

Análisis comparativo entre ecuaciones predictivas del peso en función de la talla en atún rojo, *Thunnus thynnus* (L.). Documento de trabajo (julio, 2014)

Vicente D. Estruch y José L. Cort

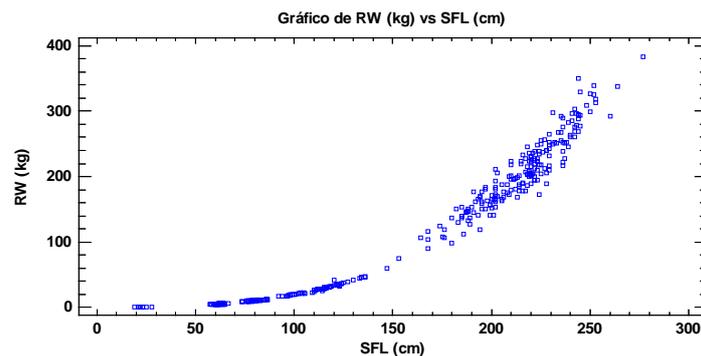
Estudio de validación comparativa

En este análisis se han comparado cuatro modelos a partir de una única muestra bivalente (SFL (cm), RW (kg)) de atunes rojos juveniles y adultos en fase reproductiva, para establecer qué modelo(s) se ajustan mejor a los datos, es decir, qué modelo(s) reflejan mejor la realidad representada por la muestra y, por tanto, tendrá(n) mayor poder descriptivo y predictivo.

Descripción de la muestra

Resumen estadístico

	SFL (cm)	RW (kg)
Recuento	407	407
Promedio	150,108	111,614
Mediana	153,0	75,0
Desviación estándar	69,8058	104,837
Mínimo	19,0	0,126
Máximo	277,0	384,0
Rango	258,0	383,874
Cuartil inferior	80,0	9,9
Cuartil superior	219,0	204,0



Modelos (ecuaciones) objeto de análisis

- Ecuación WEST BFT (en adelante **WEST**). Propuesta de Rodríguez-Marín, E. *et al.* (2014):

$$RW \text{ (kg)} = 1,59137 * e^{-5 * SFL \text{ (cm)}}^{3.020584}$$

- Ecuación EAST BFT (en adelante **EAST**). Propuesta de Rodríguez-Marín, E. y M. Ortiz (2014):

$$RW \text{ (kg)} = 3,15551 * 10^{(-5)} * SFL \text{ (cm)}^{(2,898454)}$$

Otras ecuaciones:

- Atlántico oriental y Mediterráneo occidental. Arena, P. (no publicado) en: ICCAT (2006):

$$RW \text{ (kg)} = 0.000019607 * SFL \text{ (cm)}^{3.0092} \text{ (en adelante **Ec 2**)}$$

- Atlántico occidental. Parrack, M. y P. Phares (1979):

$$RW \text{ (kg)} = 0.0000152 * SFL \text{ (cm)}^3 * 0.0531 \text{ (en adelante Ec 3)}$$

Para una mejor interpretación de los resultados, en las tablas se ha utilizado, para los modelos, la coloración progresiva de izquierda a derecha (del verde al rojo) que indicaría desde el mejor modelo (verde) hasta el peor (rojo):



Análisis

Bondad del ajuste

En la tabla 1 se muestran diversos indicadores de la bondad del ajuste. Dichos indicadores tienen en cuenta la distancia de los valores reales a los valores estimados por los modelos, sin tener en cuenta si el modelo sobreestima o subestima los valores de peso.

Resultados

- Claramente los modelos dados por Ec 2 y Ec 3 presentan un mejor ajuste global a los datos, teniendo en cuenta los valores de R^2 , el error absoluto medio y el error relativo medio.
- Obsérvese asimismo que los extremos superiores de los intervalos de confianza para los errores absolutos y relativos correspondientes a las ecuaciones Ec 2 y Ec 3, quedan por debajo de los extremos inferiores de los intervalos correspondientes a los modelos WEST y EAST. (Ver tabla 1, y figuras 2 y 3).
- Teniendo en cuenta los indicadores de bondad de ajuste descritos en las tablas 1 y 2, los modelos dados por Ec 2 y Ec 3 se ajustan mejor a los datos y, en principio tendrían un mayor poder predictivo.

Tabla 1. Indicadores descriptivos de la bondad del ajuste de las ecuaciones a los datos ($N = 407$): coeficiente de determinación (R^2), error absoluto medio ($Eam = \frac{\sum |RW_i - Estimated RW_i|}{N}$), error estándar del error absoluto (desviación estándar del error absoluto/ \sqrt{N}), error relativo medio ($Erm = \frac{\sum \frac{|RW_i - Estimated RW_i|}{Estimated RW_i}}{N} \cdot 100$) y error estándar del error relativo (desviación estándar del error relativo/ \sqrt{N}).

