

COMISIÓN INTERAMERICANA DEL ATÚN TROPICAL

COMITÉ CIENTÍFICO ASESOR

SEGUNDA REUNIÓN

La Jolla, California (EE.UU.)

9-12 de mayo de 2011

DOCUMENTO SAC-02-10

**ESTIMADORES POSESTRATIFICADOS DE LA CAPTURA TOTAL
PARA LOS DATOS DE MUESTREO EN PUERTO DE LA PESQUERÍA DE
CERCO**

Cleridy E. Lennert-Cody, Mark N. Maunder, Patrick K. Tomlinson, Alexandre Aires-da-Silva, Alejandro Pérez, Marti McCracken

1.	Resumen	1
2.	Introducción	1
3.	Antecedentes	2
3.1.	Recolección de datos.....	2
3.2.	Estimadores actuales	3
4.	Método propuesto	4
4.1.	Estimadores posestratificados candidatos	4
4.2.	Selección del estimador específico.....	4
4.3.	Factibilidad	5
5.	Trabajo futuro	6
5.1.	Tratamiento de estratos de muestreo con captura pero muestras insuficientes	6
5.2.	Estimación de la varianza	6
5.3.	Evaluación de la factibilidad para los datos de muestreo en puerto de las pesquerías cañera y de cerco antes de 2000	6
	Reconocimientos.....	6
	Referencias	6

1. RESUMEN

Aunque las metas de la estratificación para evaluación de poblaciones y la recolección de datos a menudo concuerdan, es útil poder considerar distintas estratificaciones para ambos, particularmente cuando las características de las pesquerías han cambiado a lo largo del tiempo. Para poder hacer esto, es necesario desarrollar un estimador de los totales de las pesquerías basado en la recolección posterior de estratos de datos. Se presentan dos posibles estimadores posestratificados para las especies capturadas y la composición por tamaño de los datos recientes de la pesquería de cerco. Se comenta también un método para elegir entre los dos. En el futuro se aplicarán estos métodos a los datos de cerco de 2000 a la actualidad.

2. INTRODUCCIÓN

Se usa la estratificación en las evaluaciones de las poblaciones para tratar las diferencias en la dinámica de las poblaciones y de la pesca. En general, los datos de la pesca (captura, CPUE, y composición por edad/tamaño) son estratificados (después de ser recolectados los datos) para apoyar el supuesto que los parámetros relacionados con la pesca (capturabilidad y selectividad) son constantes a lo largo del tiempo. Se puede usar la estratificación también durante la recolección de los datos para procurar evitar asigna-

ciones asimétricas de las muestras (que podría llevar a sesgos) y para minimizar la varianza de los estimadores de la población total (por ejemplo, Holt y Smith 1979; Thompson 1992). Por lo tanto, las metas de la estratificación para la evaluación de poblaciones y la recolección de datos a menudo concuerdan. No obstante, pueden ser diferentes si las características de las pesquerías han cambiado a lo largo del tiempo.

Actualmente, las evaluaciones de las poblaciones de todas las especies de atunes en el Océano Pacífico oriental (OPO) (por ejemplo, Maunder y Aires-da-Silva 2010; Aires-da-Silva y Maunder 2010) usan áreas grandes formadas al juntar los estratos espaciales de muestreo (Figura 1a-b). Sin embargo, debido a cambios recientes en la pesquería de cerco, es deseable poder considerar particiones espaciales alternativas del área del OPO de la pesquería construidas con unidades diferentes de los estratos de muestreo. Por ejemplo, la pesquería de cerco sobre objetos flotantes en el OPO se ha extendido considerablemente mar afuera desde principios de los años 1990 (Watters 1999). Por contraste, los estratos de muestreo usados para las pesquerías de superficie en el OPO fueron desarrolladas principalmente a fines de los años 1960 (Suter 2008, y sus referencias), cuando la pesquería era más costera (Watters 1999) y dominada por capturas de atunes aleta amarilla y barrilete con red de cerco sobre atunes asociados con delfines y en cardúmenes no asociados, y también por buques cañeros. Aunque estos estratos de muestreo fueron refinados a fines de los años 1990 (Suter 2008), podrían ahora no ser óptimos para uso en la evaluación en las pesquerías con un componente fuerte de alta mar, tales como la pesquería de cerco sobre objetos flotantes, que produce las mayores capturas de atún patudo (Anónimo 2010).

