salazar², L. Astudillo²

¹ *ORSTOM-HEA, B.P. 5045, 34032 Montpellier Cedex 1, France*

² *FONAIAP-CIAPES, Apartado 236, Cumaná 6101, Sucre, Venezuela*

SUMMARY

Analysis was carried out on various methods of multi-species correction of annual catches of tunas of the Venezuelan surface fishery, during the 1987-1993 period, and in particular, the application of an iterative log-lineal model for partially classified tables. Unlike previous analyses, processing was carried out on available information on species composition by commercial category, since the confusion among species seems to be linked to the size of the fish. In order to carry out this correction, two calculation methods were used. The first considers that the samples are unbiased and, consequently, extrapolate the quarterly multi-species composition, within a weight category to the total catch of this category, reported in the logbooks. The second method investigates the possibility of the use of flexible statistical techniques that take into account the particular variability of the data from the tuna fishery, for the acceptable sampling levels. When the quarterly sampling seems insufficient, substitution matrixes are developed by means of a log-lineal model that isolates the year effect (A) for the species effects (E) and category (C), or:

$$\text{Log } m_{xyz} = \mu + A_x + E_y + C_z + EC_{yz}$$

This model, although criticizable statistically, represents a notable improvement in the reconstruction of substitution matrixes. On the other hand, the use of an iterative log-lineal model, based on the "Expectation-Maximization" algorithm, allows the reconstruction of catch matrixes by species from more flexible hypotheses on the sampling than in the first case. In spite of some precautions in its use, this algorithm seems to be well adapted to the internal variability of the tuna fishery data. Notwithstanding, the discrepancies observed between the results obtained in this work and the annual estimates by species carried out previously, without considering the stratification by commercial category, emphasizes the need to evaluate the reliability of this criteria in the landings reported in the logbooks.

RESUMÉ

Cette étude porte sur les méthodes de correction multispecificques des captures annuelles de thonidés de la flotte vénézuélienne de surface durant la période 1987-93, et en particulier sur l'application d'un modèle log-linéaire itératif pour des tables partiellement classifiées. A la différence des travaux précédents, l'information disponible a été analysée au niveau de la composition spécifique par catégorie commerciale, étant donné que les confusions entre espèces semblent être liées à la taille des poissons. Deux méthodes de calcul ont été employées. La première considère que les échantillons sont sans biais et, en conséquence, extrapole la composition multispecificque trimestrielle obtenue dans une catégorie commerciale à la capture correspondante dans les livres de bord. La deuxième méthode est basée sur des techniques statistiques flexibles qui prennent en compte la forte variabilité des données de pêche thonière, pour des niveaux d'échantillonnage acceptables. Quand l'échantillonnage trimestriel apparaît insuffisant, on construit des matrices de substitution à l'aide d'un modèle log-linéaire qui isole l'effet année (A) de ceux relatifs à l'espèce (E) et à la catégorie commerciale (C), soit :

$$\text{Log } m_{xyz} = \mu + A_x + E_y + C_z + EC_{yz}$$

Ce modèle, bien que critiquable sur le plan statistique, représente une notable amélioration pour la construction de matrices de substitution. D'autre part, l'utilisation d'un modèle log-linéaire itératif, basé sur l'algorithme "Expectation-Maximization" permet de reconstruire des matrices annuelles de prises par espèce, à partir d'un corps d'hypothèses plus flexible que dans la première méthode. Malgré quelques précautions d'emploi, cet algorithme semble bien adapté à la variabilité interne des données de pêche thonière. Cependant, les différences observées entre les résultats obtenus dans ce travail et les estimations annuelles par espèces faites antérieurement sans la stratification en catégories commerciales met en évidence la nécessité d'évaluer la fiabilité de ce critère dans les déclarations de captures faites dans les livres de bord

RESUMEN

Se analizan varios métodos de corrección multiespecífica de las capturas anuales de túnidos de la flota venezolana de superficie, durante el período 1987-1993, y en particular la aplicación de un modelo log-lineal iterativo para tablas parcialmente clasificadas. A diferencia de los trabajos anteriores, se procesó la información disponible a nivel de composición específica por categoría comercial, ya que las confusiones entre especies parecen estar muy ligadas al tamaño de los peces. Para llevar a cabo esta corrección, se utilizaron dos métodos de cálculo. El primero considera que los muestreos son sin sesgos y en consecuencia, extrapola la composición multiespecífica trimestral, dentro de una categoría de peso, a la captura total de esta categoría, reportada en las bitácoras. El segundo método investiga la posibilidad del uso de técnicas estadísticas flexibles que tomen en cuenta la variabilidad particular de los datos procedentes de la pesca atunera, para los niveles de muestreo aceptable. Cuando el muestreo trimestral aparece insuficiente, se elaboran matrices de sustitución por medio de un modelo log-lineal que aísla el efecto año (A) de los efectos especie (E) y categoría (C); o sea:

$$\text{Log } m_{xyz} = \mu + A_x + E_y + C_z + EC_{yz}$$

Este modelo, aunque criticable estadísticamente, representa una mejora notable en la reconstrucción de matrices de sustitución. Por otra parte, el empleo de un modelo log-lineal iterativo, basado sobre el algoritmo "Expectación-Maximización", permite la reconstrucción de matrices de capturas por especie a partir de hipótesis más flexibles sobre los muestreos que en el primer caso. A pesar de algunas precauciones de uso, este algoritmo parece estar bien adaptado a la variabilidad interna de los datos de pesca atunera. No obstante, las discrepancias observadas entre los resultados obtenidos en este trabajo y las estimaciones anuales por especie realizadas previamente, sin considerar la estratificación en categoría comercial, ponen en evidencia la necesidad de evaluar la fiabilidad de este criterio en las declaraciones reportadas en las bitácoras.

INTRODUCCION.

El problema de los sesgos que existen entre las declaraciones de capturas por especies, hechas por los capitanes de embarcaciones atuneras y la composición real de estas capturas ha sido objeto de varios estudios. En el Atlántico este, se pueden citar los trabajos de CAYRE (1984) para la flota FIS y de PALLARES (1984) y PALLARES & MAMOLAR (1984) para los barcos españoles; más recientemente, los trabajos de PAGAVINO *et al.*, (1991) y de GAERTNER *et al.*, (1992), para los barcos venezolanos en el Atlántico oeste. De estos estudios resaltan que estos sesgos son de naturaleza comercial (confusión entre especies, o categoría de peso-especies, del mismo precio) y/o morfométrico (igual aspecto físico).