	R^2 (%)	Eam (kg)	Err Est Eam	Erm (%)	Err est Erm
WEST	95,567569	13,95742286	0,846827575	13,88723792	0,427038928
EAST	96,556263	11,74866827	0,768099007	8,93162154	0,390563809
Ec 2	98,101710	8,461995234	0,579871261	6,87618209	0,243480022
Ec 3	98,195354	8,106078686	0,570514773	6,63307435	0,268065106

Tabla 2. Intervalos de confianza al 95% para el error absoluto medio.

	IC (95%) Eam	IC (95%) Erm
WEST	13,9573 +/- 1,66473 [12,2926; 15,622]	13,8842 +/- 0,839385 [13,0448; 14,7236]
EAST	11,7487 +/- 1,50995 [10,2388; 13,2586]	8,93217 +/- 0,767806 [8,16436; 9,69997]
Ec 2	8,46201 +/- 1,13993 [7,32208; 9,60193]	6,87625 +/- 0,478666 [6,39758; 7,35492]
Ec 3	8,10602 +/- 1,12154 [6,98449; 9,22756]	6,66117 +/- 0,526787 [6,13438; 7,18796]

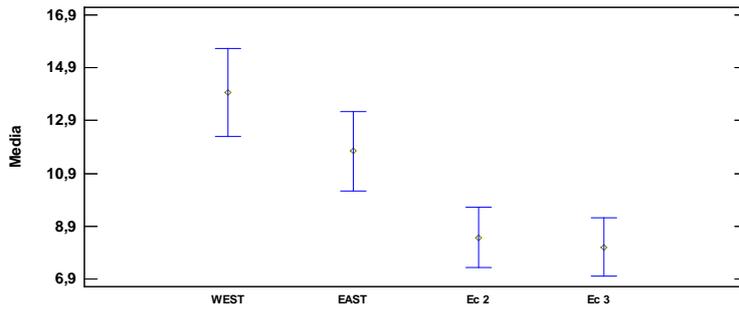


Figura 1. Medias e intervalos del confianza (95%) para los errores absolutos.

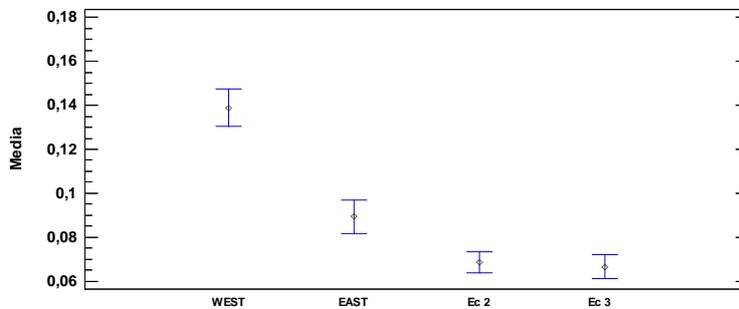


Figura 2. Medias e intervalos del confianza (95%) para los errores relativos.

Indicadores posicionales

Para un modelo correctamente ajustado a los datos reales debería cumplirse que los datos reales se distribuyen a ambos lados de la curva de forma que, aproximadamente el 50% de los datos deben quedar por encima de la curva y el 50% por debajo (la llamaremos **propiedad de equidistribución**).

Resultados

- En el caso del modelo WEST, el 94,84% de los valores reales son superiores a los valores estimados, lo cual indica que este modelo claramente subestima el peso.
- Sólo los intervalos de confianza (al 95%) para los modelos EAST y Ec 3 contienen el valor 50, lo cual indica que sólo estos modelos cumplen con la propiedad de equidistribución.
- El modelo WEST es el que viola más claramente la propiedad de equidistribución.

Tabla 3. Indicadores posicionales para valorar si los modelos proporcionan valores de estimación del peso superiores o inferiores a los valores reales.

	% valores reales inferiores a los valores estimados (% valores reales superiores a los valores estimados)	Intervalos de confianza al 95 % para el % de valores reales inferiores a los valores estimados (% de valores reales superiores a los valores estimados)
WEST	5,16 (94,84)	[3,22218;7,7796 [92,2204;96,7778]
EAST	45,95 (54,05)	[41,0304;50,9289] [49,0711;58,9696]
Ec 2	67,32 (32,68)	[62,4382;71,7882] [28,2118;37,5618]
Ec 3	0,4693 (0,5307)	[41,9964;51,9085] [48,0915;58,0036]

Análisis de los residuos (RW (kg) – valor estimado por el modelo)

Los residuos son los valores que se obtienen al calcular la diferencia entre el valor real del peso y el valor estimado por el modelo (valor real - valor estimado), teniendo en cuenta el signo (la diferencia puede ser positiva o negativa).

Un buen ajuste del modelo a los datos debe dar lugar a unos valores de los residuos con media cero y una desviación estándar no demasiado grande. Por lo tanto sería deseable que el intervalo de confianza para la media de los residuos contuviese el 0 y que la desviación estándar fuese lo más pequeña posible.

Resultados

- La diferencia entre los valores de promedio y mediana indican gran asimetría de los residuos para los modelos WEST y EAST en comparación con los modelos Ec 2 y Ec 3, lo cual se puede comprobar visualmente en la figura 3.
- Sólo el modelo Ec 3 cumple estrictamente con el requisito de que el intervalo de confianza contenga al valor 0.

