

Series Temporales: Trabajo Bloque 1

Frank Salomón Sulca Palomino

Mayo 2025

1. Resumen

Este informe analiza la serie temporal de las temperaturas en la ciudad de Málaga desde enero de 1960 hasta diciembre de 2023. Se examinan los patrones de temperatura, las tendencias y las variaciones estacionales. Además, se aplican métodos estadísticos para determinar la estacionariedad y realizar proyecciones futuras para el año 2024 utilizando modelos ARMA y SARIMA.

2. Análisis preliminar de la serie

2.1. Representación gráfica y análisis de varianza

Representa gráficamente la serie para analizar su comportamiento general. Examina si existe algún problema de heterogeneidad de la varianza y describe si la serie muestra desviaciones del comportamiento estacionario en la media.

Se cargó el archivo `.rdata` que contiene la base de datos. Se detectaron 19 valores faltantes entre 1983 y 2001, 10 de ellos consecutivos en 1990. Para completar estos datos, se utilizó la función `na_seadec()` del paquete **forecast**, respetando los patrones estacionales de la serie.

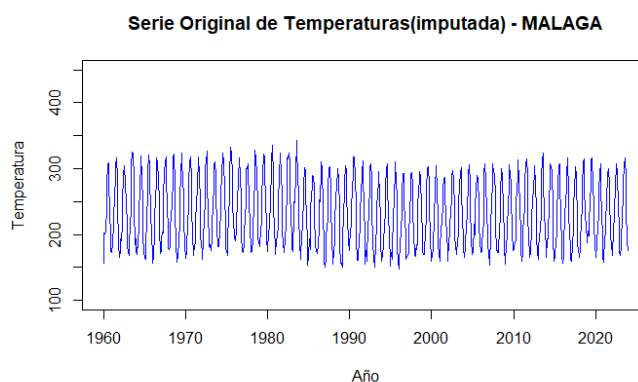
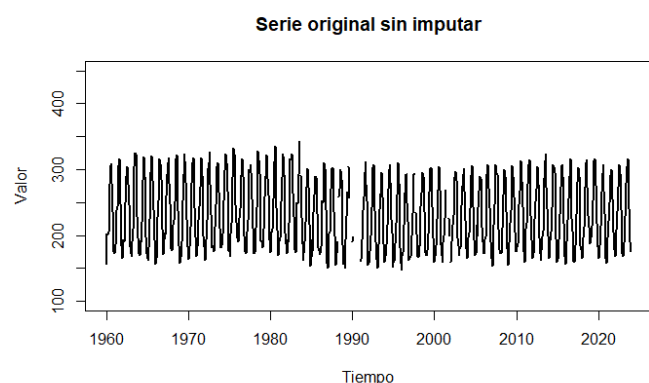
El gráfico inicial no mostró una tendencia clara. La prueba ADF arrojó un p-valor de **0.01**, lo que indica estacionariedad en la media. La prueba KPSS arrojó un p-valor de **0.1** en la serie original y también en la transformada logarítmicamente, lo que respalda la estacionariedad. Dado que la transformación logarítmica no mostró mejoras relevantes, no se utilizó.

Se ajustó un modelo de regresión lineal sobre la serie imputada. El análisis gráfico de los residuos no mostró patrones, lo que indica **homocedasticidad**. Por ello, no fue necesario aplicar diferenciación ($\text{lag} = 1$).

Dado que la serie logarítmica no aportó beneficios y los resultados de las pruebas fueron consistentes, se decidió continuar el análisis con la serie imputada, denominada `ts_malaga`.