En la zona de pesca de la flota venezolana de superficie (Mar Caribe y zonas adyacentes del Atlántico oeste), las dos principales especies registradas en las bitácoras son el atún aleta amarilla: *Thunnus albacares*, YFT, y el listado: *Katsuwonus pelamis*, SKJ. La ausencia total de varias especies de atunes en las mismas bitácoras, a pesar de que se sospechaba sus presencias en los desembarques, fue uno de los motivos para dar inicio, con carácter de prioridad, a un programa de muestreos multiespecíficos en puerto (GAERTNER *et al.*, 1988). La lista de estas especies mal identificadas incluye el ojo gordo: *Thunnus obesus*, BET; la albacora: *Th. alalunga*, ALB, el atún aleta negra: *Th. atlanticus*, BLF, la carachana: *Auxis sp.*, FRI).

Los primeros análisis sobre el tema se limitaban a una evaluación de la magnitud de los sesgos y a correcciones multiespecíficas sobre el total de la captura (PAGAVINO *et al.*, *op. cit.*; GAERTNER *et al.*, *op. cit.*). El presente trabajo tiene dos objetivos:

1. Procesar la información disponible a una escala más fina que en el pasado; es decir, al nivel de

la composición específica por categoría de peso comercial.

- Investigar las posibilidades del uso de técnicas estadísticas flexibles que tomen en cuenta la variabilidad particular de los datos procedentes de la pesca atunera.

Para llevar a cabo este segundo objetivo se tratará de utilizar un modelo Log-lineal, cuyo procedimiento para evaluar el máximo de verosimilitud de las proporciones de cada especie esta basada en el algoritmo iterativo "EM" (Expectation-maximization) de DEMPSTER et al. (1977). La aplicabilidad de este método, en temas de biología pesquera, ha sido ampliamente descrito por HOENIG y HEISEY (1987).

MATERIALES Y METODOS.

Origen de los datos.

El tipo de muestreo realizado en los puertos atuneros del Atlántico ha sido clasificado como muestreo a dos niveles ("two-stage cluster sampling") por CUEVAS *et al.* (1994). Se puede considerar, en efecto, que las unidades primarias de muestreo son los lances (o las bodegas si no se puede identificar a los lances) y que las unidades elementales de segundo nivel son los peces. El muestreo es al azar en los dos niveles, es decir, que se hacen sin distinción "a priori" de la especie, ya que tienen el doble objetivo de evaluar para cada especie por una parte, su proporción ponderal en la captura y por otra parte, su estructura demográfica. Se utiliza la longitud predorsal (LD1) para el aleta amarilla (YFT) y el Ojo gordo (BET), y la longitud a la horquilla (LH), para las demás especies. El peso total de la muestra se calcula por medio de las diferentes ecuaciones longitud-peso y se extrapola al peso de la bodega de la cual procede. El último paso consiste en sumar las capturas ponderadas, o las frecuencias de tamaño ponderadas, en el estrato espacio-temporal que le corresponde.

En el caso de que la bodega muestreada contiene varios lances, se promedia la ubicación espacio-temporal de estos lances o, según el caso, se da más peso estadístico a la información procedente del lance más importante. Esta simplificación no perjudica los resultados del presente estudio, ya que como ha sido mencionado anteriormente, se utiliza una sola área geográfica y un período de tiempo suficientemente amplio (el trimestre).

De la información reportada en las bitácoras, se utiliza las capturas totales trimestrales por categorías de peso, tal que:

- Categoría 1, para los ejemplares de menos de 3 Kg;
- Categoría 2, para los ejemplares entre 3 y 15 Kg;
- Categoría 3, para los ejemplares de más de 15 Kg.

En efecto, debido a la naturaleza de los sesgos, evocada en la parte introductoria, es lógico suponer que los errores de identificación dependen del peso medio de los peces. Por ejemplo, un mal reporte de las carachanas (FRI) es factible en la categoría 1 (peces de menos de 3 Kg), pero al contrario totalmente improbable en la categoría 3, etc., etc. Cabe mencionar que este último paso

es distinto del método empleado en los estudios anteriores a donde los porcentajes multiespecíficos se aplicaban a la captura total (sin distinción de categoría de peso).

Selección del estrato espacio-temporal sobre el cual se hará las correcciones.

El análisis de la variabilidad espacio-temporal de los muestreos en puerto, realizado por GAERTNER *et al.* (1992), demuestra la existencia de una cierta heterogeneidad entre sectores y de ahí, la dificultad de establecer estratos de igual amplitud en el tiempo. No obstante, debido a la poca extensión de la zona de pesca actual (entre los 10 y 14 grados de latitud Norte, y entre los 55 y 70 grados de longitud Oeste) y a la poca cantidad de muestreos disponibles, una estratificación trimestral, sin discriminación de zona, parece un buen compromiso (una estratificación más fina implicaría más sustitución de muestreos).

En consecuencia, se conservará este espacio de tiempo, ya utilizado en los estudios anteriores, para reconstruir las capturas específicas de la pesca atunera venezolana de superficie por cada arte (BB y PS) entre el 1er trimestre de 1987 y el 4to trimestre de 1993. Las capturas anuales son estimadas a partir de las estadísticas trimestrales corregidas. Este análisis ha sido extendido en una segunda fase a 1986; para la cual se disponía únicamente de capturas totales trimestrales.

El método de cálculo.

Dos métodos de cálculo han sido utilizados. En el primero de ellos se considera que tanto los porcentajes trimestrales de cada especie dentro de una categoría de peso (obtenidos por el muestreo) como las proporciones de cada categoría de peso (o comercial), reportadas en las bitácoras, son sin sesgos. En consecuencia las correcciones se hacen por simple producto entre las dos fuentes de información (porcentaje tomado por la especie "i", en la composición multiespecífica de un trimestre, por la captura total de la categoría de peso "j", declarada en las bitácoras, durante la misma temporada). La única diferencia entre este método y las correcciones anteriores es que aquí no se utiliza la captura total sino el total de la captura de una categoría de peso (deducido de las bitácoras y extrapolado a los desembarques estimados).