La posestratificación (por ejemplo, Holt y Smith 1979; Valliant 1993) es una técnica usada en los análisis de datos para agrupar las muestras, después de ser recolectados los datos, cuando se desean estimaciones de la población total para grupos cuya definición no forma expresamente parte del protocolo de recolección de datos. Se presentan dos posibles estimadores posestratificados de la captura total, por especie y tamaño, para los datos de muestreo en puerto recolectados desde 2000. Se propone un procedimiento de análisis de datos para elegir entre los dos. En el presente se supone que ya existen los posibles posestratos. Se podría definir estos posestratos, por ejemplo, con base en análisis de patrones espaciotemporales en las tendencias de la frecuencia por talla y/o CPUE (por ejemplo, Lennert-Cody *et al.* 2010). No obstante, se podría definir los posestratos sobre la base de otros datos y criterios también.

3. ANTECEDENTES

3.1. Recolección de datos

Datos sobre la composición por especie y tamaño de captura de atunes por los buques de cerco son obtenidos cuando los buques llegan a puerto para descargar (Tomlinson 2004; Suter 2008). Para intentar asegurar que las muestras tomadas sean representativas de la pesquería entera, se divide la pesquería de cerco en categorías, o ‘estratos de muestreo’. Se definen estos estratos de muestreo por el área de pesca (13 áreas, Figura 1a), el mes de pesca y la modalidad de pesca (6 modalidades basadas en el tipo de lance de cerco y el tamaño del buque), para un total de 936 posibles estratos. No todos los estratos contienen actividad de pesca en un año dado. Las muestras son tomadas por estrato conforme a un método de ‘dos etapas’, en el cual las bodegas del buque son la primera etapa, y los peces en la bodega la segunda etapa. Ya que no se sabe por adelantado el número de bodegas en un estrato, y ya que algunos buques podrían descargar en puertos en los que la logística hace que el muestreo sea prohibitivamente difícil, las bodegas por muestrear son seleccionados oportunamente. No obstante, se muestrea una bodega solamente si toda la captura que contiene proviene del mismo estrato de muestreo (o sea, la misma área, mes, y modalidad de pesca). Durante el transcurso de un año, se muestrearán números desiguales de bodegas por estrato. Una vez seleccionada la bodega de un buque, se muestrean peces individuales de la bodega a medida que se descarga la captura. Se mide la talla de un número de peces de cada especie (típicamente 50). De la misma bodega, e independientemente de los peces medidos, se cuenta la composición por especies de varios centenares de peces. Se seleccionan los peces muestreados de la bodega uno a la vez, a partir de un punto de partido establecido oportunamente, según lo permitan las circunstancias. En el Anexo de Su-

ter (2008) se presentan detalles de los procedimientos de toma de datos del muestreo en puerto.

3.2. Estimadores actuales

Los estimadores actuales de la captura total por especie y tamaño (Tomlinson 2004) tienen la forma general de un estimador de razón del estrato total (por ejemplo, Thompson 1992) basado en la cantidad de la captura en las bodegas muestreadas. La captura anual estimada total (en peso) de la especie i en el estrato de muestreo h es expresada de esta forma:

$$\begin{aligned} \widehat{W}_{hi} &= W_h \widehat{p}_{hi} \\ &= W_h \left[\frac{\sum_{j=1}^q W_{hj} \left(\frac{w_{hij} n_{hij}}{m_{hij} n_{h,j}} \right)}{\sum_{j=1}^q W_{hj}} \right] \\ &= W_h \frac{[\sum_{j=1}^q W_{hj} \cdot g_1(w_{hij}, m_{hij}, n_{hij}; i = 1, \dots, \#sp)]}{[\sum_{j=1}^q W_{hj}]} \end{aligned} \quad (1)$$

donde W_h es el peso total de todas las especies combinadas en el estrato de muestreo h (que se supone conocido), \widehat{p} es la estimación de la fracción de especies (derivada del peso) en el estrato, W_{hj} es el peso total de todas las especies combinadas en la j -ésima bodega muestreada del estrato de muestreo h (que igualmente se supone conocido), $j=1, \dots, q$ bodegas muestreadas, w es la suma de los pesos de los peces muestreados (convertidos de tallas), m es el número de peces muestreados, n es el número de peces contados para la composición por especies, y g_1 representa la función de los promedios y proporciones de las muestras entre paréntesis (o sea, una función de los w , m y n solamente).

