

# Análisis de Series Temporales

Maestría en Ciencia de Datos  
Universidad Austral

## Resumen

Este informe presenta un análisis detallado de tres series temporales seleccionadas, aplicando distintos métodos de modelado y diagnóstico para evaluar su comportamiento y realizar pronósticos. Se exploran conceptos de estacionariedad, correlaciones y modelos SARIMA y VAR, junto con métricas de performance y análisis de residuos.

**Autores:** Astegiano Franco, Fierro Abel, Santesteban Natalia

# Índice

<b>1. Introducción</b>	<b>2</b>
1.1. Series e Ingeniería de variables. . . . .	2
1.2. Inspección visual . . . . .	4
<b>2. Tratamiento de series y modelos estadísticos.</b>	<b>5</b>
2.1. Análisis de series . . . . .	5
2.1.1. Transformaciones aplicadas . . . . .	5
2.1.2. Estacionariedad . . . . .	5
2.1.3. Interpretabilidad . . . . .	8
2.1.4. Series transformadas y diferenciadas . . . . .	9
2.1.5. Autocovarianzas, autocorrelaciones y autocorrelaciones parciales. . . . .	9
2.2. Modelos . . . . .	14
2.2.1. Auto ARIMA - Serie Rendimientos . . . . .	14
2.2.2. ARIMA(2,0,0) - Serie Rendimientos . . . . .	15
2.2.3. Auto ARIMA - Serie proxy Temperatura . . . . .	16
2.2.4. Auto ARIMA - Serie proxy Precipitaciones . . . . .	18
2.2.5. ARIMA(2,0,0) - Serie proxy Precipitaciones . . . . .	19
2.2.6. ARIMA-X - Rendimientos . . . . .	21
<b>3. Análisis de Resultados</b>	<b>23</b>
3.1. Selección de modelos . . . . .	23
3.2. Predicciones . . . . .	24
<b>4. Modelado VAR y Análisis de Causalidad</b>	<b>27</b>
4.1. Cointegración . . . . .	29
4.2. VECM . . . . .	30
<b>5. Conclusiones</b>	<b>34</b>
<b>6. Referencias Bibliográficas</b>	<b>36</b>
6.1. Fuentes de datos . . . . .	36
<b>7. Apéndice</b>	<b>37</b>

# 1. Introducción

En esta sección se contextualiza la elección de las series, se explicará la ingeniería de variables y los criterios mediante los cuales se han unificado frecuencias temporales.

## 1.1. Series e Ingeniería de variables.

El objetivo de éste escrito es analizar y generar modelos de predicción para series de:

- Rendimientos de las campañas de soja.
- Proxy de temperatura.
- Proxy de precipitaciones.

La campaña de soja en Argentina comienza habitualmente en el mes de Octubre, con la siembra. Hasta diciembre ocurre el proceso de germinación y crecimiento vegetativo. Durante diciembre y enero ocurre la maduración de la planta y la floración. Entrado febrero comienza la etapa de llenado del grano, donde adquiere la humedad necesaria para su posterior cosecha a partir de Marzo.

La serie de rendimientos se ha obtenido del portal del Ministerio de Agricultura, Ganadería y Pesca de la República Argentina. La serie presenta un valor para cada campaña, medido en [kg/Ha]. A su vez, las series de precipitaciones y temperaturas, han sido obtenidas a través del portal del Servicio meteorológico nacional.

El foco se hará sobre la provincia de Buenos Aires ya que presenta registros climatológicos desde 1982 y es la provincia con mayor participación en la producción nacional de soja. La lista de departamentos estudiados se encuentra en el apéndice.

Para llevar a cabo el estudio, se aplicó un proceso de ingeniería de características que transforma los datos diarios de temperatura y precipitación en indicadores representativos del ciclo sojero en Buenos Aires. Para la variable de temperatura, se calculó el cociente entre la temperatura máxima y la amplitud térmica diaria, lo que permite capturar la intensidad del calor en función de la variabilidad diurna; estos valores se promedian en bloques bimensuales (Octubre–Noviembre, Diciembre–Enero y Febrero–Marzo) y luego

se combinan mediante un promedio ponderado con coeficientes de 0.4, 0.4 y 0.2, respectivamente, reflejando la importancia relativa de cada etapa del cultivo. De forma similar, para la variable de precipitación se suma el total de milímetros de lluvia en cada bloque bimensual y se obtiene un indicador final mediante un promedio ponderado con coeficientes 0.3, 0.5 y 0.2. Este enfoque permite reducir la dimensionalidad de los datos diarios, preservando la lógica de los ciclos críticos de la soja y facilitando el análisis de series de tiempo para evaluar la influencia de las condiciones climáticas en el rendimiento del cultivo.<sup>1</sup>

---

<sup>1</sup>Los coeficientes utilizados para la ponderación fueron elegidos en base a la interpretación subjetiva de la importancia de las variable en cada bimestre de los autores.

## 1.2. Inspección visual

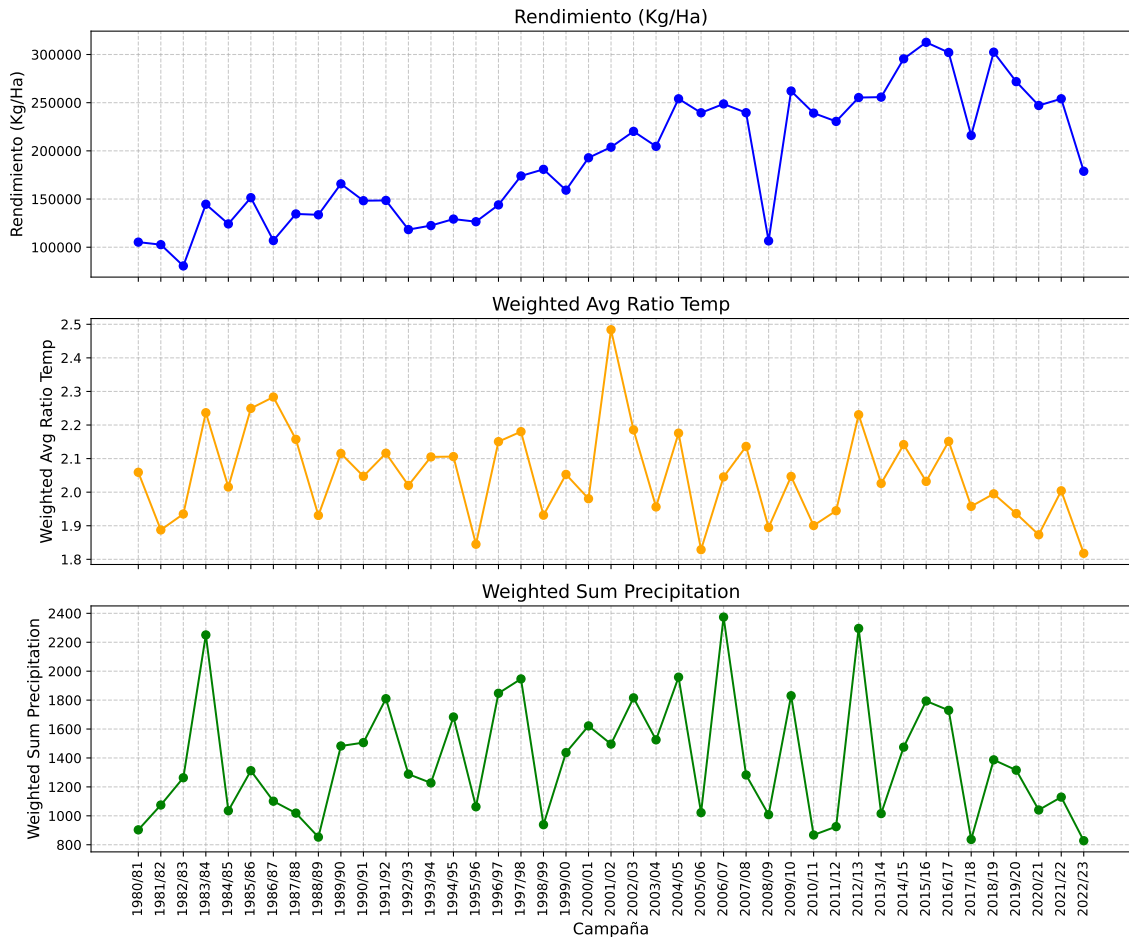


Figura 1: Series de rendimientos, proxy temperatura y proxy precipitaciones.

Se observa una tendencia creciente en los rendimientos, debida a las mejoras tecnologicas posiblemente. Se pueden ver 3 caidas pronunciadas, correspondientes a 3 sequias: 2008/09 - 2017/18 y 2022/23.

Por otro lado, no se observan intensidades de temperaturas crecientes en el tiempo como asi tampoco precipitaciones. Mas bien no pareciera haber variaciones significativas en las medias. Sin embargo, las precipitaciones presentan una alta varianza - con campanas secas de 800mm hasta campanas con mas de 2200mm (promedio ponderado de la campana completa).

## 2. Tratamiento de series y modelos estadísticos.

Esta sección estará dividida en 2 partes: análisis de las series y modelos propuestos.

