

Tarea 2 - Series de Tiempo EYP3907

Diego Aravena - Alonso Campos

17 de Noviembre de 2023

Introducción

La dendrocronología, el estudio de los anillos de crecimiento de los árboles, ofrece una ventana única para entender la respuesta de la vegetación a su entorno a lo largo del tiempo. Este informe se centra en la aplicación del Índice de Crecimiento Estandarizado (ICR) anual, una medida que captura las variaciones en el crecimiento de los anillos. Este índice, derivado de la densidad de los anillos, nos brinda una perspectiva detallada de la historia de vida de los árboles. El objetivo de este trabajo es encontrar el mejor modelo ARMA que logre explicar este fenómeno.

Metodología

Datos

Se logró emplear datos del estudio realizado por Holmes¹ extraídos de World Data Service for Paleoclimatology que administra el Banco Internacional de Datos de Anillos de Árboles. Este estudio de paleoclimatología se basa en el crecimiento de los anillos del Cedro Chileno en la Región del Libertador General Bernardo O'Higgins. Los años que serán considerados datan del 1606 hasta 1975.

Primeras impresiones

En los datos originales (véase Figura 1) se puede percibir una media constante, mas no de varianza. Además, se trataría de un proceso de corta memoria, lo cual significa que el ICR no dependerá de tiempos muy lejanos, sino que de los más recientes.

Modelamiento

En un principio se utiliza el modelo propuesto por la función auto.arima de R el cual ajusta un modelo ARMA(3,2). Para mejorar dicha propuesta, se evaluaron diferentes modelos considerando la significancia de los coeficientes y la blancura del ruido, donde finalmente se optó por un modelo ARMA(1, 36).

$$Y_t = \theta_7 Z_{t-7} + \theta_{36} Z_{t-36} + Z_t + \phi_1 Y_{t-1} \text{ con } Z \sim RB(0, \sigma^2)$$

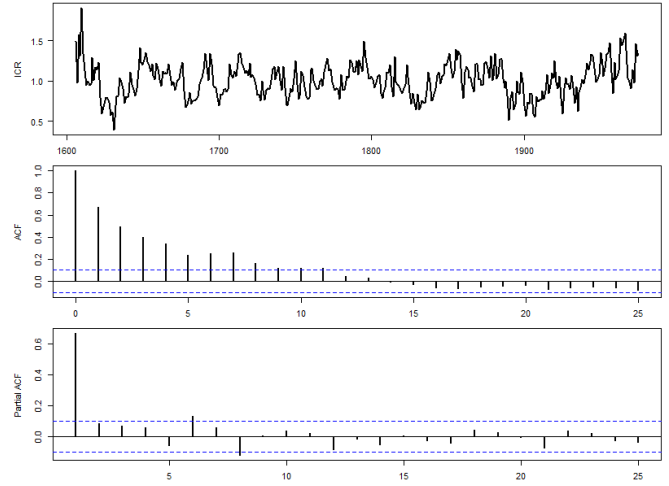


Figura 1: Serie datos originales junto a su ACF y PACF

Tabla 1: Tabla de resumen coeficientes modelo ARMA(1,36)

	Estimación	Error est.	Estadístico Z	Valor-p
ϕ_1	0.668	0.040	16.937	0.000
θ_7	0.198	0.053	3.737	0.000
θ_{36}	-0.109	0.056	-1.937	0.053

Significancia de los coeficientes Los resultados del ajuste se pueden observar Tabla 1, donde todos los coeficientes son significativos al 10 %. Adicionalmente, se debe mencionar que la media fue no significativa con un valor-p de 0.85.

Estacionaridad en media En la Figura 2 se observa que los residuos del modelo parecieran tener una media constante. Dicho de este modo, se ajusta una regresión lineal de residuos vs tiempo, obteniendo que la significancia del tiempo es 0.52, por ende, se concluye que no existe pendiente.