Similarmente, el número total estimado de peces de la especie i en el intervalo de talla k del estrato de muestreo h se expresa de esta forma:

$$\begin{aligned} \widehat{N}_{hik} &= \widehat{N}_{hi} \widehat{f}_{hik} \\ &= W_h \left[\frac{\sum_{j=1}^q W_{hj} \left(\frac{m_{hijk} n_{hij}}{m_{hij} n_{h,j}} \right)}{\sum_{j=1}^q W_{hj}} \right] \\ &= W_h \frac{[\sum_{j=1}^q W_{hj} \cdot g_2(w_{hij}, m_{hijk}, m_{hij}, n_{hij}; i = 1, \dots, \#sp)]}{[\sum_{j=1}^q W_{hj}]} \end{aligned} \quad (2)$$

donde \widehat{N} es la estimación del número total de peces en el estrato, \widehat{f} es la estimación de la fracción de especies, derivada del número de individuos, y g_2 representa una función de los promedios y proporciones de las muestras solamente. Por lo tanto, la ponderación es la misma para ambos estimadores; son diferentes solamente en la forma de g , la función de los promedios y proporciones de las muestras.

4. ENFOQUE PROPUESTO

4.1. Estimadores posestratificados candidatos

En esta sección se presentan dos candidatos de estimadores posestratificados de la captura por especies en el posestrato c que conservan la función específica g_i de la ecuación (1). Lo siguiente trata específicamente los estimadores posestratificados de la captura total por especie; los estimadores posestratificados del número total de peces en un intervalo de talla (ecuación (2)) seguiría mediante el reemplazo de g_1 con g_2 .

El primer estimador posestratificado de la captura total por especie fue desarrollado bajo el supuesto que tanto los estratos de muestreo como los posestratos podrían contener información importante con respecto a la estimación de la captura. En otras palabras, dentro de un posestrato c , se conserva la distinción entre los estratos de muestreo, o fracciones de los mismos, de tal forma que en efecto el posestrato c es subdividido de nuevo por los estratos de muestreo $\{h\}$. Se obtuvo este primer estimador posestratificado con el enfoque general de Valliant (1993) para los datos obtenidos con el muestreo estratificado de dos etapas, pero modificados para tratar una diferencia importante entre los ejemplos de Valliant (1993) y los datos de muestreo en puerto. Específicamente, en el caso de los datos de muestreo en puerto, los peces en una muestra de una bodega individual no pueden provenir de más de un estrato de muestra y se supone que no provienen de más de un posestrato; por contraste, los conglomerados de Valliant (1993), equivalentes a bodegas, podían abarcar posestratos.

El primer estimador posestratificado de la captura total de la especie i en el posestrato c , $\widehat{W}_{ps-I; ci}$, es la suma de las estimaciones de la captura de especie i de estratos de muestreo enteros, o de partes de los mismos, que pertenecen al posestrato c . En otras palabras,

$$\widehat{W}_{ps-I; ci} = \sum_{h: h \cap c} W_{h \cap c} \frac{[\sum_{j \in h \cap c} W_{hj} \cdot g_1(\dots)]}{[\sum_{j \in h \cap c} W_{hj}]} \quad (3)$$

donde la suma exterior está por encima de los estratos de muestreo h que se intersecan con el posestrato c , $h \cap c$ se refiere a la región del estrato de muestreo h que se encuentra también en el posestrato c , y $W_{h \cap c}$ es la captura total de la pesca en esa región. Nótese que este estimador posestratificado conservaría los estratos de muestreo $\{h\}$. Por ejemplo, en la Figura 1c se presenta un ejemplo de cuatro posestratos espaciales (A-D), y se puede ver que los estratos de muestreo espaciales 5, 7, 11, y 12 serían bisecados para crear el posestrato costero hipotético D, mientras que los estratos de muestreo 6 y 13 están totalmente contenidos dentro del posestrato D. $\widehat{W}_{ps-I; ci}$ podría resultar problemático si existen muchos h cuyos $h \cap c$ son pequeños y contienen pocas muestras.