A continuación se presenta el código R utilizado para crear e imputar la serie y los gráficos más importantes.

```
# Crear la serie original                                # Aplicar imputación estacional
ts_original <- ts(ciudad_df$MALAGA,                      ts_malaga <- na_seadec(ts_original)
                  start = c(1960, 1),
                  frequency = 12)
```



```
#Se aplica un modelo de regresion lineal a la serie como analisis exploratorio
modelo <- lm(ts_malaga ~ t)
residuos <- residuals(modelo)
```

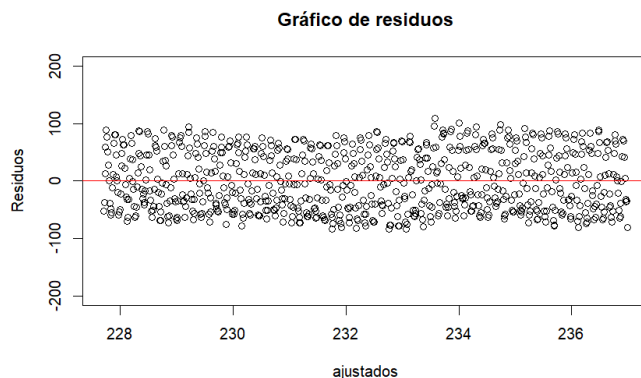


Figura 1: Gráfico de residuos del modelo de regresión lineal

2.2. Estimación de tendencia y componente estacional

Estima la tendencia y la componente estacional utilizando el método más adecuado. Indica el método de estimación empleado y los valores de los parámetros seleccionados para la estimación final. Menciona los valores que hayas ensayado, pero presenta únicamente los detalles y resultados del ajuste definitivo. Compara las estimaciones obtenidas con las proporcionadas por la función `stl`.

Para analizar la evolución de la temperatura a largo plazo, se estimó la tendencia utilizando la media móvil de 12 meses y el método STL. Ambos métodos mostraron una tendencia muy similar, con una media de **232.3**. La media móvil presentó un mínimo más bajo y un máximo ligeramente más alto, mientras que STL se adaptó mejor a los cambios. Cabe destacar que la media móvil pierde **11 datos**, mientras que STL no pierde ninguno.

Respecto a la estacionalidad, se utilizó la función `matplot()` para observar los patrones mensuales y se comparó con los resultados del análisis estacional obtenido con STL. Ambos métodos confirmaron una **estacionalidad fuerte**, con temperaturas más bajas en invierno (enero y diciembre) y más altas en verano (julio y agosto). El gráfico `matplot()` permitió visualizar estos patrones claramente, mientras que STL proporcionó una medida más precisa y detallada.

A continuación, se presentan los gráficos de la estimación de la tendencia y la estacionalidad, seguidos del último gráfico que muestra la descomposición STL.

```
# Suavizado con media móvil
suavizado_mm <- rollmean(ts_malaga, k = 12, fill = NA)
```

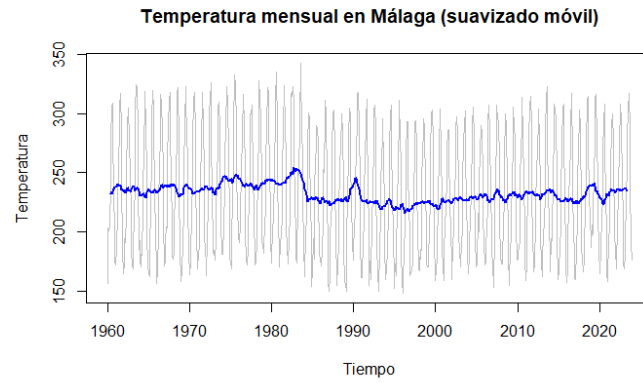


Figura 2: Estimación de la tendencia (suavizado móvil)