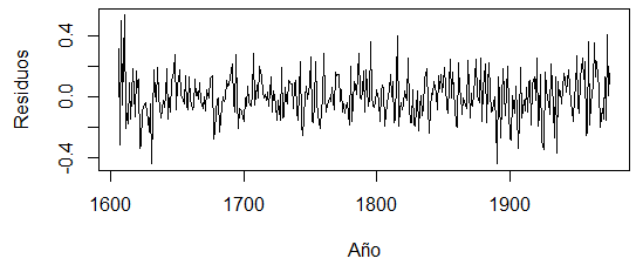


Figura 2: Residuos modelo ARMA(1, 36)

¹ Holmes, R.L. (2002-04-26): NOAA/WDS Paleoclimatology - Holmes - Santa Isabel de las Cruces - AUCH - ITRDB CHIL013. chil013-crn-noaa. NOAA National Centers for Environmental Information. <https://doi.org/10.25921/85tm-fb45>. Accessed 2023-11-17.

Homocedasticidad En primer lugar los residuos del modelo no pasaban el supuesto de normalidad, por lo que se evaluó una transformación mediante el método de Box-Cox utilizando el criterio de log-verosimilitud, el cual

propuso una transformación raíz cuadrada. Posterior a ello, primero se comprobó homocedasticidad con el test Breusch-Pagan, que entregó un valor-p de 0.58, logrando el supuesto y, como adelanto, se obtuvo normalidad.

Invertibilidad y causalidad En la Figura 3 se puede observar que todas las raíces de la componente de media móvil y autoregresivo están contenidas dentro del círculo unitario, lo que garantiza invertibilidad y causalidad del proceso respectivamente.

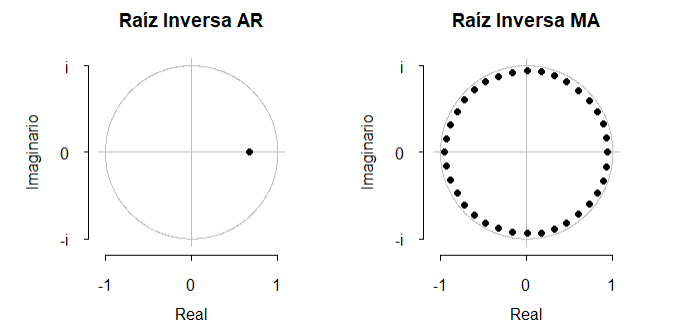


Figura 3: Gráfico raíces inversas ARMA(1,36)

Blancura de los residuos Como se puede apreciar en la Figura 4 todos los valores se encuentran sobre un nivel de significancia del 5%, por lo que no se rechaza la hipótesis que los residuos están incorrelacionados. Cabe señalar que se pueden apreciar caídas importantes en ciertos rezagos, por lo que al ser modeladas se podría lograr un mejor ajuste, sin embargo, los coeficientes resultantes resultaron ser no significativos utilizando $\alpha = 0.1$.

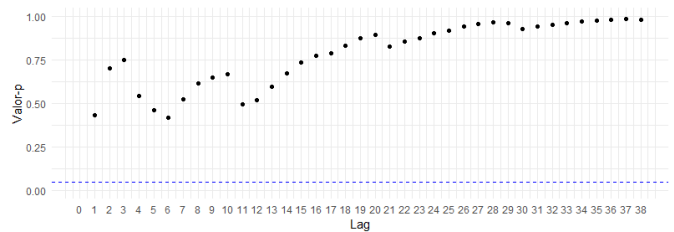


Figura 4: Gráfico Box-Ljung para residuos del modelo ARMA(1, 36)

En complemento, en la Figura 5 se aprecia que todos los valores de la autocorrelación se encuentran por debajo de la bandas de confianza, lo cual apoya la conclusión obtenida previamente. Por lo tanto, estaríamos frente a un ruido blanco. Y a modo de comentario, el gráfico PACF de la Figura 6 nos da indicios de que el modelo estaría capturando la estructura autoregresiva de los datos.

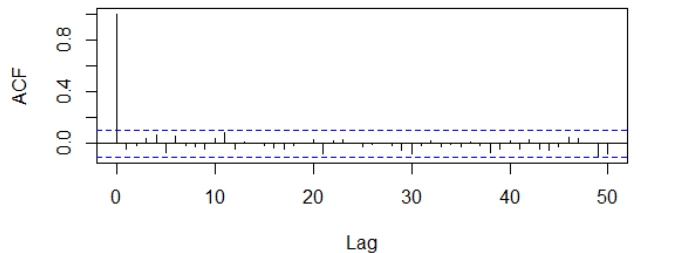


Figura 5: Gráfico ACF para residuos del modelo ARMA(1, 36)