En contrapartida, el segundo método, basado sobre modelos Log-lineales aplicados en tablas incompletas, conserva un grado de incertidumbre sobre las dos fuentes de datos. La poca utilización de las variables cualitativas en biología pesquera hace que el empleo de los modelos Log-lineales sea poco usual. Además del trabajo de HOENIG & HEISEY (1987), cuyos ejemplos de utilización van desde la elaboración de claves Edad-Longitud, hasta la clasificación entre varios stocks, se puede citar el estudio de MORAND & LAE (1992), sobre las modalidades de utilización de artes de pesca artesanales en el Delta del Níger (África) y el trabajo de GAERTNER & MEDINA-GAERTNER (1992), sobre los eventos asociados a las faenas de pesca de los cerqueros venezolanos (especies asociadas, lances en blanco, etc.).

El método preconizado por HOENIG & HEISEY (*op. cit.*) está basado sobre la idea de utilizar una primera matriz de referencia en la cual todos los elementos son conocidos (en nuestro caso la matriz construida por los muestreos trimestrales, con 6 especies y 3 categorías-peso) para construir una segunda matriz de la misma dimensión (la matriz real de las capturas por especie y por

categorías). De la segunda matriz, solo se conoce el vector línea, o sea las sumas de las capturas de cada categoría de peso reportadas en las bitácoras (sin distinción de especie, Cf. los $m_{i,j}$ en el gráfico de abajo). Sea:

	Matriz 1 (obtenida con los muestreos)				Matriz 2 (a evaluar)			Vector Solución
	Cat ₁	Cat ₂	Cat ₃	∑ lin.	Cat ₁	Cat ₂	Cat ₃	∑ lin.
Esp.								
1	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,+}$?	?	?	?
"	"	"	"	"	?	?	?	?
i	$n_{i,1}$	$n_{i,2}$	$n_{i,3}$	$n_{i,+}$?	$m_{i,2}$?	$m_{i,+}$
"	"	"	"	"	?	?	?	?
"	"	"	"	"	?	?	?	?
6	$n_{6,1}$	$n_{6,2}$	$n_{6,3}$	$n_{6,+}$?	?	?	?
∑ col	$n_{+,1}$	$n_{+,2}$	$n_{+,3}$	$n_{+,+}$	$m_{+,1}$	$m_{+,2}$	$m_{+,3}$	$m_{+,+}$

Vector de las capturas totales por categorías
(conocido desde las bitácoras)

$$\text{con } n_{+,j} = \sum_i n_{i,j}; \quad n_{i,+} = \sum_j n_{i,j}, \text{ y } n_{+,+} = \sum_i \sum_j n_{i,j}; \quad (\text{idem para m.})$$

Esta estructura inicial corresponde a un modelo Log-lineal asociado a una tabla de contingencia parcialmente clasificada. Para calcular las estimaciones del máximo de verosimilitud de las proporciones en la segunda matriz (2), se empleó el algoritmo "EM" (expectation-maximization) de DEMPSTER *et. al.*, (1977).

HOENIG & HEISEY (*op. cit.*) señalan que el uso del criterio de máximo de verosimilitud es mucho más flexible que el criterio de los mínimos cuadrados utilizado por otros autores (cf CLARK, 1981, por ejemplo). En efecto, este último supone que la matriz 1 de clasificación es conocida perfectamente, sin posibilidad de errores. Al contrario, el criterio de máximo de verosimilitud admite un grado de incertidumbre sobre todos los datos. No se plantea la hipótesis que la distribución de las especies dentro de las categorías-peso quede constante de una matriz a la otra (hipótesis utilizada en el otro método). Aquí se supone únicamente que la probabilidad de que "una tonelada" de la especie *i* cae en la categoría *j* quede constante, o sea:

$$P(j/i)_1 = P(j/i)_2;$$

Estas probabilidades condicionales (probabilidad de estar en la categoría *j*, dada la especie *i*) constituyen las tasas de clasificación. Si se anota a_{ik} , la probabilidad de que una tonelada de la matriz *k* ($k=1,2$) sea de la especie *i*, la probabilidad de que una tonelada de la matriz *k* pertenece, esta vez, no solamente a la especie *i* sino también a la categoría *j*, o sea:

$$P(i,j/k) = P(i/j) * a_{ik}$$

Los modelos Log-lineales son relativamente comparables a un ANOVA en la medida que se trata de medir el eventual efecto de un factor en términos de desviación al valor promedio, pero sobre una celda de una tabla de contingencia y no sobre una variable explicada (ver AGRESTI, 1992; para más detalles sobre modelos Log-lineales, y especialmente sobre los modelos "Logit" para los aspectos de comparación con el ANOVA).

La reformulación de esa estructura bajo la forma de un modelo Log-lineal será:

$$\text{Log (Valor de una celda } i,j,k) = \mu + E_i + C_j + O_k + (E*C)_{ij} + (E*O)_{ik}$$

con: μ = gran media; E, C, O = efectos principales Especie, Categoría, Origen de los peces (muestreos o capturas) y las interacciones entre 2 factores de interés. En realidad, una formulación más acorde con la presentación clásica de los modelos Log-lineales sería:

$$\text{Log } m_{ijk} = \mu + \lambda_i^E + \lambda_j^C + \lambda_k^O + \lambda_{ij}^{(E,C)} + \lambda_{ik}^{(E,O)},$$

$$\text{con : } \sum \lambda_i^E = 0; \sum \lambda_j^C = 0; \sum \lambda_k^O = 0; \sum \lambda_{ij}^{(E,C)} = 0; \sum \lambda_{ik}^{(E,O)} = 0;$$

o de forma abreviada [EC EO].

Las estimaciones del máximo de verosimilitud son los valores de los parámetros que maximizan la función de verosimilitud (por las cuales la probabilidad de obtener los datos observados es máxima). El algoritmo "EM" procede de manera iterativa de modo siguiente:

1) Inicialización.

Se hace una estimación preliminar (iteración 0) de las celdas de la segunda matriz, o sea:

$$m_{ij}^{(0)} = m_{+,j} * n_{i,+} / n_{+,+}$$

2) Maximización.

Se procede, dentro del ciclo de iteraciones, a la fase de maximización con el cálculo del máximo de verosimilitud de $P(j/i)$ y de $a_{i,2}$ ($a_{i,1}$ queda constante dentro del ciclo de iteraciones y no modifica la maximización del kernel del Log de verosimilitud, razón por la cual no se recalcula), sea para la iteración *s*:

$$P(j/i)^{(s+1)} = (n_{i,j} + m_{i,j}^{(s)}) / (n_{i,+} + m_{i,+}^{(s)})$$

$$a_{i,2}^{(s+1)} = m_{i,+}^{(s)} / m_{+,+}$$

de ahí se podría maximizar las estimaciones de los $m_{i,j}$ ($m_{i,j}^{(s+1)}$), ya que:

$$m_{i,j}^{(s+1)} = P(j/i)_2 * a_{i,2} * m_{+,+} = P(j/i) * m_{i,+}$$

pero como lo veremos en 3) este paso no es necesario.

