



NUEVA ESPECIE

Cymothoa facimar



DRA. MAYRA GRANO-MALDONADO Y DR. JOSÉ SALGADO-BARRAGÁN
DEDICAN UN ISÓPODO PARÁSITO A LA FACIMAR





Artículo Científico

¿La heterogeneidad en el tamaño de la muestra afecta la selección del modelo de crecimiento?

Does sample size heterogeneity affect growth model selection?



1. Guillermo Rodríguez-Domínguez



0000-0002-5612-5131

Universidad Tecnológica de Escuinapa.
Camino al Guasimal S/N, Col. Centro,
C.P. 82400, Escuinapa, Sinaloa, México
Autor de correspondencia: guirodom@uas.edu.mx



2. Jorge Payán-Alejo



0000-0003-4636-0274

Facultad de Ciencias del Mar, Universidad
Autónoma de Sinaloa. Paseo Clausen s/n
Los Pinos, C.P. 82000, Mazatlán, Sinaloa,
México. jorge.payan.facimar@uas.edu.mx

latindex



CREATIVE COMMONS



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original



¿La heterogeneidad en el tamaño de la muestra afecta la selección del modelo de crecimiento?

Does sample size heterogeneity affect growth model selection?

► RESUMEN

Para resolver el problema de datos limitados de edad-talla en organismos acuáticos se ha propuesto fortalecer la muestra con datos simulados con distribuciones normales. Se considera que el reforzamiento corrige el efecto de la heterogeneidad de tamaños al seleccionar un modelo sesgado. Para analizar estas afirmaciones se generaron datos simulados de edad-talla bajo nueve escenarios diferentes, con tamaño de muestra heterogénea y después estos mismos datos fueron reforzados con más datos simulados para homogenizar los tamaños de muestra por edad. Entonces varios modelos candidatos fueron evaluados para cada conjunto de datos (datos originales y reforzados) bajo un enfoque de selección multimodelo (MMS) con errores aditivos y multiplicativos. Entonces se estimó el índice de Akaike y la diferencia de Akaike. Las muestras homogeneizadas se repitieron 100 veces.

El Enfoque MMS seleccionó siempre el modelo generador de los datos como el mejor modelo con tamaños de muestra heterogénea. Mientras que, con las muestras homogeneizadas, no siempre el modelo generador fue seleccionado como mejor y en algunos casos fue clasificado como el peor. Se observó que la heterogeneidad de varianzas y la estructura del error asumido sí afecta la selección del mejor modelo entre muestras originales y homogeneizadas. Se propone que tanto el modelo candidato como la estructura del error se sometan al enfoque MMS cuando se modele el crecimiento de especies acuáticas y hacer un esfuerzo por complementar los datos limitados y no fortalecer con datos simulados que pueden tener un sesgo de origen.

Palabras clave: Muestras homogeneizadas, Crecimiento, Selección multimodelo, estructura del error, varianza heterogénea.



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

► ABSTRACT

To solve the problem of limited age-size data in aquatic organisms, bolstered sample, using simulated data with normal distributions has been proposed. Bolstered is believed to correct the effect of selecting a biased model, by size sample heterogeneity. To analyze these claims, simulated age-length data were generated under nine different scenarios with heterogeneous sample sizes, and then these same data were bolstered with more simulated data to homogenize sample sizes by age. Several candidate models were then evaluated for each data set (original and reinforced data) using a multi model selection (MMS) approach, with additive and multiplicative errors. The Akaike index and Akaike difference were then estimated. The bolstered samples were repeated 100 times. The MMS approach always selected the model that generated the data as the best model with heterogeneous sample sizes. While with the bolstered samples, the generating model was not always selected as the best, and in some cases, it was classified as the worst. It was observed that variance heterogeneity and the assumed error structure do affect the selection of the best model between the original and boosted samples. It is proposed that both the candidate model and the error structure be subjected to the MMS approach when modeling the growth of aquatic species, and that an effort be made to complement the limited data and not reinforce them with simulated data that may have a source bias.

Keywords: Bolstered samples, Growth, Multimodel selection, error structure, heterogeneous variance.

► INTRODUCCIÓN

En la modelación del crecimiento de especies acuáticas se utiliza frecuentemente el enfoque de selección multimodelo (MMS) (Luquin-Covarrubias et al. 2016; Góngora-Gómez et al. 2018), donde varios modelos candidatos se ajustan a los datos de crecimiento observado, a través de un procedimiento iterativo, para estimar los parámetros de cada modelo, maximizando una función de verosimilitud normal o verosimilitud log normal según sea los supuestos acerca de los errores. Una vez ajustados los modelos candidatos se selecciona el mejor utilizando el criterio de información de Akaike (AIC, por sus siglas en inglés) (Akaike 1983); el mejor modelo será aquel con el menor valor.