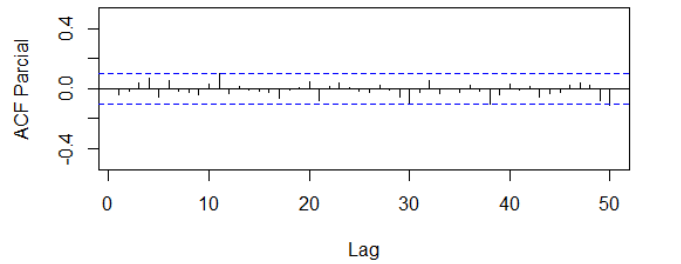


Figura 6: Gráfico PACF para residuos del modelo ARMA(1, 36)

Normalidad Finalmente, luego de transformar los datos, se comprueba la normalidad de los residuos. A partir de los resultados disponibles en la Tabla 2 se concluye que los residuos estandarizados distribuyen normalmente con media 0 y varianza 1, ya que en ambos test no se rechaza la hipótesis nula. .

Tabla 2: Valores-p para test de Normalidad

	Valor-p
Shapiro-Wilks	0.057
Kolmogorov-Smirnov	0.299

En el QQ-Plot de la Figura 7 se observa que los cuantiles teóricos son similares a los muestrales, lo que tiene sentido dado los test vistos recientemente. Además el nivel de curtosis de los residuos es de 3.62, el cual difiere levemente con el de una normal, posiblemente por los datos que se alejan en las colas

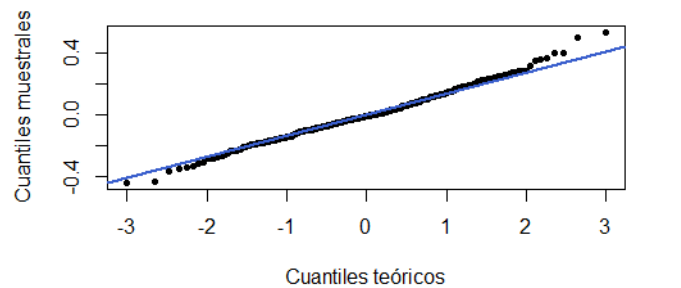


Figura 7: Gráfico cuantil-cuantil residuos ARMA(1,36)

Comparación ACF teórico En base a la Figura 8 se aprecia que el modelo logra ajustar adecuadamente a los datos de anillos de árboles, pues ninguno de los valores empíricos de las autocorrelaciones están contenidos fuera del intervalo de confianza.

Densidad espectral y periodograma La Figura 9 muestra el periodograma (línea gris), la densidad espectral (línea azul) y un gráfico suavizado de la densidad espectral (interlineado). Se puede concluir que el modelo propuesto ajusta a la serie de anillos, dado que tanto la densidad espectral como el periodograma suavizado siguen la misma tendencia, lo que nos da un buen indicio sobre la calidad de ajuste.

Ajuste Durbin-Levinson Mediante el algoritmo de Durbin-Levinson es posible obtener estimaciones de los

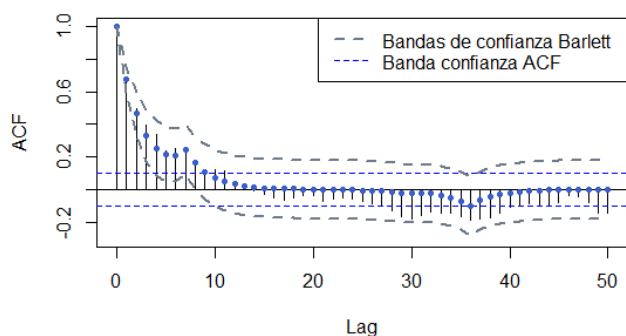


Figura 8: Comparación ACF empírico y teórico

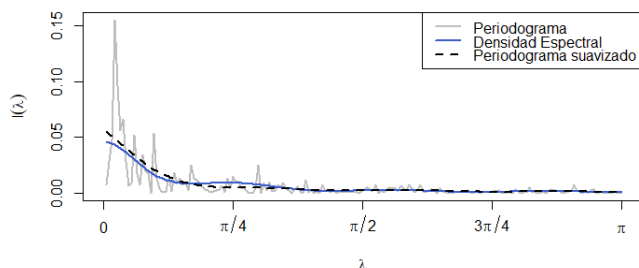


Figura 9: Comparación densidad espectral y periodograma

parámetros. Encima, como este método se basa en predicciones a un paso también proporciona los valores ajustados del modelo, de esta forma es posible realizar una comparación entre la serie de anillos y los valores ajustados mediante el método de Durbin-Levinson.