3) Expectación

En efecto, para que la suma sobre las columnas de estos valores maximizados esté conforme con los datos observados (los $m_{.j}$), se reajustan como:

$$m_{ij} = m_{ij}^{(s)} * m_{.j} / m_{ij}^{(s)}$$

De esta manera se puede obtener directamente m_{ij} , con:

$$m_{ij}^{(s+1)} = [P(j/i)^{(s+1)} * m_{i+}^{(s)} * m_{.j}] / [\sum_i \{P(j/i)^{(s+1)} * m_{i+}^{(s)}\}]$$

4) Cálculo del kernel del Log. de verosimilitud (L).

Después de algunas simplificaciones (se eliminan los términos constantes ya que no modifican los cálculos de la fase de maximización de este proceso) el kernel del Log de verosimilitud (parte de la función de verosimilitud que es afectada por los parámetros) se limita a:

$$L = \sum_{i,j} (n_{ij} + m_{ij}^{(s+1)}) * \text{Log } P(j/i)^{(s+1)} + \sum_i m_{i+}^{(s+1)} * \text{Log } a_{i2}^{(s+1)}$$

En el caso de existir algunas celdas nulas, para poder calcular los valores logarítmicos, HOENIG & HEISEY (*op. cit.*) recomiendan agregar un valor de 0.0001, a todas ellas. Este test se puede efectuar al nivel del paso 2.

5: Evaluación.

Se incrementa el número de iteración S tal que S=S+1,

si S=1, se iguala L en L' y se regresa en el paso 2);

si S>1, y |L'-L| > nivel de tolerancia, se carga L en L' y se regresa en 2);

si S>1, y |L'-L| < nivel de tolerancia, se para el proceso de iteraciones y se obtiene el vector de capturas por especie, sumando las estimaciones de la matriz especie*categorías-peso.

RESULTADOS.

El problema de las substituciones.

El sistema de bitácoras de la pesca atunera venezolana de superficie fue instalado en 1987 (GAERTNER *et al.*, 1988). Al inicio se perdieron numerosas bitácoras dando como consecuencia la ausencia de datos para los cerqueros en el 2do trimestre de 1987. Para substituir este estrato faltante, se promediaron las proporciones de cada categoría-peso, reportadas en las bitácoras durante los 2do trimestres, desde 1988 a 1993; estas proporciones fueron estimadas a 0,206, 0,307 y 0,487. Para 1986, año por el cual se disponía únicamente de las capturas totales por trimestre, se procedió al mismo tipo de estimación (% m_{ij} , promediados entre 1987 y 1993).

Por otra parte, las matrices de clasificación trimestrales (n_{ij}) se obtuvieron por agrupación de los muestreos multiespecíficos de cada trimestre entre 1987 y 1993. No obstante, para tratar de eliminar los trimestres con insuficiente información, no se conservaron las matrices trimestrales con

menos de dos muestreos ni las matrices en las cuales la suma de una columna categoría-peso aparece nula, contrariamente a lo indicado en la columna correspondiente extraída de las bitácoras, es decir, cuando $n_{.j} = 0$ y $m_{.j} > 0$ (Tab. 1).

Para estimar las matrices de substitución se recurrió igualmente a los modelos Log-lineales (ver AGRESTI, 1990; KNOKE & BURKE, 1990, para más detalles sobre este tipo de modelos). El programa OCTA (DALLAL, 1987), emplea el algoritmo de verosimilitud de Deming-Stephan, basado sobre una técnica iterativa de ajuste proporcional ("proportional iterative fitting").

Conociendo, a priori, la alta variabilidad entre los factores involucrados (años, categorías, especies y sus respectivas interacciones) y, en algunos casos, la poca cantidad de muestreos por estrato, no es nada sorprendente ver que pocos modelos Log-lineales ajustan los datos de manera significativa (Tab. 2). Es importante destacar que el modelo Log-lineal de más interés para el objetivo planteado (establecer matrices de substitución) es el modelo que aísla el efecto Año (A) de los demás (Especie E y Categoría C), o sea [A, EC]:

$$\text{Log } m_{xyz} = \mu + A_x + E_y + C_z + EC_{yz}.$$

Para una mejor ilustración del objetivo planteado, se puede hacer el paralelismo con la constitución de índices anuales de abundancia, a partir de varios CPUE locales, con el uso de un modelo lineal general. En este tipo de análisis, se busca un compromiso entre el mejor ajuste, por una parte y una formulación menos eficaz estadísticamente pero más interpretable, por otra parte. Aquí también se trata de utilizar un modelo que deja el efecto Año sin interacciones con los demás efectos, pero en este caso, para elaborar matrices trimestrales multiespecífica susceptibles de substituir la información faltante en algunos trimestres.

En la tabla 2 se nota que este modelo es significativo sólo en el caso del 4to trimestre para los cañeros. En algunos casos hay que agregar una interacción (ya sea la interacción AE, o la CA) y a veces las dos, para tener un ajuste satisfactorio estadísticamente. No obstante dado que L^2 es proporcional a N (el tamaño total de la tabla), el uso de los modelos Log-lineales sobre juegos de datos importantes tiene tendencia a seleccionar únicamente los modelos con muchos parámetros, aun cuando algunas de las interacciones de nivel superior tienen poca significación (KNOKE y BURKE, 1990). Para medir el mejoramiento de ajuste de estos modelos mas completos (sup.), con respecto al modelo de base (bas.), que es [A EC], se puede utilizar la estadística F, tal que:

$$F = 100 \{ L^2_{(bas.)} - L^2_{(sup.)} \} / L^2_{(bas.)}$$

Estos autores consideran que cuando F supera 90 %, el modelo alternativo puede ser considerado como mas explicativo que el modelo de base. Con la excepción del 4to trimestre para los cerqueros, la observación de la última columna de la tabla 2 permite rechazar esta hipótesis; además, que estos modelos no permitan reconstruir las matrices de substitución "ideales" (o, al menos, independientes del efecto A). Por otro lado, la variabilidad interanual de los muestreos es relativamente importante y no se puede decir que las composiciones multiespecíficas por categorías de peso son totalmente independientes del factor Año (A).