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

AIC es una medida de la distancia relativa de la distribución de probabilidad de los errores respecto a la distribución normal o log-normal. La diferencia de Akaike (ΔAIC) mide la distancia relativa entre modelos y se usa también como indicador del soporte del modelo en los datos. Burnham y Anderson (2002) establecieron niveles de soporte empírico de cada modelo candidato dado los datos, con base en la ΔAIC ; un valor entre 0 y 2 indica un soporte sustancial (cualquiera de los modelos en este nivel puede describir adecuadamente el crecimiento), mientras que valores entre 4 a 7 tendrían considerablemente menos soporte y valores > 10 , esencialmente ningún soporte. En esta categorización de Burnham y Anderson (2002) es evidente que quedan fuera los intervalos ΔAIC entre 2 a 4 y de 7 a 10. Los errores pueden ser aditivos con distribución de probabilidad normal o multiplicativos con distribución log normal. Generalmente cuando se observa una varianza de la talla incrementando con la edad se asume que el error es multiplicativo, mientras que si la varianza se mantiene constante entre edades se asume error aditivo. Sin embargo, la alternativa es no asumir a priori una estructura de error dada, si no evaluarla también junto con los modelos candidatos utilizando el AIC. La combinación del mejor modelo y estructura de error puede ser elegida con base al menor AIC (Rodríguez-Domínguez et al. 2024).

MMS se ha referido como un mejor enfoque para la modelación del crecimiento en especies acuáticas, sin embargo, Bolser et al. (2018) afirman que la distribución del tamaño de muestra entre edades puede afectar la selección de un modelo sesgado con consecuencias en la evaluación del stock y por deriva en el manejo de la pesquería. Estos autores proponen un procedimiento para homogeneizar el tamaño de las muestras por edad, a través de un procedimiento de simulación de datos aleatorios. Aplicando este procedimiento a datos de talla a la edad de curvina Golfina y un enfoque MMS, encontraron que el modelo de Von Bertalanffy (1938) fue el mejor para describir el patrón de crecimiento de la curvina golfina a diferencia del modelo Schnute-Richards (1990), que fue el mejor seleccionado en el análisis de los datos originales. Usaron ambos patrones de crecimiento en un modelo de rendimiento por recluta y encontraron que el modelo de Schnute-Richards predijo un stock más saludable que cuando se usa el de Von Bertalanffy, concluyendo que una selección errónea del modelo de crecimiento tiene serias implicaciones en la evaluación y manejo del stock.

**OPEN ACCESS**

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

Enciso-Enciso et al. (2022) también aplicaron el procedimiento de Bolser en datos de *Sardinops sagax* y encontraron que los únicos modelos con un suficiente soporte en los datos fueron el de Von Bertalanffy y el de Schnute, a diferencia del modelo logístico que fue seleccionado como el mejor cuando se analizaron los datos originales.

El procedimiento de Bolser et al., (2018) para homogenizar las muestras por edad, genera datos al azar a partir de una distribución normal para complementar las edades con muestras pequeñas o elige al azar una submuestra de datos donde la muestra es considerablemente grande. Como el procedimiento implica un efecto aleatorio cabe la pregunta ¿Qué pasa si se repite el procedimiento de homogenización de muestras varias veces? ¿Se obtiene siempre el mismo resultado?

Usando MMS el mejor modelo es aquel con el menor AIC así que un cambio en la selección del mejor modelo, como sucede cuando se homogeniza el tamaño de la muestra, debe implicar un cambio en el AIC de cada modelo candidato. Para analizar esta posibilidad examinemos la diferencia de Akaike entre un modelo i y el mejor modelo (b) que se puede expresar de la siguiente manera:

$$\Delta AIC = 2 * (\phi_i - \phi_b) - 2 * (LLi - LLb)$$

Donde ϕ_i y ϕ_b son el número de parámetros de los modelos i y b , mientras que LLi y LLb son la máxima log verosimilitud de ambos modelos.

El primer término no se ve afectado por el tamaño de la muestra, por lo que la ΔAIC al modificar el tamaño de la muestra depende del segundo término y específicamente de la diferencia entre log verosimilitudes. El segundo término podría expresarse así:

$$-2 * (LLi - LLb) = (n) * [\ln(2\pi) + 2 * \ln(\sigma_i) + 1] - (n) * [\ln(2\pi) + 2 * \ln(\sigma_b) + 1]$$

Donde σ_i y σ_b son las desviaciones estándar de los errores y n es el tamaño de la muestra global de todas las edades. El término no considera homogeneidad o heterogeneidad del tamaño de la muestra entre edades.

La simulación de datos aleatorios, para homogenizar el tamaño de las muestras, se obtiene de distribuciones normales con las medias y σ de los datos originales, por lo tanto, se puede considerar que σ_i y σ_b son igual a aquellas de los datos originales, pero n puede cambiar, entonces el término $-2 * (LLi - LLb)$ cambiará proporcionalmente con el tamaño de

**OPEN ACCESS**

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

la muestra, de modo que la ΔAIC cambiará, pero nunca el modelo i será mejor que el modelo b. La única manera de que el modelo i sea mejor que el modelo b es que el valor de σ_i disminuya y σ_b aumente. Por lo tanto, la homogenización del tamaño de las muestras no debería cambiar la selección del mejor modelo con el MMS. Sin embargo, la generación de datos aleatorios implica que puede cambiar aleatoriamente σ y por tanto ΔAIC , así que la selección del mejor modelo puede también cambiar, especialmente cuando la distancia entre dos modelos es pequeña.