El segundo estimador posestratificado es consecuencia de hacer caso omiso a los estratos de muestreo y en su lugar usar el enfoque tomado para el muestreo aleatorio sencillo de dos etapas (por ejemplo, Thompson 2002). Esto podría ser considerado razonable en vista de la naturaleza oportunista de la toma de muestras como tal. Esto resulta en un estimador posestratificado idéntico en forma a la ecuación (1), pero con W_h y W_{hj} reemplazado por W_c y W_{cj} , respectivamente. En otras palabras, este segundo estimador posestratificado de la captura total de la especie i en el posestrato c , $\widehat{W}_{ps-II; ci}$, se expresa de esta forma:

$$\widehat{W}_{ps-II; ci} = W_c \frac{[\sum_{j=1}^{q^*} W_{cj} \cdot g_1(\dots)]}{[\sum_{j=1}^{q^*} W_{cj}]} \quad (4)$$

donde q^* es el número de muestras en el posestrato c .

4.2. Selección del estimador específico

Para seleccionar entre los dos estimadores posestratificados (o sea, entre la ecuación (3) y la ecuación (4)) se propone que se ajusten modelos lineales generalizados a los datos de muestreo en puerto (pesos pro-

medio, fracciones de especies) a fin de evaluar si es necesario retener estratos de muestreo dentro de posestratos. En el caso del peso promedio, se puede obtener una evaluación global de la utilidad de los estratos de muestreo dentro de posestratos mediante el ajuste de los dos modelos siguientes (por especie) y la comparación del desempeño relativo de los dos modelos para cada especie con el Criterio de Información de Akaike (AIC; Burnham y Anderson 2002):

$$\text{promedio}(w_{chij}/m_{chij}) = \text{promedio general} + \text{efecto posestrato}$$

y

$$\text{promedio}(w_{chij}/m_{chij}) = \text{promedio general} + \text{efecto estrato de muestra posestrato}$$

donde ‘estrato de muestra posestrato’ indica estratos de muestra dentro de posestratos. Se puede realizar la misma evaluación global para la composición por especies mediante el ajuste de un modelo de regresión logístico (reacción polítoma) a las fracciones de especies (con las mismas variables independientes que en lo anterior). Las muestras de pesos medios y fracciones de especies son componentes usados para evaluar g . Si los modelos con el efecto estrato de muestra posestrato conducen solamente a una pequeña reducción del AIC en los modelos respectivos con el efecto posestrato más sencillo, esto sugeriría que dentro de los posestratos se podría hacer caso omiso a los estratos de muestreo. Esto es plausible para los posestratos que son definidos con base en análisis de estructura espaciotemporal en las distribuciones de frecuencia de talla y tendencias de la CPUE. Los modelos lineales generalizados anteriores serían ajustados con pesos iguales a las cantidades de captura por bodega individual (para ser consistente con la ponderación del estimador de razón). Además, y particularmente si los modelos ajustados con efectos estrato de muestra posestrato resultan en reducciones grandes del AIC, la utilidad de los estratos de muestreo dentro de posestratos serán estudiados más a fondo mediante el ajuste de modelos lineales generalizados a los pesos promedio y las fracciones de especies dentro de cada posestrato (la variable independiente sería un efecto estrato de muestra categórico). En caso necesario, pruebas del significado del efecto estrato de muestra por posestrato puede ser combinado a través de posestratos para obtener un valor p general con el método de producto truncado (Zaykin *et al.* 2002) para comparaciones múltiples.