A pesar de su limitación desde un punto de vista meramente estadístico, las estimaciones obtenidas con el modelo [A EC] no están tan alejadas de los valores observados. Para dar un ejemplo, en la tabla 3 se calcularon las estimaciones de este modelo para las estadísticas de los cerqueros en el 4to trimestre entre 1987 y 1993. Los valores estimados se obtuvieron a partir de los parámetros estimados sobre los efectos E, C y EC; o sea:

$$m_{yz} = e^{(\mu + E_y + C_z + E_{Cyz})}$$

Si se considera que el objetivo a este nivel del estudio es únicamente reconstruir matrices de sustitución, se puede observar que las estimaciones para cada especie-categoría son muy cercanas a los valores observados; con la excepción, quizás, de los pequeños listados, cuya proporción casi siempre es sobrestimada por el modelo Log-lineal. El hecho de que el mal ajuste de una sola celda perjudica el ajuste global de una tabla multifactorial es fácil de resolver desde un punto de vista técnico. En efecto, en esta situación se recomienda excluir del ajuste a esta celda (cuya estimación puede hacerse por medio de un promedio sencillo) y reajustar la tabla con el mismo modelo (AGRESTI, 1990). Otra mejora posible es la de "fijar" las celdas correspondientes a los casos de "ceros estructurales" (por ejemplo, nunca se van a observar carachanas de más de 15 kg). Aunque, como se puede notar en la tabla 3, este último problema no afecta las estimaciones de las celdas, el hecho de tomar en cuenta unos "ceros estructurales" reduciría el número de parámetros a evaluar y optimizaría los tiempos de cálculo.

En consecuencia, hemos considerado que este modelo, aunque criticable estadísticamente, representa una mejora notable para la reconstitución de matrices de sustitución (comparativamente a un promedio sencillo de las proporciones observadas en cada trimestre, en cada categoría). Estas proporciones estimadas son dadas en la tabla 4.

Comparación entre las técnicas basadas en el algoritmo Log-lineal iterativo y la corrección por ajuste proporcional directo (en cada categoría).

El análisis comparativo entre las capturas anuales de Venezuela, reportadas por el sistema de corrección global (información utilizada actualmente en los "Data Records") y corregidas de manera sencilla (a partir de las proporciones de cada especie en cada categoría de peso) o con el método "E-M", aparece en la tabla 5.

Si se considera el caso de las dos especies principales: el aleta amarilla y el listado (fig. 1) se nota que la introducción del sistema de ponderación "categoría-peso" conlleva a una subestimación del primero y una sobrestimación del segundo, con respecto a las estimaciones respectivas hechas bajo el anterior sistema de las "correcciones globales" (sin las categorías comerciales). Esta discrepancia es más nítida cuando se emplea el algoritmo E-M. Es posible que el uso más frecuente de matrices de sustitución (hemos visto que el modelo Log-lineal [A CE] tiene tendencia a sobrestimar los pequeños SKJ) y/o la falla de la hipótesis: $P(j/i)_1 = P(j/i)_2$ para algunos trimestres, podría explicar que los sesgos sean más marcados en el segundo método.

DISCUSION

Aunque realizado sobre una temporada muy breve (entre fines de 1987 y principio de 1989) el primer análisis sobre las discrepancias entre los muestreos multiespecíficos y las capturas anotadas en las bitácoras, plantea no solamente la probable declaración de especies menores (en peso o en valor comercial, como el aleta negra y las carachanas) como listado (*Katsuwonus pelamis*, SKJ), pero también reporta una porción de atún aleta amarilla (*Th. albacares*, YFT) como listado (PAGAVINO *et al.*, 1991). El análisis descrito por GAERTNER *et al.* (1992) confirma esta hipótesis y propone unas correcciones para los datos de los años 1988, 1989 y 1990.

El presente estudio tiene un enfoque distinto y trata de estimar el problema del uso de la categoría comercial como factor de ponderación de las capturas multiespecífica.

A la vista de estos resultados se puede concluir que el hecho de ponderar, o no, la composición multiespecífica al nivel de las proporciones tomadas por las categorías de peces "pequeños, medianos y grandes" en las bitácoras, tiene una gran influencia sobre el resultado final; quizás mayor que el hecho de conservar, o no, una variabilidad en las matrices multiespecíficas trimestrales.

Para las pesquerías de gran variabilidad, como es el caso de la pesca atunera, el empleo de este tipo de algoritmo, basado en el criterio de verosimilitud, es de gran interés ya que permite conservar un grado de incertidumbre sobre la matriz de clasificación (muestreos multiespecíficos en nuestro caso). En efecto, es evidente que para los niveles de muestreo "técnicamente aceptables", la pesca atunera ha sido siempre caracterizada por su alta variabilidad. En consecuencia, la integración de este concepto de variabilidad dentro de los métodos estadísticos que permiten hacer la inferencia entre el muestreo y la población, aparece como una tentativa a promoverlo. No obstante, el uso de este algoritmo necesita ser usado con cautela y tomar en consideración las siguientes recomendaciones:

En el caso de la ausencia de una especie en la matriz trimestral es recomendable eliminar esa línea de la matriz correspondiente, para acelerar la convergencia y mejorar la precisión de las demás estimaciones.

Es necesario agregar un criterio de salida del procedimiento iterativo, en la parte "5. Evaluación" del algoritmo E-M (Cf. Material y método), cuando $L^{s+1} > L^s$. Es posible, en efecto, que a pesar de alcanzar un mínimo, el algoritmo continúe la búsqueda del criterio de convergencia. Para ilustrar este punto, en la figura 2 se ha representado la evolución del logaritmo de verosimilitud "L" en función del número de iteración "S", a partir de los datos publicados, p.242, por HOENIG & HEISEY (1987). Se puede ver que a pesar de alcanzar un mínimo a S=33, el algoritmo continúa su búsqueda hasta alcanzar una convergencia a S=183.

En este trabajo, el hecho de utilizar únicamente el criterio de búsqueda de la convergencia ($|L'-L| < \text{nivel de tolerancia}$) conducía, a veces, a soluciones totalmente irreales. Para dar un ejemplo, un buen ajuste aparente (los valores residuales estaban cerca de cero) podría dar estimaciones superiores en una especie poco representada en el muestreo y al contrario subestimar una especie presente en las 3 categorías de peso muestreadas. Para que las condiciones de uso "razonable" de este algoritmo sean llenadas, en el futuro sería importante limitarse a los casos en los cuales los $n_{i,j}$

son comparables en proporción a los m_{ij} (el % de una categoría peso en la muestra tiene que ser comparable con su proporción en las bitácoras).