Los datos analizados en Bolser et al. (2018) y Enciso-Enciso. (2022) además de la heterogeneidad en el tamaño de la muestra, también contienen heterogeneidad de varianzas entre edades. Enciso-Enciso et al. (2022) asumieron error multiplicativo para estabilizar la varianza, pero en Bolser et al. (2018) no lo consideraron.

El objetivo de este estudio fue evaluar el efecto de homogenización del tamaño de la muestra con datos simulados sobre la selección multimodelo de crecimiento, comparando la capacidad del MMS para identificar el modelo verdadero bajo escenarios de heterogeneidad de muestras y varianza.

► MATERIALES Y MÉTODOS

En la tabla 1 se muestra la información de *Sardinops sagax*, de Enciso-Enciso et al. (2022), que fue la base para la generación de datos simulados analizados en esta publicación.

Tabla 1. Talla promedio, desviación estándar y tamaño de muestra por edad de *Sardinops sagax* (datos copiados de tabla 3 de Enciso-Enciso et al. (2022)).

EDAD	TALLA	DE	N
0.5	125.7	6.9	48
1.0	143.8	11.8	357
1.5	152.3	11.4	495
2.0	165.2	14.2	1090
2.5	175.5	14.7	295



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

3.0	182.0	15.1	640
3.5	186.8	13.3	152
4.0	192.0	13.1	274
4.5	193.9	10.3	55
5.0	199.2	9.8	57
5.5	199.4	6.5	21
6.0	206.4	6.8	25

Se simularon nueve conjuntos de datos de talla a la edad, todos con el mismo tamaño de muestra por edad como en la tabla 1.

El primer conjunto de datos simulados a los que se denominará “datos originales” (DO) se generaron aleatoriamente para cada edad a partir de una distribución normal con la media y la desviación estándar de la tabla 1. Para considerar el efecto de la heterogeneidad de varianzas otro conjunto de datos fue simulados como en DO, pero con desviación estándar de 11.3 constante para cada edad y a estos datos se le denominarán DOVHO.

Otros cuatro conjuntos de datos se generaron de la misma manera que DO, pero las medias fueron sustituidas por la talla estimada de modelos logísticos con cuatro vectores de parámetros distintos que se muestran en la tabla 2 y se denominaron LOGM1, LOGM2, LOGM3 y LOGM4. Las diferencias esenciales fueron en el parámetro k ; la tasa de crecimiento, con un gradiente de 0.3 a 0.9, para que el punto de inflexión del modelo logístico fuera más evidente con el incremento de k , con la idea de que con valores bajos de k los datos generados con el modelo logístico tendrían probabilidad de soportar también el modelo de Von Bertalanffy (VBM) y que a valores mayores de k la diferencia entre estos dos modelos fuera más evidente.

**OPEN ACCESS**

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

Tabla2.- Vectores de parámetros de un modelo logístico usados en la simulación de datos de crecimiento.

Modelo	L_{∞}	k	t_0
Logístico 1 (LOGM1)	250	0.3	1
Logístico 2 (LOGM2)	255	0.4	1.5
Logístico 3 (LOGM3)	255	0.5	1.5
Logístico 4 (LOGM4)	255	0.9	1.5

Otros tres conjuntos de datos fueron generados de la misma manera que DO, pero ahora las medias fueron sustituidas por la talla estimada de un modelo VBM con tres vectores distintos de parámetros como se muestra en la tabla 3 y fueron denominados VBM1, VBM2 y VBM3.

Tabla3.- Vectores de parámetros de un modelo de Von Bertalanffy usados en la simulación de datos de crecimiento.

Modelo	L_{∞}	k	t_0
VonBertalanffy1 (VBM1)	372	0.11	-2.4
VonBertalanffy1 (VBM2)	443	0.08	-2.75
VonBertalanffy1 (VBM3)	300	0.08	-1.0

Los parámetros del VBM fueron elegidos de tal forma que la ΔAIC entre el modelo VBM y los otros modelos candidatos se fuera incrementando, así en VBM1, la ΔAIC con los otros dos modelos candidatos fue menor de 2, en VBM2, varió entre 2 y 4 y en VBM3, fue mayor de 4.

Se calcularon las medias y desviación estándar por edad de los conjuntos de datos mencionados arriba (Tabla 4) y fueron usados para homogenizar a 500 datos por edad siguiendo el procedimiento de Bolser et al. (2018). Con excepción de las edades 2 y 3 se generaron datos para complementar el tamaño de la muestra, mientras que en las edades 2 y 3 donde los datos rebasaron el tamaño de 500 datos, se eligieron aleatoriamente sin remplazo 500 de los datos disponibles. El procedimiento se repitió 100 veces para cada conjunto de datos.



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

Tabla 4. Talla promedio y desviación estándar (DE) de los conjuntos de datos generados aleatoriamente.