4.3. Factibilidad

La factibilidad de aplicar un estimador posestratificado depende en parte de la resolución espacial de tanto la captura descargada total (de todas las especies combinadas) como los datos de muestreo en puerto. Se asigna la captura descargada total a los estratos de muestreo usando información de los datos de observadores y los cuadernos de bitácora de los buques. Se registra la información de los observadores y de bitácora espaciales en términos de latitud y longitud, con una resolución mínima de área de 5° (a menos que no se cuente con datos en absoluto). Por lo tanto, la captura descargada total será conocida igual de bien con respecto a los estratos espaciales de muestreo que con cualquier posestratificación espacial que sea similarmente derivada de combinaciones de áreas de 5° . La sola excepción a esto puede ocurrir con muestras asociadas con la pesca en el Golfo de California (es decir, la región dentro del golfo que forma el estrato de muestreo 8; Figura 1a).

Si los posestratos espaciales son grandes, se espera que la mayoría de la información espacial de las muestras de puerto, o toda, sea conocida al nivel de posestrato espacial. En el momento de la toma de los datos de muestreo en puerto, desde 2000 se ha registrado para la mayoría de las muestras no sólo el área de muestreo sino también el área de 5° . Una comparación preliminar de las áreas de 5° de las muestras de puerto con las posiciones reales de los lances cuya captura fue cargada en las bodegas muestreadas indica que un 82% de todas las muestras de 2000 a 2009 concordaban con la posición real del lance al nivel de área de 5° , y un 98% de todas las muestras estaban a menos de un cuadrángulo de 5° del área de 5° correspondiente.

Se señala que en el caso de una posestratificación espacial que no encaja con las áreas de estratificación de muestreo, podrían existir bodegas con captura que provino de múltiples estratos espaciales de (y por lo tanto no son muestreados) pero podrían ser de solamente un posestrato espacial. Estas bodegas no serían

muestreadas. No obstante, esto es consistente con el supuesto actual que las bodegas muestreadas representan la pesquería adecuadamente.

5. TRABAJO FUTURO

5.1. Tratamiento de estratos de muestreo con captura pero muestras insuficientes

En promedio, cada año desde 2000, aproximadamente el 20% de la captura descargada total está asociado con estratos de muestreo para los cuales no fue posible obtener datos de muestreo en puerto (debido a limitaciones logísticas) o para los cuales se dispuso de menos que dos muestras. Una vez determinada una posestratificación apropiada de los datos, el estimador posestratificado podría simplificar el tratamiento de capturas descargadas en estratos de muestreo sin datos de muestreo. Por ejemplo, suponiendo posestratos espaciales grandes, si el estimador posestratificado usado tiene la forma de la ecuación (4), entonces es probable que la captura esté siempre asociada con datos de muestreo en un área de posestrato. Se investigará en el futuro un tratamiento exacto bajo posestratificación de la captura de estratos de muestreo sin datos de muestreo.

5.2. Estimación de la varianza

Inicialmente se computarán estimaciones aproximadas de la varianza del estimador posestratificado con un procedimiento *bootstrap* de ‘media muestra’ (Efron 1982), en el que el número de muestras menos uno en un estrato es extraído (con reemplazo) a fin de eliminar sesgos (subestimación de la varianza). Si el traslape de los estratos de muestreo y los posestratos es limitado para algunos posestratos, W_{ps-I} podría tener una varianza mayor debido a tamaños de muestra más pequeños que W_{ps-II} , pero quizá tener una probabilidad menor de estar sesgado si los estratos de muestreo continúan relevantes dentro de los posestratos. Se realizará el remuestreo a partir de las distribuciones empíricas de la talla y la composición por especies, en lugar de aplicar *bootstrap* paramétricos. Se considerarán tres distintos escenarios de remuestreo: el primero usará los estratos de muestreo, el segundo los posestratos, y el tercero los estratos de muestreo dentro de los posestratos. Este procedimiento no incluye un factor finito de corrección de población. No obstante, dado que el nivel anual estimado de cobertura de muestreo de la captura es aproximadamente 8% (computado como la suma de la captura en las bodegas muestreadas dividida por la captura total de la pesca), esto probablemente no es crítico. No obstante, en el futuro se explorarán otros procedimientos de estimación de la varianza (por ejemplo, procedimientos de remuestreo para estudios de datos que incluyen un factor finito de corrección de población; por ejemplo, Sitter 1992), y se derivará una estimación de la cobertura de muestreo con base en el número de bodegas de buques.