### 2.1. Análisis de series

#### 2.1.1. Transformaciones aplicadas

**Serie Rendimientos:**

- **Remover outlier 2008/2009:** se priorizó mantener la estructura de una serie con tendencia creciente y con menos varianza, a costas de quitar el outlier causado por la sequía de 2008/09. No remover este outlier, generaba un falso positivo en la prueba ADF, indicando estacionariedad donde no la había.
- **Transformación logarítmica:** se aplica dado que todos los valores son positivos, y contribuye a disminuir la varianza de la serie con la menor pérdida de información.

**Serie Proxy Precipitaciones**

- **Transformación de Box-Cox:** se busca estabilizar la varianza de la serie buscando normalizar los datos. Trabajar sin esta transformación requiere mas de un grado de diferenciación de la serie, con la pérdida de información que eso conlleva. El lambda obtenido es  $\lambda = -0,47$ .

#### 2.1.2. Estacionariedad

La condición de estacionariedad de una serie refiere a la estabilidad en el tiempo de sus propiedades estadísticas, particularmente sobre la media y la varianza. Es decir, una serie no sera estacionaria cuando su media y/o varianza cambien en el tiempo.

Una definición sencilla de un proceso estacionario es la siguiente:

Un proceso es estacionario si, para todos los posibles rezagos  $k$ , la distribución de  $y_t, y_{t+1}, \dots, y_{t+k}$  no depende de  $t$ .

Nielsen (2019)

La no estacionariedad puede manifestarse a través de:

- Estacionalidad: patrones que se repiten en el tiempo.
- Tendencia: media creciente o decreciente en el tiempo.
- Varianza cambiante: las fluctuaciones de la serie varían en intensidad en el tiempo.
- Ciclos: patrones de larga duración en el tiempo que pueden traer cambios en la media y varianza.

Las series aquí estudiadas, debido a su frecuencia anual, carecen de estacionalidad. Asimismo, el periodo estudiado no permite detectar ciclos de largo plazo, por ello se hará foco sobre tendencias y varianzas.

Para su evaluación, se utilizarán las pruebas de Dickey Fuller aumentada y KPSS.

### **Prueba de Dickey Fuller aumentada**

La prueba de Dickey-Fuller Aumentada (ADF) es un test estadístico utilizado para determinar la estacionariedad de una serie temporal, analizando la presencia de raíces unitarias en su ecuación característica. Si la serie posee al menos una raíz unitaria, se considera no estacionaria. Sin embargo, la ausencia de una raíz unitaria no garantiza necesariamente la estacionariedad de la serie.

Matemáticamente, la prueba ADF evalúa la siguiente ecuación:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \epsilon_t \quad (1)$$

donde:

- $y_t$  es la serie temporal en el tiempo  $t$ .
- $\Delta y_t = y_t - y_{t-1}$  representa la primera diferencia de la serie.
- $\alpha$  es una constante.
- $\beta t$  es un término de tendencia opcional.
- $\gamma$  es el coeficiente asociado a  $y_{t-1}$ , clave para evaluar la raíz unitaria.
- $\delta_i$  son los coeficientes de los retardos de las diferencias de la serie.
- $\epsilon_t$  es un término de error blanco.

Luego  $H_0$ : existe al menos una raíz unitaria para la ecuación característica.

**Resultados test ADF:** Con un nivel de confianza del 95 % se rechaza la hipótesis nula cuando el valor p del test se encuentra por debajo de 0.05. Se han realizado tests para los términos de constante y tendencia.

- Log - Rendimientos: ADF(constant) p value: 0.40 // ADF(trend) p value: 0.074. En ambos casos no puede descartarse  $H_0$ .
- Box Cox - Proxy precipitaciones: ADF(constant) p value: 2.77e-09 // ADF(trend) p value: 3.7e-08. Se rechaza la  $H_0$ . La serie es estacionaria según ADF.
- Proxy temperaturas: ADF(constant) p value: 0.004 // ADF(trend) p value: 0.03. Se rechaza la  $H_0$ . La serie es estacionaria según ADF.

### Prueba KPSS

La prueba de Kwiatkowski-Phillips-Schmidt-Shin (Kwiatkowski et al. (1992)) es un test estadístico utilizado para evaluar la estacionariedad de una serie temporal. A diferencia de la prueba de Dickey-Fuller aumentada (ADF), que evalúa la presencia de una raíz unitaria, la prueba KPSS parte de la hipótesis nula de que la serie es estacionaria en tendencia o en media.

Luego  $H_0$ : La serie es estacionaria (sin raíz unitaria).

**Resultados test KPSS:** Con un nivel de confianza del 95 % se rechaza la hipótesis nula cuando el valor p del test se encuentra por debajo de 0.05.

- Log - Rendimientos: p value: 0.051 – > No se rechaza  $H_0$ . La serie es estacionaria.
- Box Cox - Proxy precipitaciones: p value: 0.049 – > Se rechaza  $H_0$ . La serie no es estacionaria.
- Proxy temperaturas: p value: 0.1. No se rechaza  $H_0$ . La serie es estacionaria.

### Conclusión estacionariedad

En la siguiente tabla se presentan los resultados, donde SI indica que la prueba concluyó que la serie era estacionaria:



	ADF	KPSS	Conclusión
Log - Rendimientos	NO	SI	NO
Box Cox - Proxy precipitaciones	SI	NO	NO
Proxy temperaturas	SI	SI	SI

Cuadro 1: Resultados de pruebas de estacionariedad

### Diferenciación

Junto con la evaluación de ambos tests, se calculó la cantidad de diferencias requeridas en cada caso para convertir la serie a estacionaria. Se obtuvieron los siguientes resultados:

- Log - Rendimientos:  $\text{ndiff (ADF)} = 1$  ;  $\text{ndiff (KPSS)} = 1$ . Diferencias realizadas: 1.
- Box Cox - Proxy precipitaciones:  $\text{ndiff (ADF)} = 1$  ;  $\text{ndiff (KPSS)} = 2$ . Diferencias realizadas: 1.
- Proxy temperaturas:  $\text{ndiff (ADF)} = 0$  ;  $\text{ndiff (KPSS)} = 0$ . No se han realizado diferencias.

Si bien el test de KPSS arroja que la serie proxy BC precipitaciones debe diferenciarse 2 veces, se continuará el estudio con 1 sólo grado de diferenciación, evitando introducir ruido innecesario.

### 2.1.3. Interpretabilidad

Las variables de rendimiento y precipitaciones debieron ser transformadas en series estacionarias para su posterior ingesta en modelos estadísticos clásicos como lo son ARIMA, ARIMAX y VAR.

Tras aplicar una transformación logarítmica y luego diferenciar, la variable de rendimientos [kg/ha] deja de expresar rendimientos absolutos, y ahora expresa la tasa de cambio porcentual anual de los rendimientos.

Por otro lado, la variable transformada de precipitaciones, si pierde interpretabilidad al aplicarle Box-Cox y diferenciarla. Pero es fácil invertir la transformación y volver a la serie original para poder interpretar valores predichos.

### 2.1.4. Series transformadas y diferenciadas

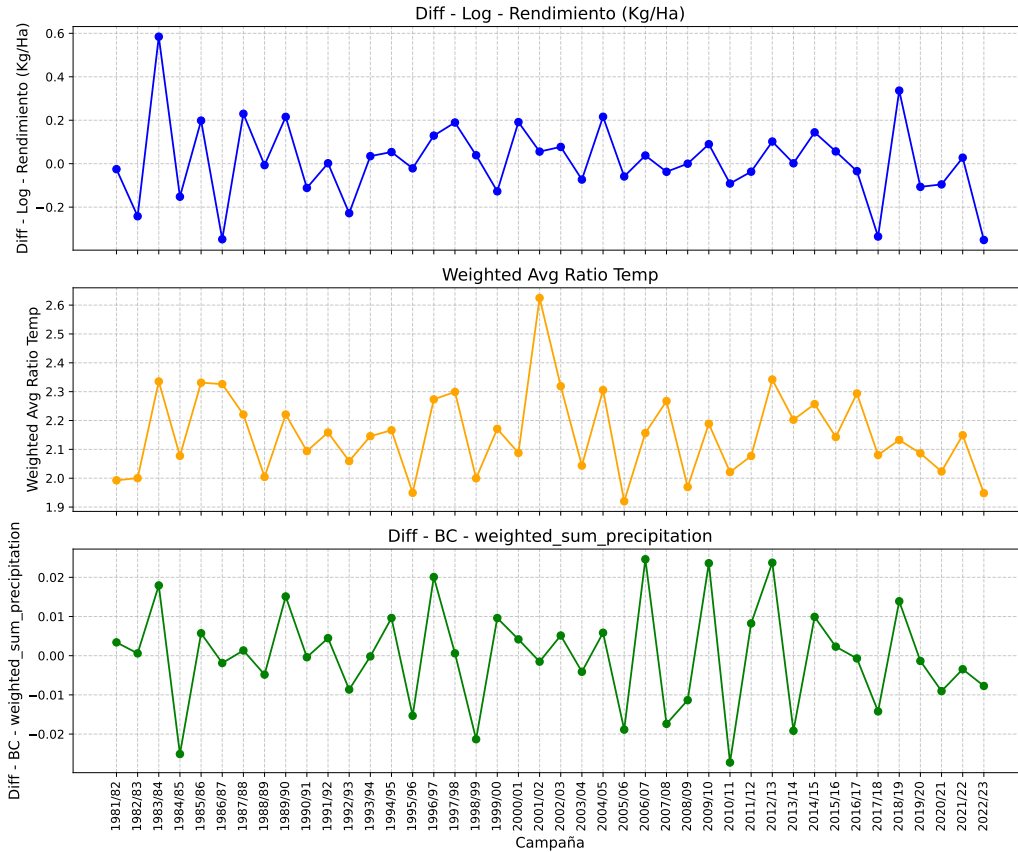


Figura 2: Series de Diff-log-rendimientos, proxy temperaturas y Diff-BC proxy precipitaciones.