	DO		DOVHO		LOGM1		LOGM2		LOGM3		LOGM4		VBM1		VBM2		VBM3		
n	Edad	Talla	DS	Talla	DS	Talla	DS	Talla	DS	Talla	DS	Talla	DS	Talla	DS	Talla	DS	Talla	DS
48.0	0.5	126.5	7.3	126.5	11.2	116.2	5.9	102.9	5.6	95.6	7.0	74.1	5.8	102.0	7.4	102.5	6.8	34.1	5.4
357.0	1.0	143.0	12.1	143.1	10.4	125.1	11.8	115.2	12.4	112.2	12.8	99.6	12.1	116.0	11.7	114.6	11.7	45.0	12.0
495.0	1.5	153.3	11.4	152.1	11.4	133.7	10.7	128.4	11.1	128.4	11.7	127.3	10.8	130.0	11.7	128.0	10.6	54.5	11.0
1090.0	2.0	164.2	14.5	165.2	11.0	143.5	13.6	140.0	14.3	142.9	14.0	156.1	13.7	142.0	14.3	140.1	13.8	64.4	14.0
295.0	2.5	175.4	14.0	175.1	11.0	152.1	14.6	151.0	13.7	159.1	15.0	181.5	15	155.0	14.6	151.9	14.4	73.2	14.0
640.0	3.0	182.7	15.2	181.7	11.3	161.4	15.3	163.7	15.4	172.7	14.6	201.9	15.4	167.0	14.7	163.4	15.5	81.8	15.0
152.0	3.5	188.0	13.0	187.8	12.3	167.4	12.4	175.9	13.8	187.5	13.8	217.9	13.9	179.0	13.6	176.5	14.1	91.5	14.0
274.0	4.0	191.9	14.1	192.6	11.2	177.1	13.4	186.0	12.9	197.1	13.3	230.8	13	189.0	12	184.5	12.8	99.5	14.0
55.0	4.5	194.5	10.0	194.9	10.0	184.1	10.2	195.2	10.8	211.4	10.3	239.8	9.2	197.0	10.6	197.1	8.4	107.1	11.0
57.0	5.0	198.9	10.3	198.1	10.7	191.5	11.2	206.9	8.5	214.7	8.4	244.6	8.2	207.0	11.4	206.1	10.3	115.5	11.0
21.0	5.5	200.7	6.3	202.4	14.1	197.7	7.9	212.4	6.2	223.2	7.7	246.8	7.0	214.0	4.6	215.4	6.5	120.2	5.1
25.0	6.0	205.3	7.4	203.4	8.1	204.2	6.2	217.3	6.8	231	7.2	250.2	7.2	224.0	6.7	221.8	5.8	127.7	8.1

El enfoque MMS fue utilizado para ajustar los modelos candidatos a los datos con tamaño de muestras heterogéneos (NHET) y sus respectivas muestras ya homogenizadas (NHOMO). Para los DO y DOVHO los modelos candidatos fueron el de VBM, el de Ruiz-Velasco (RVM) (Ruiz-Velazco et al. 2010) y el de Richards (RICHM) (Richards 1959). Los modelos de Gompertz (Gompertz 1825) y logístico (Ricker 1975), que se usaron en Enciso-Enciso et al. (2022) no se consideraron como candidatos porque el parámetro to dio valores negativos, indicando que no existe el punto de inflexión en los datos, que es característico en ambos modelos.

Para los datos generados con las cuatro versiones de un modelo logístico los modelos candidatos fueron: VBM, logístico (LOGM) y RICHM y para las tres versiones de un modelo de VBM fueron: VBM, logístico (LOGM) y Schnute (SCHM), los cuales se presentan en la tabla 5.

Tabla 5. Modelos de crecimiento candidatos utilizados para ajustar a los datos simulados.

Modelo	Ecuación
VBM	$Lt = L_{\infty} * [1 - \exp(-k1 * (t - to1))]$
RVM	$Lt = L1 + (L2 - L1) * \left[\frac{(1 - k2^t)}{(1 - k2^{tmax})} \right]^b$
RICHM	$Lt = \frac{L_{\infty}}{\left[1 + \left(\frac{1}{b} \right) \exp(-k3 * (t - to2)) \right]^b}$
LOGM	$Lt = L_{\infty} * [1 + \exp(-k4 * (t - to3))]^{-1}$
SCHM	$Lt = \left[L1^b + (L2^b - L1^b) \left[\frac{1 - e^{-a(t-\tau_1)}}{1 - e^{-a(\tau_2-\tau_1)}} \right] \right]^{1/b}$



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

En VBM: L_{∞} es la longitud asintótica, k_1 es un coeficiente de crecimiento, t_0 la edad teórica en la que la longitud sería cero. RVM: L_1 y L_2 son la talla inicial y final, t se refiere a la edad y t_{max} a la edad máxima registrada.

RICHM: L_{∞} es la longitud asintótica, k es tasa de crecimiento, t_0 la edad teórica en la que inicia el crecimiento, b es parámetro de forma o curvatura.

LOGM: L_{∞} es la longitud asintótica, k es tasa de crecimiento, t_0 es edad en la cual se encuentra el punto de inflexión ($L_{\infty}/2$).

SCHM: L_1 y L_2 son longitudes promedio para el primer y último grupo de edad, a es la tasa de crecimiento, b está relacionado con el punto de inflexión de una curva de crecimiento con forma de S.