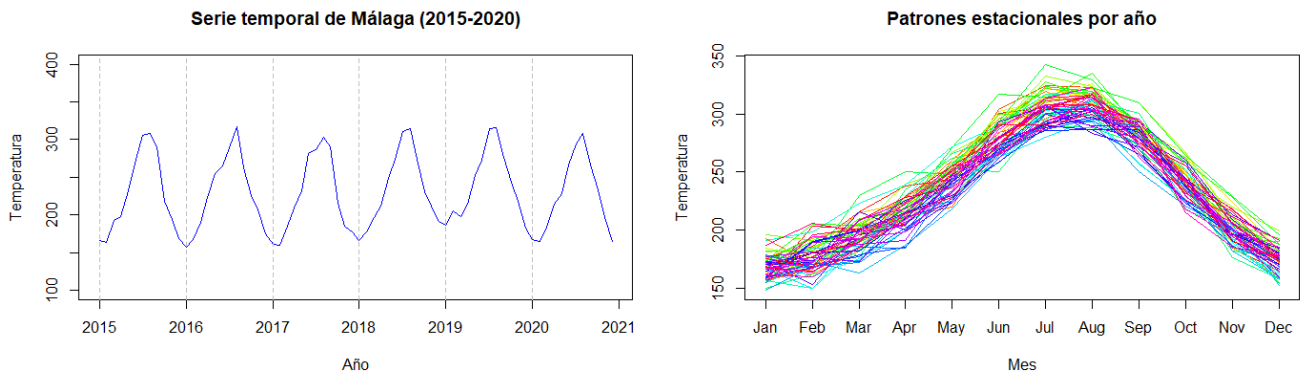


Figura 3: Componente estacional

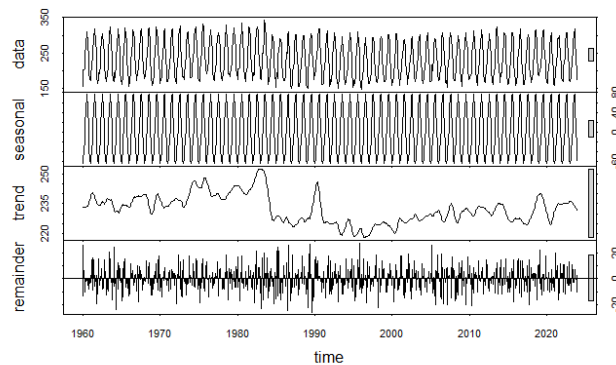


Figura 4: Descomposición STL: Tendencia, Estacionalidad y Residuos

2.3. Conversión a estacionariedad

Transforma la serie en una serie estacionaria diferenciando de forma adecuada para eliminar la componente estacional y la tendencia.

Se aplicó una diferenciación estacional a la serie con un retardo de $K = 12$ utilizando la función `diff(ts_málaga, lag`. Esta decisión se basó en los gráficos anteriores, que mostraban un componente estacional con periodicidad de 12 meses. Al comparar los gráficos ACF y PACF antes y después de la diferenciación, se observó visualmente un comportamiento más estable en la serie transformada, lo que motivó la aplicación de las pruebas ADF y KPSS.

La **prueba ADF**, en la serie original, arrojó un estadístico de **-6.7439** y un p-valor de **0.01**. Tras la diferenciación, el estadístico bajó a **-8.0537**, con un p-valor también de **0.01**, lo que refuerza la evidencia de estacionariedad.

En la **prueba KPSS**, la serie original presentó un estadístico de **0.18745** con un p-valor de **0.1**, mientras que la serie diferenciada tuvo un estadístico de **0.029264**, con p-valor también de **0.1**. Esta reducción del estadístico sugiere una mayor estabilidad de la serie diferenciada.

En resumen, aunque la serie original ya mostraba indicios de estacionariedad, la aplicación de la diferenciación estacional con **K = 12** permitió confirmar esta propiedad con mayor claridad. Se continuará el análisis con la serie transformada.

A continuación mostramos el resultado de las pruebas realizadas y los gráficos ACF y PACF antes y después de la diferenciación:

Prueba ADF para la serie original:

Augmented Dickey-Fuller Test

data: ts_malaga
Dickey-Fuller = -6.7439, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary

Prueba ADF para la serie diferenciada (Lag = 12):

Augmented Dickey-Fuller Test

data: dif12ts_malaga
Dickey-Fuller = -8.0537, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary

Figura 5: Prueba ADF a la serie original y diferenciada

Prueba KPSS para la serie original:

KPSS Test for Level Stationarity

data: ts_malaga
KPSS Level = 0.18745, Truncation lag parameter = 6, p-value = 0.1