Se observa cómo las series presentan media estable, y una varianza mas controlada.

### 2.1.5. Autocovarianzas, autocorrelaciones y autocorrelaciones parciales.

Introduutoriamente se repasará brevemente en qué consiste cada una dada su importancia para comprender la selección de modelos luego.

- **Autocovarianza:** Es una medida de la variación en una serie temporal que se debe a los valores rezagados de la misma serie. Se calcula como la covarianza entre la serie temporal y una versión rezagada de sí misma.

- **Autocorrelación:** Es una medida de la correlación entre una serie temporal y una versión rezagada de sí misma. Se calcula como la correlación entre la serie y su versión rezagada.
- **Autocorrelación parcial:** Es una medida de la correlación entre una serie temporal y una versión rezagada de sí misma, después de eliminar la influencia de los valores intermedios entre los dos puntos en el tiempo. Se calcula restando las correlaciones intermedias antes de obtener la correlación final entre la serie y su rezago.

Estos conceptos son importantes para comprender la dinámica subyacente de una serie temporal y para hacer predicciones sobre valores futuros. (Nielsen, 2019)

La gráfica de ACF<sup>2</sup>, proporciona una buena aproximación del parámetro  $q$  del componente de medias móviles que contenga la serie:  $MA(q)$ . A su vez, la gráfica de PACF<sup>3</sup> permite de igual manera obtener un valor estimado de  $p$ , el parámetro de la componente autorregresiva  $AR(p)$  de la serie.

---

<sup>2</sup>Función de autocorrelación

<sup>3</sup>Función de autocorrelación Parcial

### ACF - PACF serie de Rendimientos

Gráficas de la serie de variación porcentual anual de los rendimientos (Diff - Log - rendimientos (kg/ha) ):

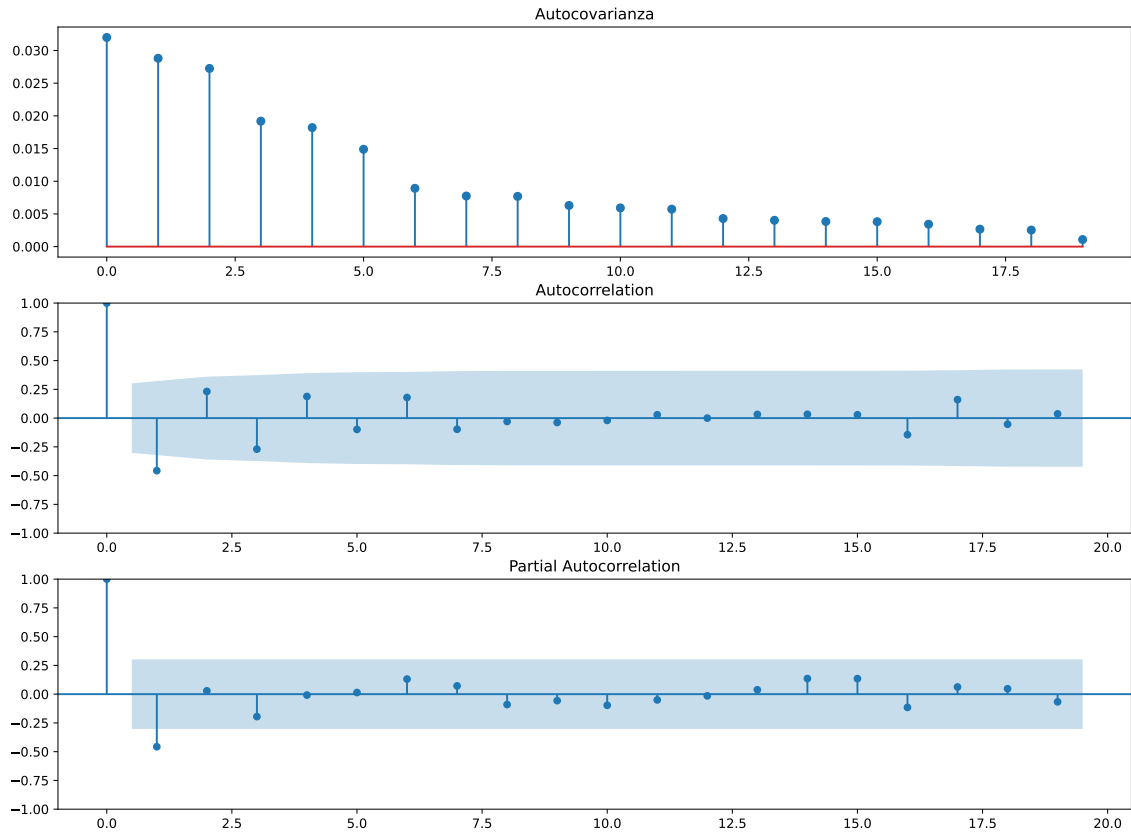


Figura 3: Auto Covarianza, autocorrelación y autocorrelación parcial.

Se observa en la gráfica de ACF un primer lag significativo y los subsecuentes lags convergiendo a 0 no muy rápido. Se podría suponer que  $q = 1$  pero en un modelo MA(1) puro, la caída a 0 en los lags subsiguientes a  $q$  debería ser abrupta. Por otro lado, en el gráfico de PACF, se observa un primer lag significativo, qué a diferencia de ACF, si permite suponer un valor  $p = 1$ . De esta forma se sugiere a priori un modelo ARIMA(1,0,0).

### ACF - PACF serie de Proxy temperaturas

Gráficas de la serie de intensidad de temperaturas:

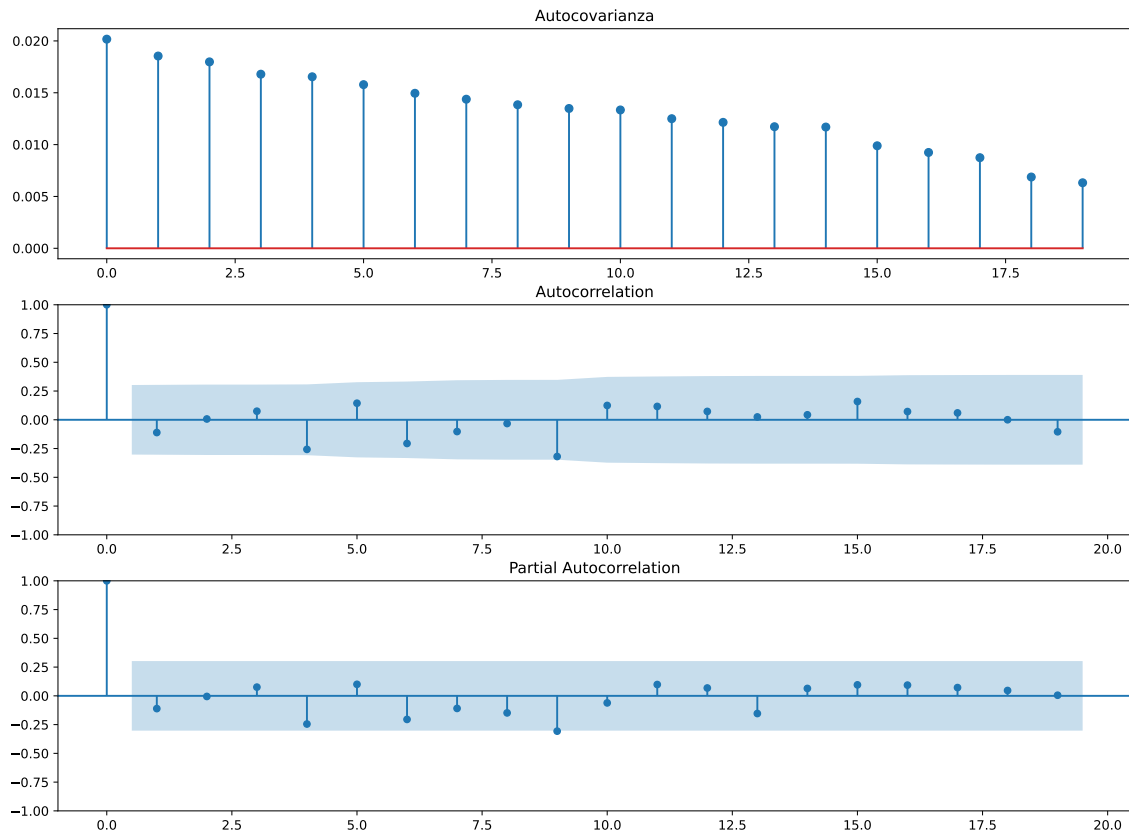


Figura 4: Auto Covarianza, autocorrelación y autocorrelación parcial.

En ambas gráficas no se detectan rezagos (lags) significativos, lo que sugiere que la serie no presenta componentes de medias móviles (MA) ni autorregresivas (AR). En su lugar, el comportamiento observado es consistente con el de un ruido blanco.