Los modelos fueron ajustados considerando errores aditivos con distribución normal, pero también errores multiplicativos con distribución log normal.

Para el procedimiento de ajuste de los modelos cuando se asumió errores aditivos la función objetivo a maximizar fue la log verosimilitud normal.

$$LL = \left(\frac{-n}{2}\right) * \left[\ln(2 * \pi) + 2 * \ln\left(\sqrt{\frac{\sum(o - e)^2}{n}}\right) + 1 \right]$$

Y cuando se asumió error multiplicativo con distribución log normal se usó la función de log verosimilitud log normal.

$$LL = \left(\frac{-n}{2}\right) * \left[\ln(2 * \pi) + 2 * \ln\left(\sqrt{\frac{\sum \ln\left(\frac{o}{e}\right)^2}{n}}\right) + 1 \right] - \sum \ln(o)$$

Donde o es la talla observada, e , es la talla estimada por el modelo candidato y n es el tamaño de la muestra.

Una vez ajustados los modelos candidatos a los datos se eligió el mejor modelo como aquel que obtuvo el menor AIC.

$$AIC = 2 * (\phi - LL)$$



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

Donde \emptyset es el número de parámetros del modelo más 1 (sigma estimada) y LL es la máxima log-verosimilitud.

También se estimó para cada modelo i la diferencia de Akaike

$$\Delta AIC_i = AIC_i - AIC_{min}$$

Donde AIC_i es el AIC de modelo a quien se estima la diferencia de Akaike y AIC_{min} es el menor AIC de los modelos candidatos evaluados en cada caso.

Además, se estimó el peso de Akaike que es una medida de la ponderación del modelo respecto a los otros modelos candidatos analizados.

$$W_i = \frac{e^{(-0.5 \cdot \Delta AIC_i)}}{\sum e^{(-0.5 \cdot \Delta AIC_i)}}$$

De las 100 veces que se repitió el procedimiento de homogenización del tamaño de las muestras, se registró la frecuencia de ΔAIC que obtuvo cada modelo, estableciendo un semáforo para clasificar el resultado: $0 < \Delta AIC < 3$ verde con soporte sustancial del modelo en los datos, $3 < \Delta AIC < 10$ amarillo, para modelos con considerablemente menos soporte y $\Delta AIC > 10$ rojo, para modelos con casi ningún soporte en los datos.

Todos los análisis estadísticos se realizaron en la plataforma R (R Core Team, 2025)

► RESULTADOS

En los nueve conjuntos de datos con NHET, analizados con el procedimiento MMS, fue evidente que los AIC de los modelos candidatos fueron consistentemente más bajos con una estructura del error aditivo que con errores multiplicativos (tabla 6).

En todos los conjuntos de datos donde se conoció el modelo de crecimiento *a priori* el mejor modelo seleccionado fue el mismo bajo el supuesto de error aditivo o error multiplicativo, sin embargo, en los DO, donde el modelo era desconocido, la selección del mejor modelo fue diferente entre ambos errores, pero por el valor más bajo de AIC , se consideró solo la estructura del error aditivo para estimar los ΔAIC y W_i .

En los DO con NHET el mejor modelo seleccionado fue RICHM,



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

seguido por el RVM y al final el VBM. RICHM obtuvo un $Wi=77.7\%$ lo que indica un buen soporte en los datos, mientras que RVM obtuvo un $Wi=21\%$ y VBM apenas un $Wi=1.3\%$. Sin embargo, al analizar las muestras homogenizadas, el 47% de las veces se eligió el VBM como el mejor modelo, aunque los tres modelos tuvieron ΔAIC menores a 3 y clasificaron en la categoría de verde.

En los cuatro casos donde se usó el modelo logístico para la generación de datos con NHET, el procedimiento MMS seleccionó acertadamente el modelo LOGM como el mejor con un Wi entre 50 y 75%. Con excepción de los datos LOGM1, el modelo clasificado en segundo lugar fue el RICHM y en tercero el VBM. Con los datos LOGM1 los tres modelos tuvieron un soporte sustancial en los datos y clasificaron en la categoría de verde. El análisis de las muestras homogenizadas también identificó frecuentemente al LOGM como el mejor modelo (83 a 99%), con excepción de los datos LOGM2 donde el modelo RICHM fue el mejor en el 99%, mientras que el verdadero modelo LOGM fue clasificado en la categoría de amarillo y rojo en el 89% de las muestras.

En los tres casos donde se usó el modelo de Von Bertalanffy y NHET, para simular los datos, el MMS detectó adecuadamente que el VBM era el mejor o con un peso sustancial en los datos; clasificado en la categoría de verde. Solo en los datos VBM1 el modelo clasificado en primer lugar fue el LOGM, pero la diferencia con el VBM fue apenas de una décima. En cambio, con el análisis de las muestras homogenizadas, solo en los datos VBM1 se identificó acertadamente que el VBM era el mejor modelo (76% como el mejor y 86% en categoría verde), pero con los datos VBM2 y VBM3 se identificó erróneamente que el SCHM fue el mejor (57%) o en la categoría verde (100%), mientras que el VBM fue erróneamente clasificado en categoría amarilla y roja en el 100% de las muestras a pesar de ser el modelo generador de la muestra analizada.

