

Estimación temporal del ancho del anillo del árbol *smo66b* de Bolivia

Bladimir Morales,¹[★]

¹Departamento de Estadística, Pontificia Universidad Católica de Chile, Chile

2024 Julio 07

1 MOTIVACIÓN

El Centro Nacional de Información Ambiental (National Centers for Environmental Information, NCEI)¹ ofrece datos, productos y servicios ambientales que abarcan desde las profundidades oceánicas hasta la superficie solar, promoviendo la resiliencia, prosperidad y equidad para las generaciones actuales y futuras. El NCEI gestiona uno de los archivos más completos del planeta, con datos oceánicos, atmosféricos y geofísicos. Cada mes, almacena más de 229 terabytes de datos provenientes de más de 130 plataformas de observación y fuentes ambientales.

Para este proyecto, nos enfocaremos en datos específicos de Bolivia, siguiendo la publicación de Lopez & Villalba (2022). Los datos fueron descargados de la página <https://www.ncei.noaa.gov/access/paleo-search/study/36007> el 2 de julio de 2024. En particular, se estudiará el ancho total del anillo de un árbol con identificación *smo66b*, medido en milímetros (mm), con las siguientes características: latitud -15.967661, longitud -62.227181, elevación mínima de 420 metros (m), elevación máxima de 420 m, el año más antiguo es 103 años calibrados AP (1847 d. C.), y el año más reciente es -56 años calibrados AP (2006 d. C.). En la Figura 1 se presenta la serie de tiempo anual del ancho del anillo del árbol *smo66b* desde 1847 hasta 2006, con un total de 160 observaciones. Al realizar el test de Dickey-Fuller aumentado, se rechaza la hipótesis nula de raíz unitaria, proporcionando evidencia estadística al 5% de significancia de que la serie es estacionaria. Esta conclusión se refuerza con el resultado del test KPSS, que también indica la estacionariedad de la serie.

2 AJUSTE MODELOS

En esta sección, proponemos modelos para ajustar la serie temporal. Dado que se comprobó la estacionariedad, según la Figura 2, se propone ajustar dos modelos: AR(3) y ARMA(3,3), propuesto por Box & Jenkins (1976). Al ajustar ambos modelos, se verificó el supuesto de homocedasticidad mediante el test de Breusch-Pagan, encontrando evidencia estadística al 5% de significancia de que los residuos no son homocedásticos en ambos modelos. Además, los residuos no cumplen el supuesto de normalidad según el test de Lilliefors (Kolmogorov-Smirnov), rechazándose la hipótesis nula de normalidad de los residuos al 5% de significancia. Por esta razón, se aplicó una transformación de Box-Cox a los datos, obteniéndose un $\lambda = -0.0429$ mediante el método propuesto por Guerrero, lo cual sugiere la necesidad de esta transformación. La serie transformada cumple con el supuesto de estacionariedad al 5% de significancia, según los tests de Dickey-Fuller aumentado y KPSS. La Figura 3

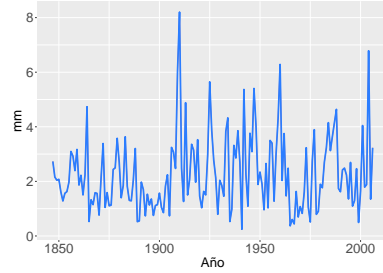


Figure 1. Serie de tiempo del ancho total del anillo del árbol *smo66b* medido en milímetros (mm)

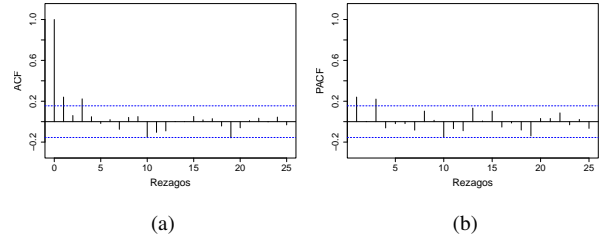


Figure 2. (a) Función de autocorrelación (ACF) y (b) función de autocorrelación parcial (PACF) del ancho total del anillo del árbol *smo66b* desde 1847 hasta 2006.

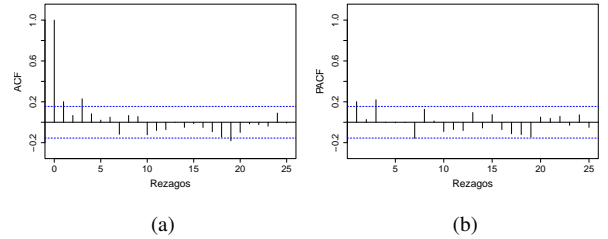


Figure 3. (a) Función de autocorrelación (ACF) y (b) función de autocorrelación parcial (PACF) de la serie transformada por Box-Cox.

muestra el ACF y PACF de la serie transformada, manteniendo la propuesta de los modelos AR(3) y ARMA(3,3).