### ACF - PACF serie proxy de precipitaciones

Gráficas de la serie de Box-Cox diferenciada para el proxy de precipitaciones:

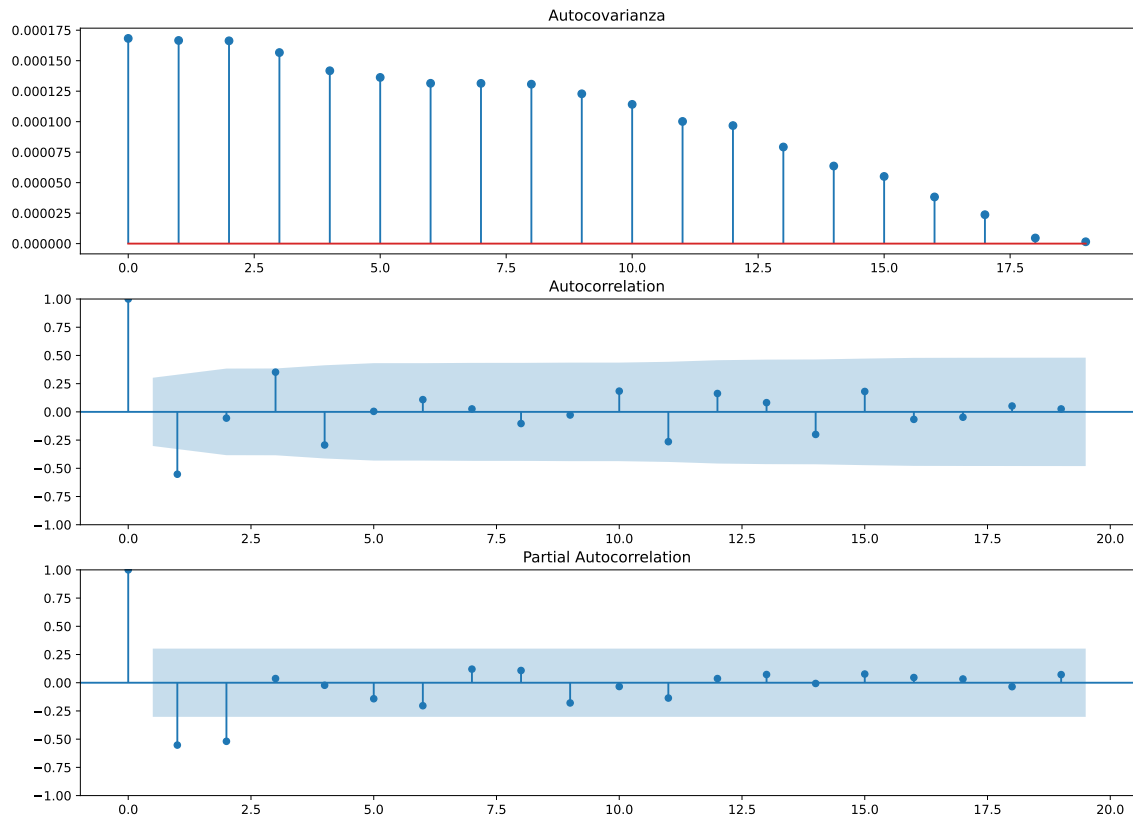


Figura 5: Auto Covarianza, autocorrelación y autocorrelación parcial.

Se observa en la gráfica de ACF un primer lag significativo pero los lags subsiguientes no decaen a 0 instantáneamente. Luego, en PACF se observan los primeros 2 lags significantes.

A priori se propone un modelo  $ARIMA(2,0,0)$ .

## 2.2. Modelos

### 2.2.1. Auto ARIMA - Serie Rendimientos

Tras observar las funciones de autocorrelación y autocorrelación parcial en la Fig.3, se ha propuesto un primer modelo ARIMA(1,0,0). Se ha corrido un modelo auto arima<sup>4</sup> permitiendo la exploración de parámetros p y q, desde el orden 0 al 8.

El modelo obtenido es el siguiente:

ARIMA(1,0,0)

SARIMAX Results	
<b>Dep. Variable:</b>	<i>Diff – Log – Rendimientos(kg/ha)</i>
<b>No. Observations:</b>	34
<b>Sample:</b>	11-01-1981 - 11-01-2014
<b>Model:</b>	SARIMAX(1, 0, 0)
<b>Log Likelihood:</b>	16.580
<b>AIC:</b>	-29.160
<b>BIC:</b>	-26.108
<b>HQIC:</b>	-28.119

Cuadro 2: Resumen del modelo ARIMA(1,0,0)

	coef	std err	z	P>  z	[0.025]	[0.975]
ar.L1	-0.4777	0.180	-2.660	0.008	-0.830	-0.126
sigma2	0.0219	0.006	3.966	0.000	0.011	0.033

Cuadro 3: Coeficientes del modelo

<sup>4</sup>Función del paquete pmdarima de Python que permite de forma iterativa probar múltiples grados de p y q, y devuelve el modelo que minimiza criterios de información como AIC, BIC.

<b>Ljung-Box (L1) (Q):</b>	0.05
<b>Prob(Q):</b>	0.83
<b>Jarque-Bera (JB):</b>	2.47
<b>Prob(JB):</b>	0.29
<b>Heteroskedasticity (H):</b>	0.19
<b>Prob(H) (two-sided):</b>	<b>0.01</b>

Cuadro 4: Pruebas Diagnósticas del Modelo ARIMA(1,0,0)

### Interpretación del modelo

Auto arima corrobora la hipótesis de un modelo ARIMA(1,0,0), de ecuación:

$$y_t = -0,4777 y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 0,0219)$$

El coeficiente  $ar.L1 = -0,4777$  presenta un valor  $p < 0,05$ , rechazando la hipótesis nula de que el coeficiente no es significativo. Esto indica que por cada punto porcentual que aumentó el rendimiento del período anterior, el porcentaje de cambio actual disminuye en 0.47. Contribuyendo a un proceso de reversión a la media.

Luego, Ljung-Box obtiene un estadístico  $Q = 0,05$ , y un valor  $p = 0,83$ . Este test evalúa la hipótesis nula de que no existe autocorrelación entre los residuos del modelo. Con  $p > 0,05$ , se acepta la  $H_0$ , y se descarta autocorrelación en los residuos, indicando que el modelo captura la estructura temporal de la serie.

De igual manera se acepta la  $H_0$  del test Jarque-Bera, indicando que la distribución de los residuos puede considerarse normal.

Sin embargo, en lo que a homocedasticidad respecta, la  $H_0$  plantea que la varianza de los residuos es constante, y se rechaza con  $p = 0,01 < 0,05$ .

### 2.2.2. ARIMA(2,0,0) - Serie Rendimientos

Se probó un modelo con mayor componente autorregresiva. El resultado ha sido un modelo donde el único término significativo es  $ar.L1$ , indicando que el modelo ARIMA(1,0,0) proporciona un mejor ajuste.

Se evidencian valores más altos de criterios AIC y BIC, así como ligeramente mejores estadísticos Q y JB. Se descarta la implementación de este modelo.



## ARIMA(2,0,0)

SARIMAX Results	
<b>Dep. Variable:</b>	Diff - Log - Rendimiento (Kg/Ha)
<b>No. Observations:</b>	34
<b>Sample:</b>	11-01-1981 - 11-01-2014
<b>Model:</b>	ARIMA(2, 0, 0)
<b>Log Likelihood:</b>	18.181
<b>AIC:</b>	-28.361
<b>BIC:</b>	-22.256
<b>HQIC:</b>	-26.279

Cuadro 5: Resumen del modelo ARIMA(2,0,0)

	coef	std err	z	P>  z	[0.025]	[0.975]
const	0.0296	0.019	1.553	0.121	-0.008	0.067
ar.L1	-0.5056	0.194	-2.608	0.009	-0.886	-0.126
ar.L2	0.0278	0.219	0.127	0.899	-0.401	0.457
sigma2	0.0199	0.005	3.904	0.000	0.010	0.030

Cuadro 6: Coeficientes del modelo

<b>Ljung-Box (L1) (Q):</b>	0.00
<b>Prob(Q):</b>	0.99
<b>Jarque-Bera (JB):</b>	2.01
<b>Prob(JB):</b>	0.37
<b>Heteroskedasticity (H):</b>	0.16
<b>Prob(H) (two-sided):</b>	0.01
<b>Skew:</b>	0.16
<b>Kurtosis:</b>	4.14

Cuadro 7: Pruebas Diagnósticas del Modelo ARIMA(2,0,0)

### 2.2.3. Auto ARIMA - Serie proxy Temperatura

Como se ha observado en la Fig.4 de ACF y PACF correspondientes, no se aprecian lags significativos indicando componentes de medias móviles ni autorregresivos. A priori la variable proxy temperatura pareciera comportarse como un ruido blanco, indicando que no existe correlación significativa

entre observaciones consecutivas ni más distantes.