**OPEN ACCESS**

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original



Tabla 6. Valores de AIC, Δ AIC y W_i obtenidos del ajuste de los modelos candidatos a cada conjunto de datos con NHET y frecuencia de veces en que un modelo candidato obtuvo un Δ AIC en un intervalo dado, en las 100 repeticiones de muestras homogenizadas (NHOMO). En negritas y subrayados se señalan los mejores modelos.

Data	Model	TAMAÑO DE MUESTRA HETEROGENEO				TAMAÑO DE MUESTRA HOMOGENIZADA			
		ERROR ASUMIDO				NUMERO DE VECES QUE Δ AIC =			
		ADITIVO	MULTIPLICATIVO	Δ AIC	$W_i(\%)$	0	0<3	3<10	>10
DO	VBM	28290.6	28312.3	8.6	1.1	47	34	6	0
	RVM	28284.7	28306.3	2.6	21.2	22	65	0	0
	RICHM	28282.0	28309.4	0	77.7	18	69	0	0
DOVHO	VBM	26890.7	26981.4	9.2	0.6	2	22	47	29
	LOGM	26882.7	26973.4	1.2	35.3	16	84	0	0
	RICHM	26881.5	26972.6	0	64.1	82	18	0	0
LOGM1	VBM	28061.5	28103.9	1.5	25.4	17	24	55	4
	LOGM	28060.0	28101.8	0	54.1	83	16	1	0
	RICHM	28061.9	28103.8	1.9	20.5	0	100	0	0
LOGM2	VBM	28206.7	28312.7	6.5	2.7	0	0	0	100
	LOGM	28200.2	28309.6	0.0	69.1	1	10	37	37
	RICHM	28202.0	28311.5	1.8	28.2	99	1	0	0
LOGM3	VBM	28230.6	28399.1	18.6	0	0	0	0	100
	LOGM	28212.0	28386.7	0.0	72.6	87	12	1	0
	RICHM	28213.9	28388.6	2.0	27.4	14	86	0	0
LOGM4	VBM	28292.3	28542.8	179.0	0	0	0	0	100
	LOGM	28113.3	28335.3	0.0	70.3	84	16	0	0
	RICHM	28115.03	28337.2	1.7	29.7	16	84	0	0
VBM1	VBM	28140.1	28235.9	0.1	34.3	74	12	14	0
	LOGM	28140.0	28235.7	0.0	35.2	0	0	100	0
	SCHM	28140.3	28236.3	0.3	30.5	26	15	59	0
VBM2	VBM	28087.0	28164.7	0.0	64.6	0	0	15	85
	LOGM	28090.0	28167.8	3.1	14.1	43	41	16	0
	SCHM	28088.7	28167.1	1.7	21.3	57	43	0	0
VBM3	VBM	28061.9	28704.9	0.0	90.9	37	48	15	0
	LOGM	28072.7	28716.8	10.8	0.4	0	0	0	100
	SCHM	28066.6	28710.0	4.7	8.7	63	37	0	0

DISCUSIÓN

Los menores valores de *AIC* de los modelos candidatos, asumiendo errores aditivos en comparación a cuando se asumieron errores multiplicativos, era de esperarse, ya que los datos fueron generados con distribuciones de probabilidad normal. Cuando se asume que los errores son aditivos se asume también que tienen una distribución normal, pero cuando se asume errores multiplicativos se asume también que su distribución es log normal. Así que, si la verdadera distribución de los errores es normal, pero se asumen errores multiplicativos, la transformación logarítmica sesga la distribución de los errores hacia los valores negativos (al inverso de una distribución log normal) por lo que se aleja de la distribución normal generando valores de *AIC* más grandes que cuando se asumió errores aditivos.

El análisis de los datos DO con tamaños de muestra heterogéneos demostró que la estructura de error asumida afecta la selección del mejor



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

modelo, por lo que la estructura de error debiera evaluarse también junto a los modelos candidatos en un enfoque MMS. También se observó que con tamaño de muestras homogenizadas la selección del mejor modelo varió aleatoriamente en cada repetición. Por lo que tomar como mejor modelo aquel que resulta de una sola homogenización del tamaño de muestra puede resultar en una selección sesgada del verdadero modelo como sucedió con el análisis de las muestras homogenizadas VBM2 y VBM3.

La comparación de AIC de los modelos candidatos ajustados a los datos DO y DOVHO entre muestras heterogéneas y sus respectivas muestras homogenizadas no son coincidentes, especialmente con el modelo VBM, pero el desconocimiento del verdadero modelo a priori no permite concluir si las muestras homogenizadas dan un resultado más confiable. Pero la variación de *AIC* del VBM obtenida en muestra homogenizados fue mayor que la que obtuvieron los otros dos modelos candidatos, lo cual indica una mayor incertidumbre.