Finalmente, como en muchos métodos que emplean el criterio de verosimilitud, el alcance de un mínimo local es difícil de detectar.

CONCLUSIONES.

Aunque es lógico suponer que la naturaleza de los sesgos depende del tamaño de los atunes, es difícil saber si las proporciones de cada categoría reportada en las bitácoras es suficientemente precisa para su uso en las correcciones multiespecífica. Es posible, en efecto, que el problema de algunas fallas en el uso del algoritmo E-M sea debido a algunas discrepancias de las proporciones de cada categoría entre las bitácoras y los muestreos.

Si, por el contrario, se considera que los capitanes de pesca reportan de manera conveniente el peso medio de la captura de cada lance y si, en consecuencia, se adopta el método de ponderación al nivel de cada categoría, entonces el uso del algoritmo "E-M", bajo algunas precauciones de uso parece justificarse. De igual manera el empleo de modelos Log-lineales para la construcción de matrices de sustitución, aunque criticable desde un punto de vista meramente estadístico, es un logro que debería ser más empleado en el futuro.

BIBLIOGRAFIA.

- AGRESTI, A., (1990).- *Categorical Data Analysis*. Ed. John Wiley & Sons, New York. 558 p.
- CAYRE P. (1984).- Procédure suivie pour la révision de la composition spécifique des statistiques thonnières FISM (France, Cote d'Ivoire, Sénégal et Maroc). Coll. Vol. Sci. Pap. ICCAT 21 (2): 102-107.
- CLARK, W.G., (1981).- Restricted least-squares estimates of age composition from length composition. Can. J. Fish. Aquat. Sci. 38: 297-307.
- CUEVAS A., DE LA HORRA J., & ROMO J. (1994).- Statistical analysis of tuna fisheries in eastern Atlantic. A preliminary report. Doc. ICCAT SCRS/94/9, 48 p.
- DALLAL, G.E., (1987).- OCTA. A program for the Log-linear analysis of multidimensional contingency tables. Am. Stat. 41: 328.
- DEMPSTER A.P., LAIRD N.M. & RUBIN D.B. (1977).- Maximum likelihood via the EM algorithm. Journ. Royal Stat. Soc. Ser. B Method. 39: 1-22.
- GAERTNER D., CASTILLO C., GUTIERREZ X., SALAZAR H., RODRIGUEZ O. y ASTUDILLO L. (1992).- Nota sobre los sesgos entre las bitácoras y los muestreos multiespecíficos de la flota atunera venezolana de superficie. Col. Vol. Sci. Pap. ICCAT 38:254-261
- GAERTNER D. y MEDINA-GAERTNER M., (1992).- An overview of the relationship between tunas and floating objects in the south of Caribbean Sea. In International workshop on fishing for tunas associated with floating objects. IATTC, La Jolla February 11-14, 1992. 20 pp.
- GAERTNER D., MEDINA-GAERTNER M., CASTILLO C., y MARTINEZ L. (1988).- Instalación de un sistema de bitácoras para la pesca atunera de superficie en Venezuela y análisis de los primeros resultados. Col. Vol. Sci. Pap. ICCAT 28/ 130-140.
- HOENIG J.M., & HEISEY D. M. (1987).- Use of a Log-Linear model with the EM algorithm to correct estimates of stock composition and to convert length to age. Trans. Am. Fish. Soc. 116: 232-243.
- KNOKE D. & BURKE P. J. (1990).- *Log-linear Models*. Sage University Paper series on Quantitative Applications in the Social Sciences 07-001. Beverly Hills & London. Sage Pubns. 80 p.
- MORAND P. et LAE R. (1992).- Vers un traitement approfondi de l'information qualitative des enquêtes halieutiques: le modèle Log-linéaire et l'analyse logistique, appliquées au delta central du Niger. In, Seminfo 5: Statistique Impliquée. Textes réunis par F. Laloé. Ed. ORSTOM, Coll. Colloques et séminaires p.343-354.
- PAGAVINO M., GAERTNER D., SALAZAR H., ASTUDILLO L. y CASTILLO C. (1991).- Composición de las capturas atuneras de superficie de Venezuela deducida a partir de muestreos multiespecíficos en puerto. In Report of the Yellowfin Year Program. Coll. Vol. Sci. Pap. ICCAT 36:551-563.
- PALLARES P. (1984).- Nota sobre el método seguido para la corrección de las estadísticas españolas. Coll. Vol. Sci. Pap. ICCAT, 21 (2): 142
- PALLARES P. & MAMOLAR (1984).- Análisis comparativo de los tipos de bancos de tunidos del Golfo de Guinea según los datos provenientes de los muestreos y de los cuadernos de pesca (1979-1983). Coll. Vol. Sci. Pap. ICCAT, 21 (2): 184-185

	BB							PS						
	87	88	89	90	91	92	93	87	88	89	90	91	92	93
T1	(0)	4	(0)	4	(1)	5	(0)	(0)	(0)	7	3	14	10	27
T2	(0)	2	(1)	(0)	9	(1)	2	(0)	4	8	9	3	4	17
T3	(1)	5	(1)	9	3	4	3	(0)	(4)	8	17	10	20	43
T4	(0)	(1)	6	(1)	2	6	6	2	8	9	7	6	3	89

Tabla 1 Repartición de los muestreos en el tiempo. Cuando se realizaron menos de 2 muestreos, o cuando se encontraron capturas dentro de una categoría de peso (bitacoras) que no estaba representada en el muestreo, se hace la substitución; casos identificables por los ().

Arte	Trim.	Modelo	L2	g.l.	p	F %
BB	I	A CE	78,60	51	0,008	59,62
		CE EA	31,74	36	0,672	
	II	A CE	135,23	68	0,000	69,22
		CE EA	41,62	48	0,730	
	III	A CE	191,46	102	0,000	73,81
		CE CA	50,14	90	0,999	
	IV	A CE	70,44	85	0,872	*****
			****	*****	***	
PS	I	A CE	703,01	68	0,000	77,67
		CE CA EA	157,19	40	0,000	
	II	A CE	578,39	85	0,000	88,38
		CE CA EA	67,21	50	0,053	
	III	A CE	934,67	85	0,000	85,50
		CE CA EA	135,50	50	0,000	
	IV	A CE	724,63	102	0,000	92,67
		CE CA EA	53,28	60	0,718	

Tabla 2. Estadísticas comparativas entre varios modelos Log-lineales utilizados para construir las matrices de substitución de los muestreos; los efectos estudiados son : A = año, C = categoría-peso, E = especie (Ver texto).