5.3. Evaluación de la factibilidad para los datos de muestreo en puerto de las pesquerías de caña y de cerco antes de 2000

Al igual que con los datos recientes de muestreo en puerto, la factibilidad de posestratificar los datos de muestreo en puerto anteriores a 2000 obtenidos de buques de cerco y de caña dependerá en parte de la resolución de la información espacial asociada con esas muestras. Esto será investigado en el futuro.

RECONOCIMIENTOS

Agradecemos a Nancy C. H. Lo nuestras útiles discusiones.

REFERENCIAS

- Aires-da-Silva, A. and Maunder, M.N. 2010. Status of bigeye tuna in the eastern Pacific Ocean in 2008 and outlook for the future. Inter-American Tropical Tuna Commission Stock Assessment Report 10, pages 116-228.
- Anonymous, 2010. Tunas and billfishes in the eastern Pacific Ocean in 2009. Fishery Status Report 8.
- Burnham, K.P. and Anderson, D.R. 2002. Model Selection and Multimodel Inference. Springer.
- Efron, B. 1982. The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics.

- Holt, D. and Smith, T.M.F. 1979. Post stratification. *Journal of the Royal Statistical Society, Series A* 142, Part 1: 33-46.
- Lennert-Cody, C.E., Maunder, M.N., Aires-da-Silva, A. 2010. Preliminary analysis of spatial-temporal pattern in bigeye tuna length-frequency distributions and catch-per-unit-effort trends. Document BET-01-02a, External Review of IATTC Bigeye Tuna Assessment, 3-7 May, 2010.
- Maunder, M.N. and Aires-da-Silva, A. 2010. Status of yellowfin tuna in the eastern Pacific Ocean in 2008 and outlook for the future. *Inter-American Tropical Tuna Commission Stock Assessment Report* 10, pages 3-109.
- Sitter, R.R. 1992. A resampling procedure for complex survey data. *Journal of the American Statistical Association* 87: 755-765.
- Suter, J.M. 2008. An evaluation of the area stratification used for sampling tunas in the eastern Pacific Ocean and implications for estimating total annual catches. Masters thesis, San Diego State University, San Diego, California, USA.
- Thompson, S.K. 1992. *Sampling*. Wiley.
- Tomlinson, P.K. 2004. Sampling the tuna catch of the eastern Pacific Ocean for species composition and length-frequency distributions. *Inter-American Tropical Tuna Commission Stock Assessment Report* 4, pages 311-333.
- Valliant, R. 1993. Poststratification and conditional variance estimation. *Journal of the American Statistical Association* 88: 89-96.
- Watters, G.M. 1999. Geographical distributions of effort and catches of tunas by purse-seine vessels in the eastern Pacific Ocean during 1965-1998. *Data Report* 10.
- Zaykin, D.V., Zhivotovsky, L.A., Westfall, P.H. and Weir, B.S. 2002. Truncated product method for combining p-values. *Genetic Epidemiology* 22: 170-185.