En los otros conjuntos de datos donde se conoció a priori el modelo de crecimiento, la heterogeneidad del tamaño de la muestra no afectó el enfoque MMS en seleccionar como mejor modelo el mismo con el que fueron generados. La única excepción fue con los datos VBM1 donde el enfoque MMS detectó erróneamente que el mejor modelo fue LOGM cuando en realidad era el VBM, sin embargo, la elección fue marginal porque las AIC de los tres modelos candidatos estuvieron con buen soporte en los datos y clasificados en la categoría de verde.

La heterogeneidad de varianzas, más que la heterogeneidad en el tamaño de las muestras, parece ser más importante de considerar cuando se ajusta un modelo. El análisis de los datos DOVHO, que son una repetición de los datos DO, pero con varianza homogénea, no mostro diferencias en la elección del mejor modelo tanto en datos con tamaños de muestras heterogéneas como en las respectivas muestras homogenizadas, ni entre la estructura del error asumida.

Bolser et al. (2018) y Enciso-Enciso et al. (2022) concluyeron que el VBM fue el mejor modelo para describir el patrón de crecimiento en sus datos, a partir de una sola muestra homogenizada, pero en ambos trabajos los datos tuvieron además una alta heterogeneidad de varianzas

**OPEN ACCESS**

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original



y esto puede afectar la selección del mejor modelo cuando se homogenizaron los tamaños de muestra.

Muchos investigadores del crecimiento en organismos acuáticos tienen la preferencia por el VBM porque tiene una base teórica biológica, aunque los datos observados no soporten el modelo, sin embargo, otros modelos bifásicos han sido explicados teóricamente en términos energéticos (Quince et al, 2008). Los modelos SCHM, RICHM, LOGM y RVM son del tipo bifásicos y en muchas especies parecen seguir trayectorias de crecimiento diferentes a las de VBM (Katsanevakis y Maravelias, 2008)

El hecho de que el modelo de Schnute-Richards en modelos de rendimiento por recluta pueda pronosticar un estado más saludable del recurso que con otros modelos (Bolser et al. 2018), no es prueba de que sea un modelo sesgado. La evidente heterogeneidad de varianzas entre las clases de edad de sus datos puede influir en la selección del mejor modelo como se vio en el análisis de datos DO y DOVHO.

El tratamiento frecuente que se le da a las muestras cuando tienen varianzas heterogéneas es considerar a priori errores de tipo multiplicativo Log normales, (Enciso-Enciso et al., 2022, Echavarría-Heras et al 2024,) pero la transformación logarítmica no siempre logra estabilizar la varianza (Feng et al., 2014) y es otro tema que se debe analizar más detenidamente por el efecto que puede tener en el enfoque MMS en la modelación del crecimiento.

El problema del muestreo de datos de crecimiento son los tamaños de muestra en edades extremas (en las primeras y últimas edades), por el efecto de la selectividad o por mortalidad natural (Bolster et al. 2018)

Es común en los datos disponibles para modelar el crecimiento de especies acuáticas que el tamaño de la muestra de las primeras o últimas edades sea menor de 30 (Secor et al., 2009; Luquin-Covarrubias et al., 2016 ; Bolster et al., 2018; Lang et al., 2019; Enciso-Enciso et al. 2022). En estadística generalmente se asume que muestras mayores de 30 son necesarias para obtener una buena estimación de la media y la varianza de los datos. En estos casos, la homogenización del tamaño de muestras entre edades, asumiendo que la media y la varianza son inesgadas, puede ponderar un sesgo real. Nada puede sustituir la certidumbre de



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original

estos datos extremos más que un esfuerzo extraordinario por obtener la mayor cantidad de datos, o acumulando datos de otros estudios.

El efecto de seleccionar un modelo sesgado sin duda tiene consecuencias cuando esta información se usa en evaluaciones del recurso, por lo que hay que ser muy cuidadosos cuando se eligen los modelos candidatos, la estructura de errores y alternativas de ajuste cuando los datos no cumplen con el requisito de homogeneidad de varianzas.

En conclusión, la estructura del error al ajustar un modelo de crecimiento a datos observados no debiera ser asumido *a priori*, es mejor evaluarlo con un enfoque MMS eligiendo aquel que genere el menor AIC ya que la elección errónea de una estructura de error *a priori* puede influir en el enfoque MMS para elegir el mejor modelo. El procedimiento MMS sobre datos simulados con un modelo de crecimiento conocido y heterogeneidad en el tamaño de muestra seleccionó siempre como mejor modelo aquel con el que fueron simulados los datos, en cambio con las muestras homogenizadas siempre hubo una probabilidad de elegir el modelo erróneo. Se comprobó que la heterogeneidad de varianzas más que la heterogeneidad en el tamaño de las muestras afecta la selección del modelo de crecimiento con el enfoque MMS y es un caso que debe ser analizado con detalle en un trabajo aparte.

► LITERATURA CITADA

Akaike, H. (1983). Information measures and model selection. *Int Stat Inst.* 44:277-291.