Para realizar el ajuste, se consideró la serie transformada con Box-Cox y además se centralizó. Se utilizó el algoritmo de Durbin-Levinson para ajustar los modelos AR(3) y ARMA(3,3), como se describe en Palma (2016). En la Figura 4a, se puede observar la función de autocorrelación (ACF) empírica frente a la estimada, con sus respectivas bandas de confianza, para el modelo AR(3). De manera similar, en la Figura 4b se presenta para el modelo ARMA(3,3). La

¹ <https://www.ncei.noaa.gov/>

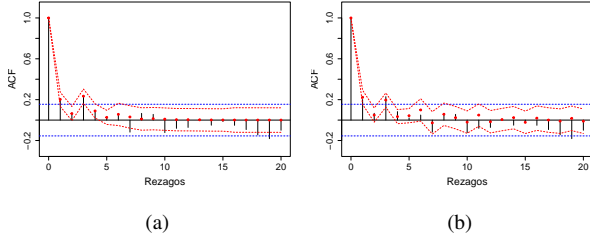


Figure 4. Función de autocorrelación (ACF) empírica versus estimada por (a) AR(3) y (b) ARMA(3,3).

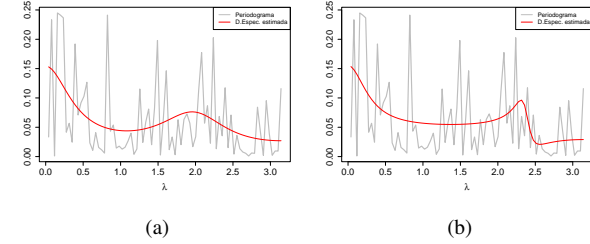


Figure 5. Dendrogram versus densidad espectral estimada (a) AR(3) y (b) ARMA(3,3).

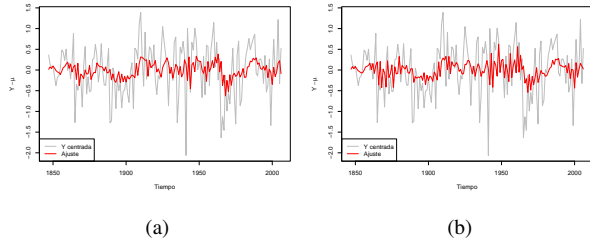


Figure 6. Serie transformada centrada versus ajuste estimado por (a) AR(3) y (b) ARMA(3,3).

Figura 5a muestra el periodograma comparado con la densidad espectral estimada del modelo AR(3), y la Figura 5b hace lo propio para el modelo ARMA(3,3). Finalmente, en las Figuras 6a y 6b, se visualiza el ajuste de los modelos AR(3) y ARMA(3,3), respectivamente, utilizando la serie transformada y centralizada.

La Tabla 1 presenta el vector de parámetros ϕ con sus respectivos estimadores $\hat{\phi}$ del modelo AR(3), así como el estadístico Z_0 , que distribuye normal por sus propiedades asintóticas, y su correspondiente p -valor. En esta tabla se puede observar que el estimador ϕ_2 , correspondiente al rezago 2, no es significativo al 5% para el modelo. La Tabla 2 muestra el vector de parámetros ϕ y θ , con sus respectivos estimadores $\hat{\phi}$ y $\hat{\theta}$ del modelo ARMA(3,3), junto con el estadístico Z_0 y su p -valor. Aquí se puede ver que los estimadores ϕ_2 y θ_2 no son significativos al 5% para el modelo. En el análisis de diagnóstico de los residuos, al realizar el test de Dickey-Fuller aumentado para verificar la estacionariedad, se rechaza la hipótesis nula con un p -valor de 0.01 tanto para el modelo AR(3) como para el modelo ARMA(3,3). Esto proporciona evidencia estadística para afirmar que los residuos son estacionarios. La misma conclusión se obtiene con el test KPSS. En las Figuras 7a y 7b, correspondientes a la función de autocorrelación ACF de los modelos AR(3) y ARMA(3,3), se observa que los residuos son estacionarios. En las Figuras 8a y 8b, todos

ϕ	$\hat{\phi}$	Z_0	p -valor
ϕ_1	0.19	2.50	0.01
ϕ_2	-0.02	-0.27	0.79
ϕ_3	0.23	2.91	0.00

Table 1. Parámetros estimados con el algoritmo de Durbin Levinson y significancia del modelo AR(3)