Especie	Categ.	1987			1988		
		C 1	C 2	C 3	C 1	C 2	C 3
YFT	Obs.	2.1	38.7	46.7	2.4	293.2	244.2
	Calc.	1.2	44.7	51.1	5.7	223.1	254.6
SKJ	Obs.	20.7	20.0	0.0	59.4	53.6	0.0
	Calc.	22.1	17.3	0.0	110.0	86.2	0.0
FRI	Obs.	2.8	0.0	0.0	9.2	0.2	0.0
	Calc.	2.2	0.0	0.0	11.0	0.1	0.0
ALB	Obs.	0.0	0.0	0.0	0.0	0.0	14.3
	Calc.	0.0	0.0	0.7	0.0	0.0	3.4
BET	Obs.	0.0	0.0	0.0	0.0	1.0	3.1
	Calc.	0.1	0.1	0.0	0.6	2.6	0.3
BLF	Obs.	0.0	14.4	0.0	5.6	41.0	0.0
	Calc.	1.0	5.0	0.0	4.8	24.9	0.0

Especie	Categ.	1989			1990		
		C 1	C 2	C 3	C 1	C 2	C 3
YFT	Obs.	2.5	198.1	372.0	5.7	244.3	271.3
	Calc.	5.6	218.3	249.2	5.7	220.6	251.8
SKJ	Obs.	31.8	69.0	0.0	50.1	74.2	0.0
	Calc.	107.6	84.4	0.0	108.7	85.3	0.0
FRI	Obs.	14.1	0.0	0.0	21.4	0.2	0.0
	Calc.	10.7	0.1	0.0	10.8	0.1	0.0
ALB	Obs.	0.0	0.0	0.0	0.0	0.0	5.1
	Calc.	0.0	0.0	3.3	0.0	0.0	3.3
BET	Obs.	0.0	2.3	0.0	0.5	4.8	0.0
	Calc.	0.6	2.6	0.3	0.6	2.6	0.3
BLF	Obs.	2.7	19.2	0.0	5.8	35.7	0.0
	Calc.	4.7	24.4	0.0	4.8	24.7	0.0

Especie	Categ.	1991			1992		
		C 1	C 2	C 3	C 1	C 2	C 3
YFT	Obs.	2.8	245.1	306.3	1.3	127.8	89.8
	Calc.	5.5	215.3	245.7	2.3	88.9	101.4
SKJ	Obs.	27.5	82.3	0.0	25.7	12.5	0.0
	Calc.	106.1	83.2	0.0	43.8	34.3	0.0
FRI	Obs.	10.1	0.0	0.0	0.2	0.0	0.0
	Calc.	10.5	0.1	0.0	4.4	0.0	0.0
ALB	Obs.	0.0	0.0	0.0	0.0	0.0	0.0
	Calc.	0.0	0.0	3.2	0.0	0.0	1.3
BET	Obs.	0.9	1.9	0.0	0.0	0.0	0.0
	Calc.	0.5	2.6	0.3	0.2	1.1	0.1
BLF	Obs.	6.5	18.2	0.0	1.8	30.2	0.0
	Calc.	4.7	24.1	0.0	1.9	9.9	0.0

Especie	Categ.	1993		
		C 1	C 2	C 3
YFT	Obs.	39.1	1023.0	1146.6
	Calc.	29.8	1159.2	1323.2
SKJ	Obs.	854.5	527.0	0.0
	Calc.	571.4	448.0	0.2
FRI	Obs.	48.4	0.7	0.0
	Calc.	56.7	0.7	0.2
ALB	Obs.	0.0	0.0	13.0
	Calc.	0.2	0.2	17.4
BET	Obs.	3.9	15.6	0.0
	Calc.	2.9	13.7	1.8
BLF	Obs.	24.5	83.8	0.0
	Calc.	25.0	129.6	0.2

Tab. 3. Comparación entre las capturas observadas y las capturas estimadas por medio del modelo Log-lineal [A CE] para los cerqueros (PS) en el 4to trimestre (1987-1993).

Art	Tri	Cat	YFT	SKJ	FRI	ALB	BET	BLF
BB	I	C 1	1,326	4,063	0,037	0,037	0,096	0,780
		C 2	4,446	17,150	0,037	0,037	0,407	1,587
		C 3	20,126	0,037	0,037	0,037	0,169	0,189
	II	C 1	2,838	2,394	0,058	0,034	0,034	0,063
		C 2	11,740	4,632	0,034	0,034	0,365	1,327
		C 3	23,641	0,143	0,034	0,144	0,567	0,034
	III	C 1	0,150	2,924	0,057	0,040	0,040	0,238
		C 2	43,075	5,018	0,040	0,040	0,096	1,381
		C 3	17,340	1,054	0,040	0,040	0,908	0,040
	IV	C 1	0,433	3,360	0,051	0,035	0,062	0,116
		C 2	42,606	6,436	0,035	0,035	1,422	3,497
		C 3	15,831	0,108	0,034	0,035	0,260	0,100
PS	I	C 1	21,911	69,895	32,040	0,079	6,341	7,822
		C 2	96,255	55,534	0,222	0,041	4,559	31,722
		C 3	506,234	0,476	0,041	12,897	12,146	0,243
	II	C 1	17,409	69,200	21,693	0,043	0,453	4,084
		C 2	69,200	22,579	0,466	0,042	0,360	11,212
		C 3	360,323	0,042	0,042	8,474	0,120	0,216
	III	C 1	3,721	118,392	30,084	0,039	2,042	4,683
		C 2	279,779	4,204	0,373	0,037	7,345	49,600
		C 3	366,501	0,568	0,038	4,195	1,775	0,038
	IV	C 1	5,029	96,931	9,621	0,032	0,494	4,238
		C 2	195,195	75,490	0,121	0,032	2,326	21,846
		C 3	222,294	0,032	0,032	2,956	0,302	0,032

Tabla 4. Valores estimados por medio de los modelos Log-lineales [A CE], para la construcción de las matrices de sustitución de los muestreos trimestrales multispecíficos por categorías de peso (C 1 < 3 Kg; C 2 entre 3 y 15 Kg; C 3 > 15 Kg) y por los 2 artes de superficie (BB y PS).