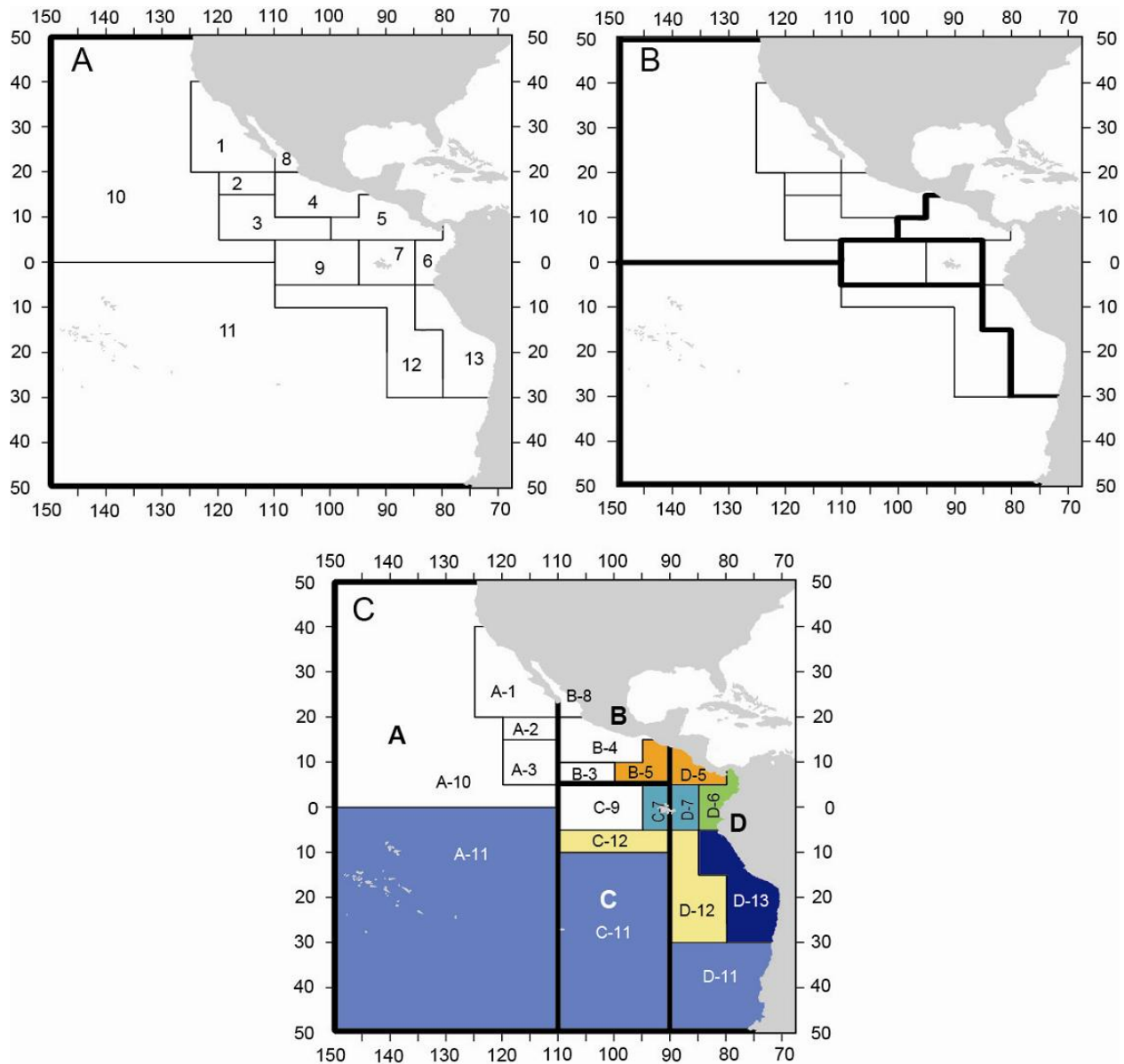


FIGURE 1. a) Sampling strata (Tomlinson 2004); b) stock assessment strata for bigeye tuna in floating-object sets (Aires-da-Silva and Maunder 2010); and c) an example of hypothetical spatial poststrata (A-D) based on results of a multivariate regression tree analysis of bigeye tuna length-frequency distributions (Lennert-Cody *et al.* 2010). The spatial sampling strata that are partially or totally contained in poststratum D are colored to illustrate spatial sampling strata that cross poststratum boundaries.

FIGURA 1. a) Estratos de muestreo (Tomlinson 2004); b) estratos de la evaluación de la población de atún patudo en lances sobre objetos flotantes (Aires-da-Silva and Maunder 2010); y c) un ejemplo de posestratos espaciales hipotéticos (A-D) basados en los resultados de un análisis de árbol de regresión multivariable de las distribuciones de frecuencia de talla de atún patudo (Lennert-Cody *et al.* 2010). Se coloran los estratos espaciales de muestreo parcial o totalmente contenidos en el posestrato D para ilustrar estratos espaciales de muestreo que cruzan límites de posestrato.