ϕ	$\hat{\phi}$	Z_0	p -valor	θ	$\hat{\theta}$	Z_0	p -valor
ϕ_1	-0.55	-2.63	0.01	θ_1	0.80	3.48	0.00
ϕ_2	0.16	0.69	0.49	θ_2	-0.03	-0.10	0.92
ϕ_3	0.60	4.50	0.00	θ_3	-0.46	-2.73	0.01

Table 2. Parámetros estimados con el algoritmo de Durbin Levinson y significancia del modelo ARMA(3,3)

los estimadores están dentro del círculo unitario, cumpliendo así con el supuesto de invertibilidad para ambos modelos. En las Figuras 9a y 9b, se puede observar que hasta el rezago 20 se cumple el supuesto de blancura para los residuos, según el test de Box Ljung. En este test, no se rechaza la hipótesis nula de ruido blanco al 5% de significancia para ambos modelos. Para evaluar la homocedasticidad, se aplicó el test de Breusch-Pagan, donde la hipótesis nula establece que los residuos son homocedásticos. Para el modelo AR(3), se obtuvo un valor p de 0.00812, lo cual no es suficiente estadísticamente para rechazar la hipótesis nula (H_0). De manera similar, para el modelo ARMA(3,3), se obtuvo un valor p de 0.00459, también sin suficiente evidencia estadística para rechazar H_0 . Por lo tanto, se dispone de evidencia estadística al nivel del 5% para concluir que los errores en ambos modelos son homocedásticos. Este hallazgo se respalda visualmente en las Figuras 10a y 10b. Para evaluar la normalidad de los residuos, se aplicó el test de Lilliefors (Kolmogorov-Smirnov), donde la hipótesis nula establece que los residuos siguen una distribución normal. En el caso del modelo AR(3), se obtuvo un valor p de 0.08557, mientras que para el modelo ARMA(3,3) fue de 0.05884. En ambos casos, no se encontró suficiente evidencia para rechazar la hipótesis nula (H_0) al nivel del 5%, lo que sugiere que los errores se distribuyen normalmente. Los QQ-plots correspondientes se presentan en las Figuras 11a y 11b. En cuanto a la bondad de ajuste, se observa que el modelo AR(3) tiene un AIC de 286.97 y un BIC de 290.04, mientras que el ARMA(3,3) presenta un AIC de 283.99 y un BIC de 287.06. En términos de significancia, el modelo AR(3) muestra un parámetro no significativo, mientras que el ARMA(3,3) presenta dos parámetros no significativos. En el análisis de diagnóstico, el modelo AR(3) exhibe un mejor comportamiento, cumpliendo con todos los supuestos con una significancia del 1%. Por otro lado, el ARMA(3,3) muestra resultados cercanos al límite del 5% de significancia en algunos tests. En cuanto a la bondad de ajuste, el modelo ARMA(3,3) presenta un mejor ajuste. Basándonos en los principios de parsimonia, significancia estadística y cumplimiento de supuestos, seleccionamos al modelo AR(3) como el mejor candidato para ajustar la serie en estudio.

Procedemos ahora con el ajuste del modelo AR(3) utilizando el método de Whittle (ver Palma (2016)). En la Tabla 3, se observa que una vez más el parámetro ϕ_2 no resulta significativo para el modelo. A pesar de esto, los estimadores obtenidos son bastante similares a los del algoritmo de Durbin Levinson, como se muestra en la Tabla 1. Para evaluar la precisión de los diferentes métodos utilizados en la estimación, la Tabla 4 presenta medidas como el error medio

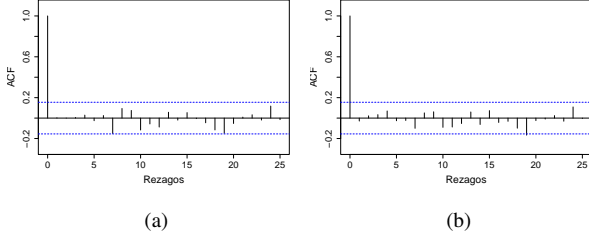


Figure 7. Función de autocorrelación ACF de los residuos del modelo (a) AR(3) y (b) ARMA(3,3).

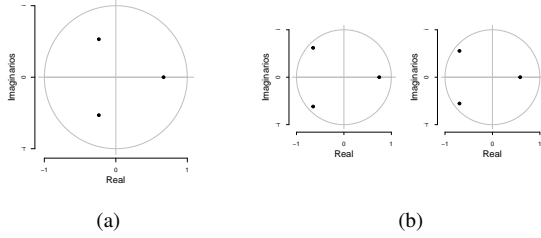


Figure 8. Círculo unitario de los estimadores del modelo (a) AR(3) y (b) ARMA(3,3).