A continuación se presentan los resultados de la función auto arima:

### ARIMA(0,0,0)

SARIMAX Results	
<b>Dep. Variable:</b>	<i>ProxyTemperatura</i>
<b>No. Observations:</b>	34
<b>Sample:</b>	11-01-1981 - 11-01-2014
<b>Model:</b>	SARIMAX
<b>Log Likelihood:</b>	16.516
<b>AIC:</b>	-29.032
<b>BIC:</b>	-25.980
<b>HQIC:</b>	-27.991

Cuadro 8: Resumen del Modelo ARIMA(0,0,0)

	<b>coef</b>	<b>std err</b>	<b>z</b>	<b>P&gt;  z </b>	<b>[0.025]</b>	<b>[0.975]</b>
intercept	2.1650	0.028	78.212	0.000	2.111	2.219
sigma2	0.0222	0.005	4.367	0.000	0.012	0.032

Cuadro 9: Coeficientes del Modelo ARIMA(0,0,0)(0,0,0)[0] con Intercepto

<b>Ljung-Box (L1) (Q):</b>	0.64
<b>Prob(Q):</b>	0.42
<b>Jarque-Bera (JB):</b>	2.81
<b>Prob(JB):</b>	0.25
<b>Heteroskedasticity (H):</b>	1.08
<b>Prob(H) (two-sided):</b>	0.90

Cuadro 10: Pruebas Diagnósticas del Modelo ARIMA(0,0,0)

El modelo ARIMA(0,0,0) con intercepto sugiere que la serie de la Proxy de temperatura no tiene dependencia temporal significativa de sus valores pasados o errores anteriores. En su lugar, la serie fluctúa aleatoriamente alrededor de su media (*intercept* = 2,1650), lo que es consistente con un proceso de ruido blanco. Las pruebas diagnósticas respaldan la adecuación del modelo,

ya que no se detectan autocorrelación, no normalidad ni heteroscedasticidad en los residuos.

#### 2.2.4. Auto ARIMA - Serie proxy Precipitaciones

De la inspección de la Fig.5 se ha propuesto un modelo ARIMA(2,0,0) ya que presenta los primeros 2 lags significativos en la función de autocorrelaciones parciales.

A continuación se presenta el modelo obtenido con auto arima:

**ARIMA(0,0,1)**

SARIMAX Results	
<b>Dep. Variable:</b>	<i>y</i>
<b>No. Observations:</b>	34
<b>Sample:</b>	11-01-1981 - 11-01-2014
<b>Model:</b>	SARIMAX(0, 0, 1)
<b>Log Likelihood:</b>	109.438
<b>AIC:</b>	-214.877
<b>BIC:</b>	-211.824
<b>HQIC:</b>	-213.836

Cuadro 11: Resumen del Modelo ARIMA(0,0,1)

	<b>coef</b>	<b>std err</b>	<b>z</b>	<b>P &gt;  z </b>	<b>[0.025]</b>	<b>[0.975]</b>
ma.L1	-0.9526	0.092	-10.361	0.000	-1.133	-0.772
sigma2	8.738e-05	2.92e-05	2.994	0.003	3.02e-05	0.000

Cuadro 12: Coeficientes del Modelo ARIMA(0,0,1)

<b>Ljung-Box (L1) (Q):</b>	2.23
<b>Prob(Q):</b>	0.13
<b>Jarque-Bera (JB):</b>	1.25
<b>Prob(JB):</b>	0.54
<b>Heteroskedasticity (H):</b>	1.93
<b>Prob(H) (two-sided):</b>	0.29

Cuadro 13: Pruebas Diagnósticas del Modelo ARIMA(0,0,1)

El modelo ARIMA(0,0,1) sugiere que la serie Proxy de precipitaciones tiene una dependencia temporal significativa a través de los errores pasados, capturada por el término de media móvil (*ma.L1*). Este término indica que los errores anteriores influyen en el valor actual de la serie, lo que puede interpretarse como un proceso de ajuste basado en la historia de los residuos. Las pruebas diagnósticas respaldan la adecuación del modelo, ya que no se detectan autocorrelación, no normalidad ni heteroscedasticidad en los residuos. Sin embargo, el estadístico  $Q = 0,13$  es incómodamente bajo, y si bien no se rechaza  $H_0$  - introduce la sospecha de que el ajuste no sea óptimo.

### 2.2.5. ARIMA(2,0,0) - Serie proxy Precipitaciones

Se decide ajustar un nuevo modelo basado en el análisis de ACF y PACF con 2 rezagos significantes para la componente autorregresiva.

#### ARIMA(2,0,0)

SARIMAX Results	
<b>Dep. Variable:</b>	Diff - BC - weighted_sum_precipitation
<b>No. Observations:</b>	34
<b>Sample:</b>	11-01-1981 - 11-01-2014
<b>Model:</b>	SARIMAX(2, 0, 0)
<b>Log Likelihood:</b>	110.974
<b>AIC:</b>	-213.948
<b>BIC:</b>	-207.842
<b>HQIC:</b>	-211.866

Cuadro 14: Resumen de Resultados del Modelo ARIMA(2,0,0)

	coef	std err	z	P>  z	[0.025]	[0.975]
const	0.0004	0.001	0.635	0.526	-0.001	0.002
ar.L1	-0.9315	0.145	-6.432	0.000	-1.215	-0.648
ar.L2	-0.6031	0.135	-4.478	0.000	-0.867	-0.339
sigma2	8.224e-05	2.74e-05	2.999	0.003	2.85e-05	0.000

Cuadro 15: Coeficientes del Modelo ARIMA(2,0,0)

<b>Ljung-Box (L1) (Q):</b>	0.00
<b>Prob(Q):</b>	0.96
<b>Jarque-Bera (JB):</b>	1.25
<b>Prob(JB):</b>	0.54
<b>Heteroskedasticity (H):</b>	1.01
<b>Prob(H) (two-sided):</b>	0.99
<b>Skew:</b>	0.03
<b>Kurtosis:</b>	2.06

Cuadro 16: Pruebas Diagnósticas del Modelo ARIMA(2,0,0)

El modelo ARIMA(2,0,0) se puede expresar como:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$$

donde:

- $y_t$  es el valor de la serie en el tiempo  $t$ ,
- $\phi_1 = -0,9315$  es el coeficiente autorregresivo de primer orden,
- $\phi_2 = -0,6031$  es el coeficiente autorregresivo de segundo orden,
- $\epsilon_t$  es el término de error en el tiempo  $t$ , que sigue una distribución normal con media cero y varianza  $\sigma^2 = 8,224 \times 10^{-5}$ .

Sustituyendo los valores estimados, la ecuación queda:

$$y_t = -0,9315 y_{t-1} - 0,6031 y_{t-2} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 8,224 \times 10^{-5})$$

Los coeficientes autorregresivos son altamente significativos con *pvalue*  $\ll 0,05$ , y contribuyen negativamente sobre el valor actual. El término constante, resulta no significativo con *pvalue*  $> 0,05$  implicando que la serie oscila en torno a una media igual a cero.

Ljung-Box obtiene un estadístico  $Q = 0,00$ , y un valor  $p = 0,96$ . Este test evalúa la hipótesis nula de que no existe autocorrelación entre los residuos del modelo. Con  $p > 0,05$ , se acepta la  $H_0$ , y se descarta autocorrelación en los residuos, indicando que el modelo captura la estructura temporal de la serie.

De igual manera se aceptan las  $H_0$  del test Jarque-Bera y Homocedasticidad, indicando que la distribución de los residuos puede considerarse normal y con varianza constante.

### 2.2.6. ARIMA-X - Rendimientos

El último modelo utilizado para realizar estimaciones sobre la serie de Rendimientos fue un ARIMA con variables exógenas: Proxy Precipitaciones y Proxy temperatura.

#### ARIMA-X (3,0,2)

SARIMAX Results	
<b>Dep. Variable:</b>	Diff - Log - Rendimiento (Kg/Ha)
<b>No. Observations:</b>	34
<b>Sample:</b>	11-01-1981 - 11-01-2014
<b>Model:</b>	SARIMAX(3, 0, 2)
<b>Log Likelihood:</b>	26.379
<b>AIC:</b>	-34.758
<b>BIC:</b>	-21.021
<b>HQIC:</b>	-30.074

Cuadro 17: Resumen de Resultados del Modelo ARIMAX

	coef	std err	z	P>  z	[0.025]	[0.975]
intercept	-1.7165	0.983	-1.747	0.081	-3.642	0.209
Proxy-precipitacion	3.7350	1.752	2.132	0.033	0.302	7.168
Proxy-temp	0.2915	0.156	1.870	0.062	-0.014	0.597
ar.L1	-0.7505	0.344	-2.179	0.029	-1.425	-0.076
ar.L2	-0.4675	0.372	-1.258	0.208	-1.196	0.261
ar.L3	-0.6223	0.150	-4.157	0.000	-0.916	-0.329
ma.L1	0.4011	0.402	0.997	0.319	-0.388	1.190
ma.L2	0.1775	0.316	0.562	0.574	-0.441	0.796
sigma2	0.0118	0.004	2.802	0.005	0.004	0.020

Cuadro 18: Coeficientes del Modelo ARIMA-X (3,0,2)

El modelo ARIMA-X (3,0,2) sugiere que la serie de variaciones porcentuales de rendimientos está influenciada tanto por su propia estructura temporal como por las variables exógenas incluidas. En particular, por el proxy de precipitaciones que tiene un efecto positivo y estadísticamente significativo ( $p = 0,033$ ) sobre el rendimiento, indicando que un aumento en las precipitaciones se asocia con un incremento en la variación porcentual del rendimiento.