A partir del algoritmo E-M y de un minimo de 2 muestras por trimestre

	1986	1987	1988	1989	1990	1991	1992	1993
BB								
YFT	1735,4	2968,1	3185,6	2482,9	2870,8	2780,5	2289,6	2812,8
SKJ	1384	1519,8	2540,3	1568,5	973,2	3031	1708,3	1196,7
FRI	3,1	3,6	0	15,7	0,5	0	1,9	0
ALB	0,6	0,2	0	13	0	0,3	3,9	0
BET	125,2	114	209,9	329,6	154,9	133,6	36,3	120,9
BLF	930,3	1084,8	928,4	622,2	818,6	607,1	755,9	624,5

	1986	1987	1988	1989	1990	1991	1992	1993
PS								
YFT	5310,6	5370,3	4136,4	10278,9	4814,9	8348,7	7310,9	9597,8
SKJ	5114,2	4936,2	3227,6	3234,5	3955,1	8950	8297,5	10781,8
FRI	448,6	840,6	431	388	1142,2	1910,4	319,9	695,1
ALB	25,4	21	14	4,5	32,5	36,2	2,8	46,8
BET	175,6	192,5	108	43,8	66,4	412,9	172,2	733,9
BLF	1584,4	1760	1876,9	1169,2	1708	1839,5	2881,8	3178,8

A partir de los % por especies y para cada categoria de peso

	1986	1987	1988	1989	1990	1991	1992	1993
BB								
YFT	2887,4	3907,8	4226,1	3457,5	3765,3	4190,3	3615,9	3295,7
SKJ	991,5	1420,8	2198,8	1293,2	777,3	1951,6	941	1122,5
FRI	5	5,8	1,5	9,4	1,3	0	5,2	0
ALB	2,4	3,3	0	5,5	0,6	2,3	3,9	0
BET	82,6	94,3	118,3	65,7	58,8	55,7	86,8	122,5
BLF	209,8	268,7	310,6	200,7	214,8	357,1	243,1	214

	1986	1987	1988	1989	1990	1991	1992	1993
PS								
YFT	7276,8	6582,9	5991,9	11612,2	6533,2	11966,9	9693,3	12659,3
SKJ	3977,3	4329,4	2309,5	2429,5	3014,4	6185,6	6892,7	10048,9
FRI	489,5	816,5	439,6	598,6	1067,1	1710,2	322,4	881,2
ALB	58	42,9	23,9	7,8	56,1	73,3	3,4	111,3
BET	120,7	170,4	100,8	21,5	53	320,5	168,6	325,9
BLF	736,5	1178,7	928,7	449,5	935,2	1241	1905,1	1006,7

A partir de los % por especies y la captura total

	1986	1987	1988	1989	1990	1991	1992	1993
BB								
YFT	1409	2285	4395	3611	3941	5089	3843	3741
SKJ	1819	2068	2132	1168	731	1075	665	688
FRI		31		1		0	11	0
ALB					1	4	0	0
BET	23		156	9	2	152	93	156
BLF		44	171	243	444	235	187	170

	1986	1987	1988	1989	1990	1991	1992	1993
PS								
YFT	8161	8303	6759	12234	8628	16110	13575	15739
SKJ	5548	5512	1717	1834	1774	3243	3357	7433
FRI		356	303	479	1969	812	191	523
ALB	22		97	12	1	142	7	228
BET			40	20	64	183	129	355
BLF		71	879	490	71	706	1727	734

Tab. 5. Comparacion entre varios metodos de correccion de las capturas venezolanas de superfi

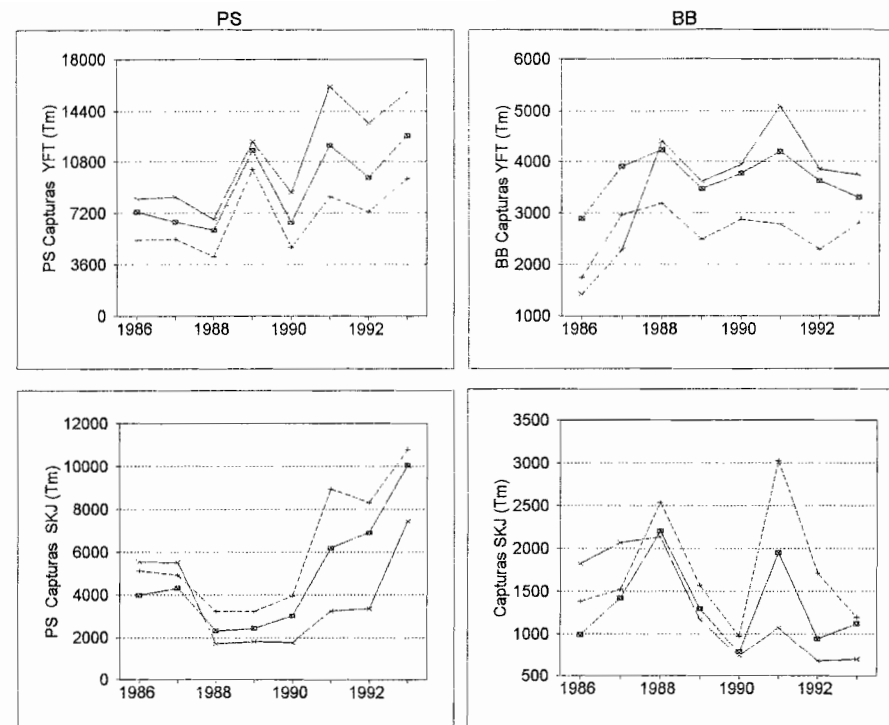


Fig. 1 Capturas anuales de los artes de superficie venezolanos para el YFT y el SKJ en funcion del tipo de correccion efectuado: con el Algoritmo E-M (+ - - - +), con los % por especie por cada categoria comercial (■ - - - ■), y con los % por especie sobre la Captura total (x - - - x).

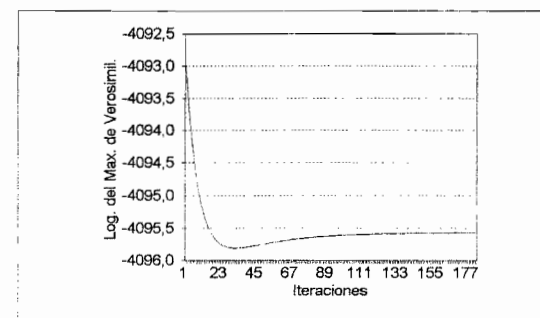


Fig. 2.- Evolucion del Log de verosimilitud en funcion del numero de iteraciones; a partir del trabajo de HOENIG et HEISEY (1987); p 242.