Prueba KPSS para la serie diferenciada (Lag = 12):

KPSS Test for Level Stationarity

data: dif12ts_malaga
KPSS Level = 0.029264, Truncation lag parameter = 6, p-value = 0.1

Figura 6: Prueba KPSS a la serie original y diferenciada

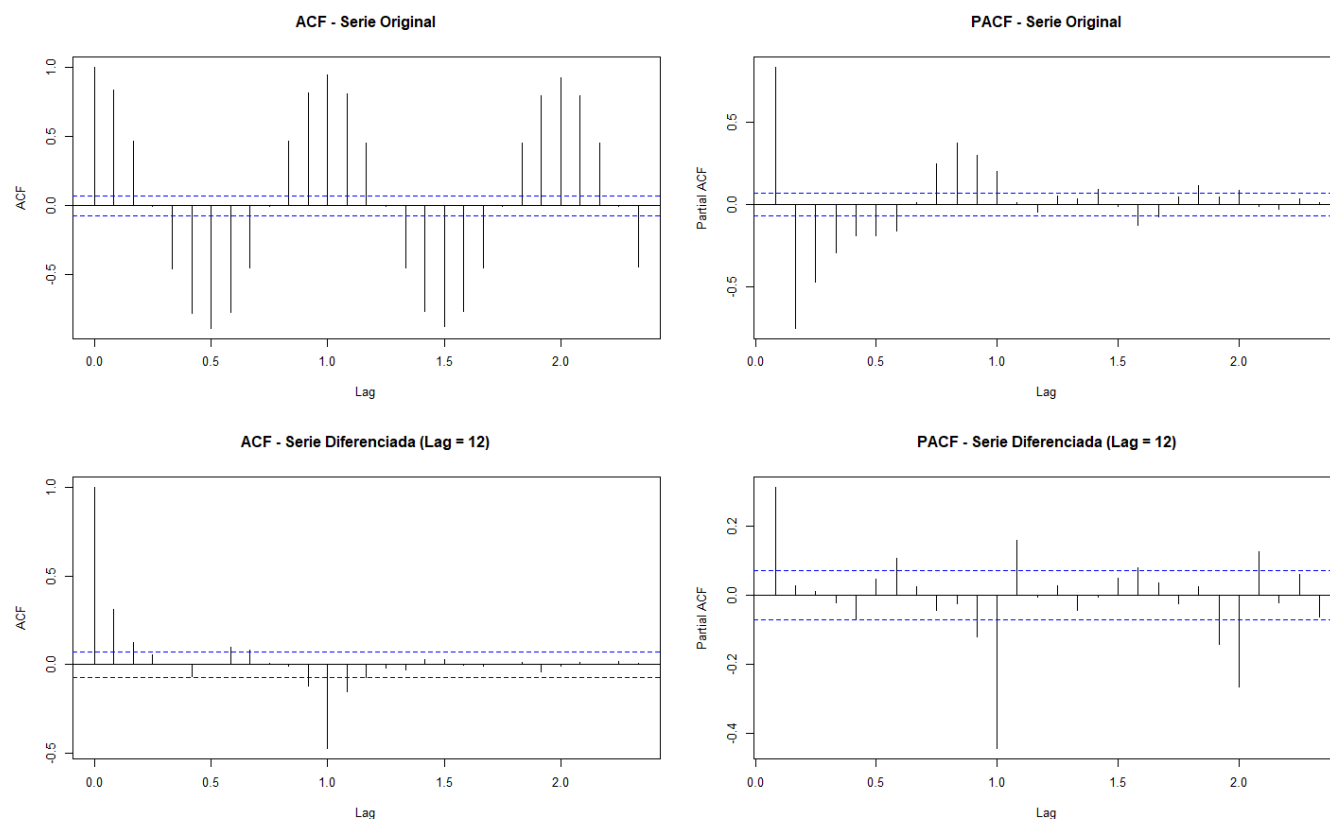


Figura 7: Gráficos ACF y PACF a la serie original y diferenciada

3. Modelado ARMA y SARIMA

3.1. Evaluación de modelo ARMA

Analiza si es útil plantear un modelo tipo ARMA para la serie.