<b>Ljung-Box (L1) (Q):</b>	0.31
<b>Prob(Q):</b>	0.58
<b>Jarque-Bera (JB):</b>	0.57
<b>Prob(JB):</b>	0.75
<b>Heteroskedasticity (H):</b>	0.23
<b>Prob(H) (two-sided):</b>	0.02
<b>Skew:</b>	0.25
<b>Kurtosis:</b>	2.61

Cuadro 19: Pruebas Diagnósticas del Modelo ARIMA-X (3,0,2)

Por otro lado, la relación ponderada de la temperatura promedio muestra un efecto positivo marginalmente significativo ( $p = 0,062$ ), sugiriendo que la temperatura también podría influir en el rendimiento, aunque con menor certeza estadística.

En cuanto a la estructura temporal del modelo, los términos autorregresivos AR(1) ( $p = 0,029$ ) y AR(3) ( $p < 0,001$ ) son significativos y negativos, lo que indica un proceso de reversión a la media en la serie. Sin embargo, los términos de medias móviles (MA(1) y MA(2)) no son estadísticamente significativos ( $p > 0,05$ ), lo que sugiere que los errores pasados no contribuyen de manera significativa al valor actual de la serie.

Las pruebas diagnósticas respaldan la adecuación del modelo en términos de autocorrelación y normalidad de los residuos (Ljung-Box  $p = 0,58$ , Jarque-Bera  $p = 0,75$ ). Sin embargo, se detecta heteroscedasticidad en los residuos ( $p = 0,02$ ), lo que indica que la varianza de los errores no es constante a lo largo del tiempo.

### 3. Análisis de Resultados

#### 3.1. Selección de modelos

Para cada serie se ha elegido un mejor modelo de los anteriormente expuestos, basados en el cumplimiento de supuestos estadísticos (residuos sin autocorrelación, homocedásticos y normalmente distribuidos) así como también criterios de información AIC y BIC.

##### Rendimientos

Métrica	ARIMA(1,0,0)	ARIMA(2,0,0)	SARIMAX(3,0,2)
AIC	-29.160	-28.361	-34.758
BIC	-26.108	-22.256	-21.021
Ljung-Box (Q)	0.05	0.00	0.31
Prob(Q)	0.83	0.99	0.58
Jarque-Bera (JB)	2.47	2.01	0.57
Prob(JB)	0.29	0.37	0.75
Heteroscedasticidad (H)	0.19	0.16	0.23
Prob(H)	0.01	0.01	0.02

Cuadro 20: Comparación de modelos - Serie de rendimientos.

##### Precipitaciones

Métrica	ARIMA(0,0,1)	ARIMA(2,0,0)
AIC	-214.877	-213.948
BIC	-211.824	-207.842
Ljung-Box (Q)	2.23	0.00
Prob(Q)	0.13	0.96
Jarque-Bera (JB)	1.25	1.25
Prob(JB)	0.54	0.54
Heteroscedasticidad (H)	1.93	1.01
Prob(H)	0.29	0.99

Cuadro 21: Comparación de modelos - Serie Proxy precipitaciones.



### Modelos elegidos:

- Serie de rendimientos: modelo elegido ARIMA(1,0,0) - modelo más simple y mejor BIC.
- Serie proxy temperatura: modelo elegido ARIMA(0,0,0) - modelo de constante.
- Serie proxy precipitaciones: modelo elegido ARIMA(2,0,0) - BIC y AIC muy bajos, excelente resultados en diagnósticos de residuos.

### 3.2. Predicciones

Dado que la frecuencia de las series es anual, el horizonte de predicción será de 1 período hacia delante. Se evaluaron las series de rendimientos y precipitaciones utilizando el enfoque de validación cruzada *Walk-Forward*<sup>5</sup>. Se han entrenado los modelos con datos hasta la campaña de 2014 y se testeó la predicción de 2015, luego se re-entreno incluyendo los datos hasta 2015, y se testeó la campaña de 2016. Así sucesivamente hasta la predicción de la campaña 2022.

Las predicciones se evaluaron utilizando:

- Mean Absoluto Error(MAE): Mide el promedio de los errores absolutos entre los valores reales y los predichos. Está en la misma unidad que la variable original.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- Root Mean Squared Error (RMSE): Raíz cuadrada del mean squared error (MSE). Es menos sensible a los outliers. Está en la misma unidad que la variable original.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

---

<sup>5</sup>La validación cruzada walk-forward es un método en el que se entrena el modelo con datos históricos para predecir el siguiente período, luego se incorpora ese período al conjunto de entrenamiento y se repite el proceso. Esto respeta el orden temporal y proporciona una evaluación realista del desempeño del modelo.

### Serie de Rendimientos

A continuación se presentan los resultados del entrenamiento CV Walk-Forward, incluyendo tabla de predicciones y errores, como el gráfico de valores predichos y valores reales.

Train / Pred	MAE	RMSE	y_test	y_pred
2014-2015	0.125379	0.125379	0.056440	-0.068939
2015-2016	0.008431	0.008431	-0.034396	-0.025965
2016-2017	0.351190	0.351190	-0.335338	0.015852
2017-2018	0.186378	0.186378	0.336440	0.150061
2018-2019	0.062354	0.062354	-0.106604	-0.168959
2019-2020	0.147201	0.147201	-0.095417	0.051785
2020-2021	0.017327	0.017327	0.027841	0.045169
2021-2022	0.338002	0.338002	-0.351153	-0.013151

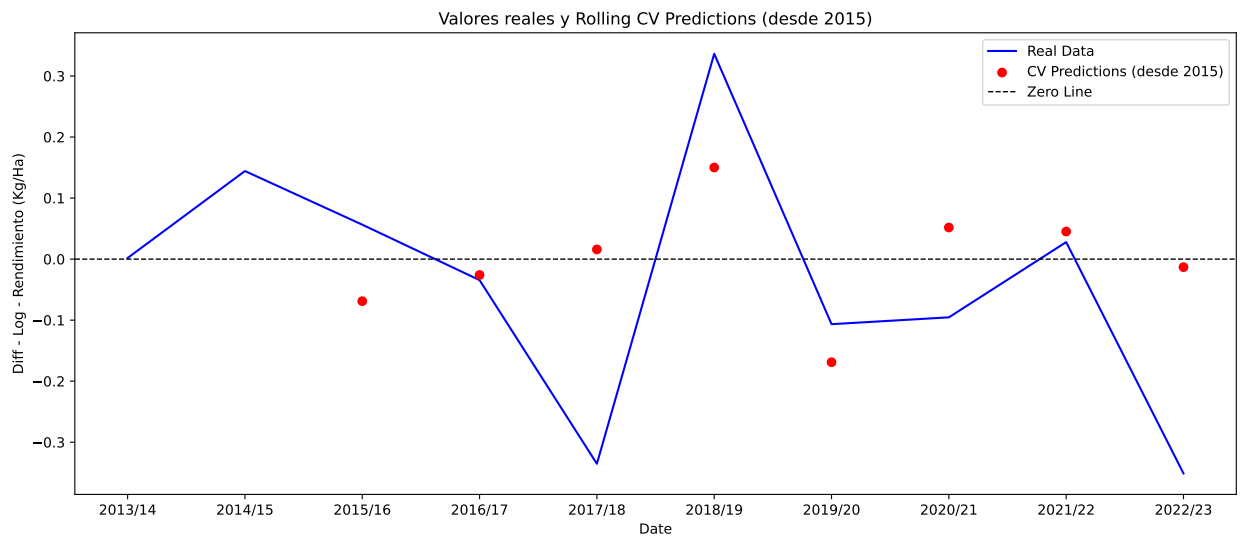


Figura 6: Real vs Fitted values de Diff-log-rendimientos CV.

Al evaluar los resultados obtenidos de la validación cruzada utilizando las métricas expuestas anteriormente, se obtiene:

- MAE: 0.1545.
- RMSE: 0.1981.

Recordando que la transformación de la serie puede interpretarse como la tasa de cambio porcentual de los rendimientos, estas métricas indican un rango de error entre 15 % y 20 % en las predicciones con horizonte de un periodo. Si llevásemos estos resultados a la escala original, la media de los errores se ubicará en torno al 16 % y el error de pronóstico (más sensible a outliers), estará cercano al 21 %. Cabe destacar sin embargo, que las 2 campañas donde el modelo arroja peores resultados coinciden con 2 sequías muy fuertes: 2017/2018 y 2022/2023. Es decir, podría decirse que el modelo generaliza bien a pesar de fallar en la identificación de eventos anómalos.

### **Serie proxy Precipitaciones**

Se presentan los resultados del entrenamiento CV Walk-Forward, incluyendo tabla de predicciones y errores, como el gráfico de valores predichos y valores reales.