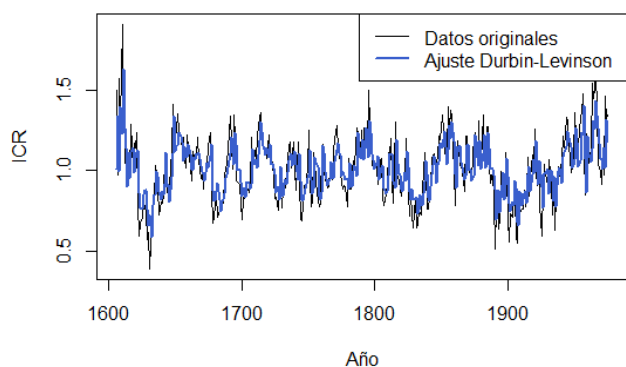


Figura 10: Datos originales y ajuste Durbin-Levinson

La Figura 10 se aprecia que el método de Durbin-Levinson logra ajustar a la serie de crecimiento de tiempo, pues al comparar reales con los valores ajustados, éstos se superponen y siguen el mismo patrón.

Estimación método de Whittle En esta sección se utilizará el método de Whittle para estimar los coeficientes del modelo ARMA(1, 36).

En la Tabla 3 se puede observar que si bien existen diferencias en las estimaciones, éstas no difieren notoriamente. Aunque la estimación de σ , considerando la magnitud, sí posee una mayor diferencia ya que la estimación de Arima es el doble que mediante Whittle.

En la Tabla 4 se comparan diferentes métricas de error para los métodos de estimación de la función Arima y el método Whittle. En las métricas RMSE, MAE, MAPE y MASE menores valores representan un mejor ajuste del

Tabla 3: Comparación estimaciones coeficientes método Arima y Whittle

	Arima	Whittle
ϕ_1	0.6685	0.6738
θ_7	0.198	0.2086
θ_{36}	-0.109	-0.0691
σ	0.1492	0.074

Tabla 4: Comparación medidas de error, estimación Arima y método Whittle

	Arima	Whittle
ME	0.005	0.005
RMSE	0.150	0.151
MAE	0.115	0.115
MPE	-1.750	-1.731
MAPE	11.728	11.757
MASE	0.885	0.888
ACF1	-0.061	-0.065

modelo. Se aprecia que el método de estimación de la función Arima posee menores valores en RMSE, MAPE y MASE, mientras que MAE en ambos métodos coinciden, por lo que el modelo ARIMA daría señales leves de un mejor ajuste. El ME y el MPE corresponden a medias de los errores y errores porcentuales respectivamente, por lo que son estadísticos de resumen de los residuos, por otro lado la métrica ACF1 es la autocorrelación de los errores con un desfase, por lo que un menor valor (en valor absoluto) es indicio de un mejor ajuste. De este modo, el modelo Arima tendría un mejor ajuste.

Finalmente, es posible concluir que el método de estimación de la función Arima de R presenta un mejor ajuste en comparación que la estimación del método de Whittle, pues posee resultados superiores en las métricas, o al menos igual. No obstante, cabe señalar que ninguna de las métricas obtuvo un resultado notoriamente diferente.

Conclusión

En síntesis, se realizó un ajuste mediante la función `auto.arima()` como base para realizar mejoras que satisficieran la significancia de los coeficientes y la incorrelación de los residuos. De esta forma se ajusta un modelo ARMA(1, 36), donde solo se especificaron algunos coeficientes de tanto del componente autorregresivo y media móvil. Posteriormente, se realiza un estudio de los supuestos y bondad de ajuste del modelo mediante las diferentes técnicas. Luego, se comparan métricas de error de las estimaciones realizadas por la función Arima y el método Whittle. Finalmente, se concluye que el modelo logró ajustarse a los datos, e incluso para futuras investigaciones en conjunto con mayor información climática/vegetal podremos conocer el comportamiento que tuvo o tendrá el Cedro Chileno dentro de un período de tiempo en el que estemos interesados.