- Aunque el intervalo de confianza al 95% para Ec 2 no contiene a 0, los extremos son valores relativamente pequeños, claramente más cercanos a 0 que en el caso de los modelos WEST y EAST.
- Los valores promedio de los residuos son claramente menores para los modelos Ec 3 y Ec 4.
- Los modelos WEST y EAST subestiman sensiblemente el peso real (tabla 4, figuras 3 a y 3 b), mientras que el modelo Ec 2 lo sobrestima pero sólo ligeramente (tabla 4, figura 3 c)

Tabla 4. Resumen estadístico para los residuos correspondientes a los distintos modelos analizados respecto de los datos globales e intervalos de confianza al 95% para la media de los residuos.

	WEST	EAST	Ec 2	Ec 3
Recuento	407	407	407	407
Promedio	13,1879	111,155	-2,49932	-0,237931
Desviación estándar	17,6863	105,1520	14,2258	14,0814
Mediana	2,49364	0,143725	-0,651763	0,0487585
Intervalo de confianza del 95,0% para la media:	13,1879 +/- 1,7234 [11,4645; 14,9113]	111,155 +/- 10,2463 [100,908; 121,401]	-2,49932 +/- 1,3862 [-3,88552; -1,11312]	-0,237931 +/- 1,37213 [-1,61006; 1,1342]

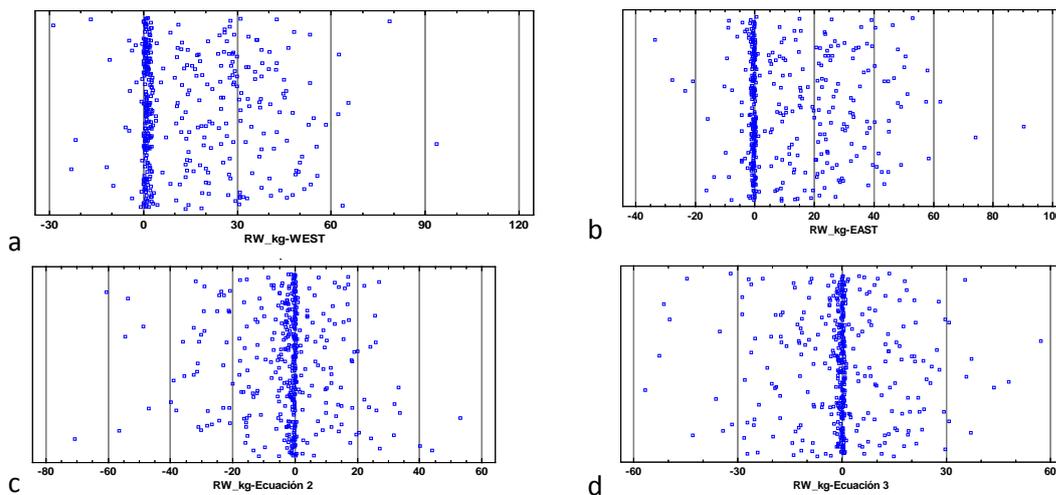


Figura 3. Gráficos de dispersión de los residuos. Las imágenes a, b, c, y d corresponden a los residuos de los modelos correspondientes a las ecuaciones WEST, EAST, Ec 2 y Ec 3, respectivamente.

Conclusiones

- El resultado aceptable para EAST respecto a la equidistribución de los datos a un lado y otro de la curva no se contradice con el mal resultado para el mismo modelo al considerar los residuos ya que los residuos también miden la lejanía de los datos a la curva y no sólo la posición de los datos respecto de la curva.
- A la vista de los resultados obtenidos, podemos afirmar que el modelo que mejor se ajusta a los datos es Ec 3, seguido por Ec 2.

- Los análisis realizados permiten asegurar que los modelos WEST y EAST no son apropiados para explicar el comportamiento de los datos de la muestra.

Referencias

Arena, P. (no publicado) en: ICCAT (2006). Length-weight relationships adopted by the SCRS for major species.

<http://www.iccat.int/Documents/SCRS/Manual/Appendices/Appendix%204%20III%20Length-weight.pdf>.

Parrack, M. y P. Phares (1979). Aspects of the growth of Atlantic bluefin tuna determined from mark-recapture data. *Col. Vol. Sci. Pap. ICCAT*, 8: 356–366.

Rodríguez-Marín, E. y M. Ortiz (2014). Further analysis results of biometric relationships of Atlantic bluefin tuna, attending to the recommendation of ICCAT bluefin tuna Group (2014 ICCAT Bluefin Data Preparatory Meeting). ICCAT, SCRS/2014/053 Rev: 17 p.

Rodríguez-Marín, E., J. M. Ortiz de Urbina, N. Abid, E. Alot, I. Andrushchenko, S. Deguara, A. Di Natale, M. Gatt, W. Golet, S. Karakulak, A. Kimoto, D. Macías, P. Quelle, S. Saber, M. N. Santos, J. Walter y R. Zarrad (2014). Length weight relationships for Atlantic bluefin tuna (*Thunnus thynnus*). ICCAT, SCRS/2014/053: 19 p.