<b>Train / Pred</b>	<b>MAE</b>	<b>RMSE</b>	<b>y_test</b>	<b>y_pred</b>
2014-2015	0.000063	0.000063	0.002308	0.002245
2015-2016	0.007346	0.007346	-0.000700	-0.008046
2016-2017	0.013522	0.013522	-0.014215	-0.000693
2017-2018	0.000816	0.000816	0.013890	0.013074
2018-2019	0.003167	0.003167	-0.001354	-0.004521
2019-2020	0.002447	0.002447	-0.009037	-0.006590
2020-2021	0.012319	0.012319	-0.003457	0.008862
2021-2022	0.015722	0.015722	-0.007725	0.007997

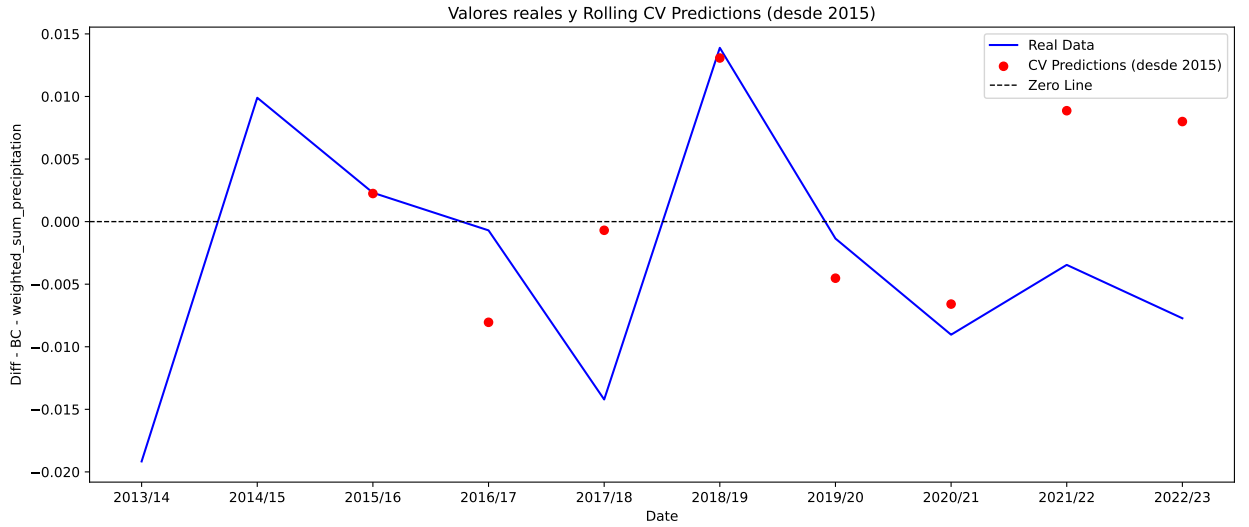


Figura 7: Real vs Fitted values de Proxy precipitaciones CV.

Al evaluar los resultados obtenidos de la validación cruzada utilizando las métricas expuestas anteriormente, se obtiene:

- MAE: 0.007.
- RMSE: 0.009.

Observar que los errores parecen muy pequeños debido a que están expresados en la misma escala que la serie transformada. Sin embargo, la Fig.7 permite observar claramente el desajuste entre la serie y los pronósticos.

## 4. Modelado VAR y Análisis de Causalidad

Un modelo VAR (Vector Autoregresivo) es un modelo estadístico multi ecuacional utilizado para analizar y pronosticar datos de series temporales. Es un tipo de modelo de regresión que extiende el modelo AR (Autoregresivo) para incluir múltiples variables y rezagos. En un modelo VAR, los valores de una serie temporal se predicen en función de los valores de otras series temporales, donde la predicción es una función de los valores pasados de todas las series temporales. El modelo se representa típicamente en forma matricial, donde la ecuación predictiva para cada variable es una combinación de los valores pasados de todas las variables (Nielsen (2019)).

Una condición necesaria para la construcción de un VAR es que las series involucradas tengan el mismo grado de integración y que sean estacionarias, por ello trabajamos con las series transformadas de rendimientos y precipitaciones. La serie de intensidad de temperaturas no presentó componentes autorregresivos significativos, por lo que no será incluida en este análisis.

Se realizó una exploración de la cantidad de lags a incluir, minimizando criterios de información. El ensayo se corrió explorando hasta 7 lags, a continuación los resultados:

Order	AIC	BIC	FPE	HQIC
0	-12.92	-12.83	2.458e-06	-12.89
1	-13.17	-12.90	1.917e-06	-13.07
2	<b>-13.50*</b>	<b>-13.05*</b>	<b>1.381e-06*</b>	<b>-13.34*</b>
3	-13.37	-12.74	1.581e-06	-13.16
4	-13.18	-12.38	1.930e-06	-12.91
5	-13.16	-12.17	2.021e-06	-12.82
6	-13.24	-12.08	1.924e-06	-12.85
7	-13.09	-11.75	2.342e-06	-12.64

Cuadro 22: Selección de orden del VAR (\* indica el mínimo)

La optimización según los 4 criterios sugiere incluir hasta lags de 2do orden. Esto coincide con el valor p obtenido en el modelo elegido para la serie proxy de precipitaciones.

El modelo propuesto entonces es el siguiente:

$$y_{1t} = c_1 + \phi_{11}^{(1)} y_{1,t-1} + \phi_{12}^{(1)} y_{2,t-1} + \phi_{11}^{(2)} y_{1,t-2} + \phi_{12}^{(2)} y_{2,t-2} + \epsilon_{1t},$$

$$y_{2t} = c_2 + \phi_{21}^{(1)} y_{1,t-1} + \phi_{22}^{(1)} y_{2,t-1} + \phi_{21}^{(2)} y_{1,t-2} + \phi_{22}^{(2)} y_{2,t-2} + \epsilon_{2t},$$

donde:

- $y_{1t}$ : **Diff - Log - Rendimiento (Kg/Ha)** en el tiempo  $t$ .
- $y_{2t}$ : **Diff - BC - weighted\_sum\_precipitation** en el tiempo  $t$ .
- $c_1, c_2$ : Constantes (términos independientes) para cada ecuación.
- $\phi_{ij}^{(k)}$ : Coeficientes para el  $k$ -ésimo rezago de la variable  $j$  en la ecuación de la variable  $i$ .

- $\epsilon_{1t}, \epsilon_{2t}$ : Términos de error (ruido blanco) para cada ecuación.

### Estabilidad del modelo

Una vez definido el modelo, se procede a verificar la estabilidad. Esta refiere a la capacidad del sistema de volver al equilibrio tras una perturbación en alguna de las variables. Es decir, si el modelo es estable, las perturbaciones en una variable generan efectos transitorios que se desvanecen con el tiempo, permitiendo que las series regresen a su media. Por el contrario, si el modelo es inestable, una perturbación puede causar un comportamiento explosivo en las variables, haciendo que diverjan de su media y crezcan sin límite. En este caso, el modelo no es útil para análisis o pronósticos.

La estabilidad se verifica a través de los autovalores de la matriz  $F$  asociada al VAR:

$$F = \begin{bmatrix} \Phi_1 & \Phi_2 \\ I_2 & 0 \end{bmatrix} = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{11}^{(2)} & \phi_{12}^{(2)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{21}^{(2)} & \phi_{22}^{(2)} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

Donde:

- $\phi_{ij}^{(k)}$ : Representa el coeficiente del  $k$ -ésimo rezago de la variable  $j$  en la ecuación de la variable  $i$ .
- $I_2$ : Es la matriz identidad de tamaño  $2 \times 2$ , que asegura la transferencia de los rezagos.
- $0$ : Es una matriz de ceros de tamaño  $2 \times 2$ , que completa la estructura de la matriz de compañía.

Tras calcular los autovalores de dicha Matriz  $F$ , se obtuvo que los 4  $\lambda > 1$ . Es decir, que el sistema no es estable. Se repitió el proceso para VARs con  $lags = 5, 3, 1$  y el modelo no cambió su condición. Por este motivo no se siguió adelante con el VAR.

## 4.1. Cointegración

Dado que el modelo VAR no resultó adecuado para hallar una relación de corto plazo entre las variables, se realizó un chequeo de cointegración entre la serie de Rendimientos con transformación logarítmica y la serie proxy de precipitaciones con transformación de Box Cox. Ambas sin diferenciar, es

decir, no siendo estacionarias.

La cointegración testea si existe una combinación lineal de 2 o más series de tiempo no estacionarias, qué al combinarse, la serie resultante es estacionaria. Permite determinar si existe un equilibrio compartido de largo plazo, a pesar de desviaciones de corto plazo.

Entonces, dos series no estacionarias  $y_t$  y  $x_t$  ( $I(1)$ ) están **cointegradas** si existe una combinación lineal estacionaria ( $I(0)$ ):

$$u_t = y_t - \beta_1 x_t - \beta_0,$$

donde  $\beta_1$  y  $\beta_0$  definen la relación de equilibrio a largo plazo y  $\mu_t$  es el error estacionario (representa desviaciones del equilibrio).

Para evaluar dicha relación se llevo a cabo la prueba de Johansen. La hipótesis nula plantea que hay como máximo  $n$  relaciones de cointegración en el sistema, y se testeó con  $n = 0$  y  $n = 1$ . Los resultados arrojaron:

- Primer estadístico: 25.49 vs 15.49. Se rechaza  $H_0$ , existe al menos una relación de cointegración.
- Segundo estadístico: 5.08 vs 3.84. Se rechaza  $H_0$ , existe más de 1 relación de cointegración.

## 4.2. VECM

Dado que se detecta cointegración entre la variable de *Log-Rendimientos* y *BC-precipitaciones*, se procede a utilizar un modelo de vectores con corrección de error (VECM).