Bolser, D. G., Grüss, A., López, M. A., Reed, E. M., Mascareñas-Osorio, I., & Erisman B. E. (2018). The influence of sample distribution on growth model output for a highly exploited marine fish, the Gulf Corvina (*Cynoscion othonopterus*). *PeerJ.* 6: e5582. <https://doi.org/10.7717/peerj.5582>

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.). Springer



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original



- Echavarría-Heras H., E. Villa-Diharce, A. Montesinos-López, C. Leal-Ramírez (2024).** An extended multiplicative error model of allometry: Incorporating systematic components, non-normal distributions, and piecewise heteroscedasticity. *Biology Methods and Protocols*, b p a e 0 2 4 . <https://doi.org/10.1093/biomethods/bpae024>
- Enciso-Enciso, C., Nevárez-Martínez, M. O., Sánchez-Cárdenas, R., Marín-Enríquez, E., Salcido-Guevara L. A., & Minte-Vera, C. (2022).** Allometry and Individual Growth of the Temperate Pacific Sardine (*Sardinops sagax*) Stock in the Southern California Current System. *Fishes*.7(5): 226; <https://doi.org/10.3390/fishes7050226>
- Feng Ch., H. Wang, N. Lu, T. Chen, H. He, Y. Lu, X. M Tu. 2014.** Log-transformation and its implications for data analysis. *Shanghai Arch Psychiatry*. 26(2):105–109. doi: [10.3969/j.issn.1002-0829.2014.02.009](https://doi.org/10.3969/j.issn.1002-0829.2014.02.009)
- Gompertz, B. (1825).** [On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies.](#) *Philos. Trans. R. Soc. Lond.* 115:513-583.
- Góngora-Gómez, A. M., Leal-Sepúlveda, A. L., García-Ulloa, M., Aragón-Noriega, E. A., & Valenzuela-Quinónez, W. (2018).** Morphometric relationships and growth models for the oyster *Crassostrea corteziensis* cultivated at the southeastern coast of the Gulf of California, Mexico. *Lat. Am. J. Aquat. Res.* 46(4): 735-743. DOI: 10.3856/vol46-issue4-fulltext-11
- Katsanevakis S., Ch. D. Maravelias (2008).** Modelling fish growth: multi-model inference as a better alternative to a priori using von Bertalanffy equation. [Fish and Fisheries](#) 9(2)178-187. [10.1111/J.1467-2979.2008.00279.X](https://doi.org/10.1111/J.1467-2979.2008.00279.X)
- Lang E.T., B. J. Falterman, L. L. kitchens, and C.D. Marshall (2019)** Age and growth of yellowfin tuna (*thunnus albacares*) in the northern Gulf of Mexico. *Collect. Vol. Sci. Pap. ICCAT*, 73(1): 423-433.



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original



- Luquin-Covarrubias, M. A., Morales-Bojórquez, E., González-Peláez, S. S., Hidalgo-De-La-Toba, J. Á., & Lluch-Cota, D. B. (2016).** Modeling of growth depensation of geoduck clam *Panopea globosa* based on a multimodel inference approach. *Journal of Shellfish Research* 35 (2): 379-387. <https://doi.org/10.2983/035.035.0212>
- Quince CH., P. A. Abrams, B. J. Shuter, N. P. Lester (2008).** Biphasic growth in fish I: Theoretical foundations. *Journal of Theoretical Biology*, 254 (2) : 197 - 206 . <https://doi.org/10.1016/j.jtbi.2008.05.029>
- R Core Team (2025).** R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.
- Richards, J. (1959).** A flexible growth function for empirical use. *Journal of Experimental Botany* 10: 290-300
- Rodríguez-Domínguez, G., Aragón-Noriega, E. A., Payán-Alejo, J., Mendivil-Mendoza, J. E., Curiel-Bernal, M. V., Valenzuela-Quinonez, W., & Urías-Sotomayor, R. (2024).** The Fractal Approach to Describe Growth of Farmed Marine Species: Using Double and Triple Logistic Models. *Fishes*, 9:106. <https://doi.org/10.3390/fishes9030106>.
- Ruíz-Velazco, J. M. J., Hernández-Llamas, A., Gómez-Muñoz, V. M., & Magallón, F. J. 2010.** Dynamics of intensive production of shrimp *Litopenaeus vannamei* affected by white spot disease. *Aquaculture* 300: 113–119.
- Schnute, J. (1981).** A versatile growth model with statistically stable parameters. *Can. J. Fish. Aquat. Sci.* 38(9): 1128-1140. <https://doi.org/10.1139/f81-153>
- Schnute, J.T. & L. J. Richards. (1990).** A unified approach to the analysis of fish growth, maturity, and survivorship data. *Can. J. Fish. Aquat. Sci.* 47: 24-40.
- Secor D.H., R.L. Wingate, J.D. Neilson, J.R. Rooker, and S.E. Campana (2009):** Growth of Atlantic bluefin tuna: direct age estimates. *Collect. Vol. Sci. Pap. ICCAT*, 64(2): 405-416.
- Von Bertalanffy, L. (1938).** A quantitative theory of organic growth. *Human Biology*, 10(2): 181-213.



OPEN ACCESS

Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons Atribución-No Comercial-Compartir igual (CC BY-NC-SA 4.0), que permite compartir y adaptar siempre que se cite adecuadamente la obra, no se utilice con fines comerciales y se comparta bajo las mismas condiciones que el original