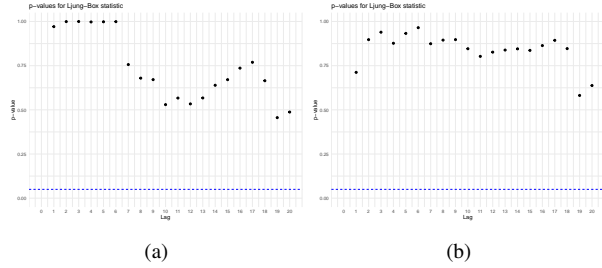


Figure 9. Test de Blancura de Box Ljung del modelo (a) AR(3) y (b) ARMA(3,3).

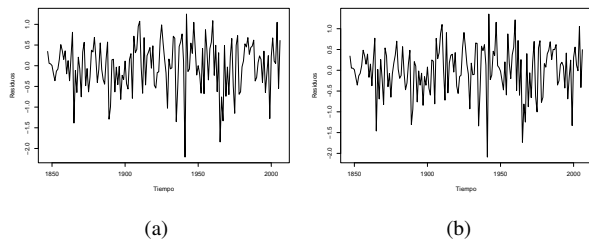


Figure 10. Residuos versus el tiempo del modelo (a) AR(3) y (b) ARMA(3,3).

(ME), la raíz cuadrada del error cuadrático medio (RMSE), el error absoluto medio (MAE), el error porcentual medio (MPE), el error absoluto medio porcentual (MAPE) y el error absoluto medio escalado (MASE) para cada método empleado por la función Arima() de R, el algoritmo de Durbin Levinson y el método de Whittle en el modelo AR(3). Se observa que varias de estas medidas son similares entre los métodos, aunque el método de Whittle muestra diferencias, especialmente en ME, MPE y MAPE, donde presenta menores valores de error.

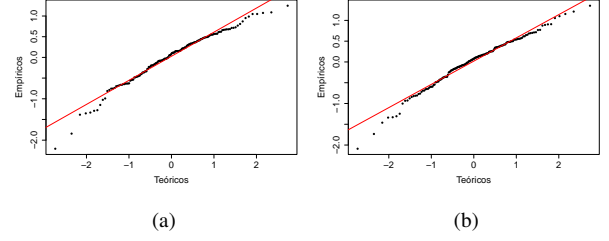


Figure 11. QQ-plot de los residuos del modelo (a) AR(3) y (b) ARMA(3,3).

ϕ	$\hat{\phi}$	Z_0	p -valor
ϕ_1	0.19	2.42	0.02
ϕ_2	0.00	0.06	0.95
ϕ_3	0.22	2.84	0.00
σ	0.60	-	-

Table 3. Parámetros estimados con el método de Whittle y significancia del modelo AR(3)

Medida	Arima()	Durbin-Levinson	Whittle
ME	0.0023	0.0023	0.0022
RMSE	0.5891	0.5891	0.5893
MAE	0.4643	0.4643	0.4651
MPE	80.701	80.721	68.590
MAPE	203.93	203.91	194.59
MASE	0.7877	0.7877	0.7889

Table 4. Medidas de evaluación del error de la función Arima() de R, el algoritmo de Durbin Levinson y el método de Whittle para el modelo AR(3).

3 CONCLUSIONES

Este estudio investigó métodos de estimación para la serie del ancho total del anillo de un árbol identificado como *smo66b*. Inicialmente, los datos fueron transformados mediante Box-Cox debido a la falta de homocedasticidad y normalidad, confirmada mediante pruebas de significancia estadística. Se ajustaron dos modelos propuestos, AR(3) y ARMA(3,3), utilizando el algoritmo de Durbin Levinson, evaluando la significancia de los parámetros, verificando los supuestos del modelo y analizando la bondad de ajuste de ambos modelos.

Se optó por el modelo AR(3) debido a su parsimonia, significancia y cumplimiento de los supuestos al nivel de significancia del 1%. Se estimó utilizando el método de Whittle y se compararon las estimaciones utilizando medidas de error. Se observaron medidas de error similares entre los métodos, aunque el método de Whittle mostró menores valores en error medio ME, error porcentual medio MPE y error absoluto medio porcentual MAPE.

REFERENCES

- Box G., Jenkins G. M., 1976, Time Series Analysis: Forecasting and Control. Holden-Day
- Lopez L., Villalba R., 2022, NOAA National Centers for Environmental Information
- Palma W., 2016, Time Series Analysis. Wiley Series in Probability and Statistics, Wiley, <https://books.google.cl/books?id=hubGBwAAQBAJ>