Para estas dos variables, el modelo tiene la siguiente estructura:

$$\begin{aligned}\Delta y_t &= \alpha_y \underbrace{(y_{t-1} - \beta_1 x_{t-1} - \beta_0)}_{\text{Desviación del equilibrio}} + \sum_{i=1}^{p-1} \phi_{yy}^{(i)} \Delta y_{t-i} + \sum_{i=1}^{p-1} \phi_{yx}^{(i)} \Delta x_{t-i} + \epsilon_{1t}, \\ \Delta x_t &= \alpha_x \underbrace{(y_{t-1} - \beta_1 x_{t-1} - \beta_0)}_{\text{Desviación del equilibrio}} + \sum_{i=1}^{p-1} \phi_{xy}^{(i)} \Delta y_{t-i} + \sum_{i=1}^{p-1} \phi_{xx}^{(i)} \Delta x_{t-i} + \epsilon_{2t},\end{aligned}$$

### Explicación de Términos

- $\alpha_y, \alpha_x$ : Velocidades de ajuste ( $\alpha_y < 0$  indica corrección activa).

- $\beta_1, \beta_0$ : Coeficientes de cointegración (relación de equilibrio).
- $\phi_{yy}^{(i)}, \phi_{yx}^{(i)}$ : Efectos de rezagos de  $\Delta y_t$  y  $\Delta x_t$  en  $\Delta y_t$ .
- $\phi_{xy}^{(i)}, \phi_{xx}^{(i)}$ : Efectos de rezagos de  $\Delta y_t$  y  $\Delta x_t$  en  $\Delta x_t$ .
- $\epsilon_{1t}, \epsilon_{2t}$ : Términos de error (ruido blanco).



## Resultados obtenidos:

Equation	Variable	coef	$P >  z $
Log - Rendimiento (Kg/Ha)	L1.Log - Rendimiento (Kg/Ha)	-0.4869	0.001
	L1.BC - weighted_sum_precipitation	0.1338	0.943
BC - weighted_sum_precipitation	L1.Log - Rendimiento (Kg/Ha)	-0.0179	0.111
	L1.BC - weighted_sum_precipitation	-0.4461	0.003

Cuadro 23: Parámetros fuera de la relación de cointegración y variables endógenas rezagadas

Variable	$\alpha$	std err	$P >  z $
ec1	-0.1546	0.052	0.003

Cuadro 24: Coeficientes de ajuste (alpha) - Log - Rendimiento (Kg/Ha)

Variable	$\alpha$	std err	$P >  z $
ec1	-0.0025	0.004	0.542

Cuadro 25: Coeficientes de ajuste (alpha) - BC - weighted\_sum\_precipitation

Variable	coef	std err	$P >  z $
beta.1	1.0000	0.000	0.000
beta.2	-6.1448	0.068	0.000

Cuadro 26: Relaciones de cointegración

## Análisis de resultados

### 1. Dinámica de Corto Plazo (Variables Endógenas Rezagadas)

- • Ecuación de Log - Rendimiento (Kg/Ha):

- El coeficiente para `L1.BC - weighted_sum_precipitation` es **0.1338** con un p-valor de **0.943**, lo que indica que no es estadísticamente significativo. Esto sugiere que la precipitación rezagada no tiene un efecto inmediato significativo sobre el rendimiento.
- El coeficiente para `L1.Log - Rendimiento (Kg/Ha)` es **0.4869** con un p-valor de **0.001**, lo que indica que los valores pasados del rendimiento predicen significativamente los valores futuros. Esto refleja una reversión a la media en el corto plazo.
- **Ecuación de BC - weighted\_sum\_precipitation:**
  - El coeficiente para `L1.Log - Rendimiento (Kg/Ha)` es **0.0179** con un p-valor de **0.111**, lo que indica que no es estadísticamente significativo. Esto sugiere que el rendimiento rezagado no tiene un efecto inmediato significativo sobre la precipitación.
  - El coeficiente para `L1.BC - weighted_sum_precipitation` es **-0.4461** con un p-valor de **0.003**, lo que indica que los valores pasados de la precipitación predicen significativamente los valores futuros. Esto también refleja una reversión a la media en el corto plazo.

**Interpretación:** En el corto plazo, no hay un efecto significativo inmediato de una variable sobre la otra. Cada variable parece estar impulsada principalmente por sus propios valores pasados, lo que sugiere una dinámica independiente en el corto plazo.

## 2. Equilibrio de Largo Plazo (Relación de Cointegración)

- **Coeficiente de Cointegración:**
  - El coeficiente de cointegración para `BC - weighted_sum_precipitation` ( $\beta_2$ ) es **-6.1448** con un p-valor de **0.000**, lo que indica que la relación de largo plazo entre el rendimiento y la precipitación es estadísticamente significativa.
- **Términos de Ajuste (Alpha):**
  - El coeficiente de ajuste ( $\alpha$ ) para `Log - Rendimiento (Kg/Ha)` es **-0.1546** con un p-valor de **0.003**, lo que indica que el rendimiento se ajusta activamente para restaurar el equilibrio de largo plazo después de una perturbación.

- El coeficiente de ajuste ( $\alpha$ ) para BC - `weighted_sum_precipitation` es **-0.0025** con un p-valor de **0.542**, lo que indica que la precipitación no se ajusta significativamente en respuesta a desviaciones del equilibrio.

**Interpretación:** Existe una relación de largo plazo significativa entre el rendimiento y la precipitación. El rendimiento depende de la precipitación en el largo plazo y se ajusta para corregir desviaciones del equilibrio. Sin embargo, la precipitación no responde a desviaciones del equilibrio, lo que sugiere que es exógena en el sistema.

## 5. Conclusiones

El análisis de las series temporales de rendimientos de soja y precipitaciones en la provincia de Buenos Aires evidencia que, en el **corto plazo**, ambas variables siguen un patrón de reversión a la media, lo que indica que sus fluctuaciones se rigen principalmente por dinámicas internas y factores locales (como las prácticas de manejo y la variabilidad climática). En contraste, a **largo plazo** se detecta una relación de cointegración significativa: los rendimientos de soja se ajustan activamente para restablecer el valor de equilibrio ante desviaciones, subrayando el papel crucial de las precipitaciones, determinadas por factores externos y fenómenos climáticos como El Niño.

Si bien el modelo VECM logra capturar esta relación de equilibrio, los pronósticos obtenidos presentan un margen de error considerable en la serie de rendimientos (MAE 15 % y RMSE 20 %), lo que limita su aplicabilidad en la práctica. Además, la inestabilidad detectada en el modelo VAR y la heteroscedasticidad de los residuos en el ARIMA/X sugieren la necesidad de explorar enfoques alternativos—por ejemplo, modelos no lineales o técnicas que incorporen la dinámica de la volatilidad (como GARCH) o emplear modelos más complejos como redes neuronales — y considerar la inclusión de variables adicionales (tecnológicas, económicas o relacionadas con eventos climáticos extremos).

En este contexto, la serie proxy de temperatura se comporta como un ruido blanco, evidenciando la ausencia de patrones autoregresivos o correlaciones significativas. Su comportamiento sugiere que, en el corto plazo, la variable temperatura no aporta valor predictivo ni influye directamente en los rendimientos o en las precipitaciones.

En conjunto, estos enfoques no se confunden entre sí, sino que se complementan: mientras el análisis a corto plazo mediante ARIMA con validación cruzada nos proporciona una visión detallada de la precisión predictiva y la dinámica operativa inmediata de las series, el VECM revela la estructura subyacente y el mecanismo de ajuste que vincula rendimientos y precipitaciones a largo plazo. Si bien los modelos actuales ya ofrecen información valiosa, la incorporación de metodologías predictivas más complejas podrían potenciar significativamente la precisión. Esto permitiría desarrollar una herramienta robusta para formular estrategias de gestión hídrica y de adaptación en el sector agrícola, aprovechando tanto las fortalezas operativas a corto plazo como la capacidad de anticipar y corregir desviaciones estructurales a largo plazo.

## 6. Referencias Bibliográficas

### Referencias

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3):159–178.

Nielsen, A. (2019). *Practical Time Series Analysis*. O'Reilly Media, Inc.

### 6.1. Fuentes de datos

- Ministerio de Agricultura, Ganadería y Pesca de Argentina. *Monitor de Estimaciones Agrícolas*. Disponible en: <https://www.magyp.gob.ar/sitio/areas/estimaciones/monitor/>.
- Dirección Nacional de Agricultura. *Reportes de Estimaciones de rendimientos*. Disponible en: <https://datosestimaciones.magyp.gob.ar/reportes.php?reporte=Estimaciones>
- Servicio Meteorológico Nacional. *Registros climatológicos* <https://www.smn.gob.ar/estadisticas>

### Código utilizado

- Preparación de series:  
<https://colab.research.google.com/drive/1GInTY1qzu4QdPk0CIfRcx8X-vDFHwGcy>
- Análisis de modelos:  
[https://colab.research.google.com/drive/145yt3G8mDlt7pWK1svRV9QNCxG\\_RjeKc](https://colab.research.google.com/drive/145yt3G8mDlt7pWK1svRV9QNCxG_RjeKc)

## 7. Apéndice

Lista de departamentos estudiados - Provincia de Buenos Aires, Argentina.

- AZUL
- BAHÍA BLANCA
- BENITO JUÁREZ
- BOLÍVAR
- CORONEL PRINGLES
- CORONEL SUÁREZ
- DOLORES
- JUNÍN
- LA PLATA
- LAS FLORES
- MERLO
- MORÓN
- OLAVARRÍA
- PEHUAJO
- PUNTA INDIO B.A.
- SAN FERNANDO
- TANDIL
- TRENQUE LAUQUEN
- TRES ARROYOS