Después de aplicar las pruebas ADF y KPSS a la serie original y a la diferenciada con rezago 12, se concluyó que la **serie diferenciada** es la más adecuada, ya que presenta **mejores indicadores de estacionariedad**.

Esta conclusión se respalda también con los gráficos de **ACF** y **PACF**: mientras que la serie original mostraba dependencia temporal, en la serie diferenciada la mayoría de las autocorrelaciones quedaron dentro del rango esperado, indicando que la tendencia y estacionalidad fueron eliminadas.

Por otro lado, los residuos de la serie original ya mostraban **homocedasticidad**, y esta característica se mantiene tras la diferenciación. Esto permite seguir adelante con la aplicación de un modelo **ARMA** con mayor confianza.

En resumen, la serie diferenciada cumple con los requisitos necesarios y está lista para ser modelada con un ARMA.

3.2. Selección y ajuste del modelo ARMA

Selecciona de forma razonada el modelo ARMA que consideres más adecuado (NO se puede utilizar como criterio la orden auto.arima) y presenta el modelo ajustado. Contrasta si los coeficientes del modelo son iguales a cero (cada uno de forma individual).

Con base en los gráficos ACF y PACF de la serie diferenciada, se observó que solo el **primer rezago** es significativamente diferente de cero, lo que sugiere que un modelo **ARMA(1,1)** podría ser apropiado. Sin embargo, al evaluar este modelo, el coeficiente del término MA no fue significativo (**p-valor = 0.42**), por lo que se descartó incluirlo.

Se compararon los AIC de ambos modelos: **ARMA(1,1)** obtuvo un valor de **6145.65**, mientras que **ARMA(1,0)** alcanzó **6144.28**, siendo este último más bajo. Esto respalda la elección del **modelo ARMA(1,0)**, al ofrecer un mejor ajuste con menor complejidad.

No se consideraron modelos con órdenes más altos, ya que no aportaban mejoras significativas en los coeficientes ni en los criterios de información, aumentando innecesariamente el riesgo de sobreajuste. Se optó así por un modelo más simple y robusto.

```
AIC para ARMA(1,1): 6145.649
Call:
arima(x = dif12ts_malaga, order = c(1, 0, 1))

Coefficients:
      ar1      ma1  intercept
    0.3964 -0.0926    0.0558
s.e.  0.1058   0.1149    0.7658

sigma^2 estimated as 196.5:  log likelihood = -3068.82,  aic = 6145.65
```

Figura 8: Resumen del modelo ARMA(1,1)

```
AIC para ARMA(1,0): 6147.648
Call:
arima(x = dif12ts_malaga, order = c(1, 0, 0))

Coefficients:
      ar1  intercept
    0.3129    0.0471
s.e.  0.0345    0.7418

sigma^2 estimated as 196.6:  log likelihood = -3069.14,  aic = 6144.28
```

Figura 9: Resumen del modelo ARMA(1,0)

3.3. Validación del modelo

Valida el modelo utilizando las herramientas adecuadas e indica cuál consideras la comprobación más importante.

Se evaluó el modelo ARMA(1,0) mediante pruebas estadísticas y análisis de residuos. La prueba de Shapiro-Wilk arrojó un p-valor de 0.6444, mayor a 0.05, por lo que no se rechaza la hipótesis nula, indicando que los residuos siguen una distribución normal.

Por otro lado, la prueba de Ljung-Box dio un p-valor menor a 2.2×10^{-16} , lo que indica autocorrelación significativa en los residuos. Esto se confirma en el gráfico ACF, donde se observa un pico en el rezago 12, señal de una componente estacional no capturada.

Dado este resultado, se considerará ajustar un modelo SARIMA, que permita incorporar componentes estacionales y corregir la autocorrelación detectada, mejorando así el ajuste del modelo a la serie.

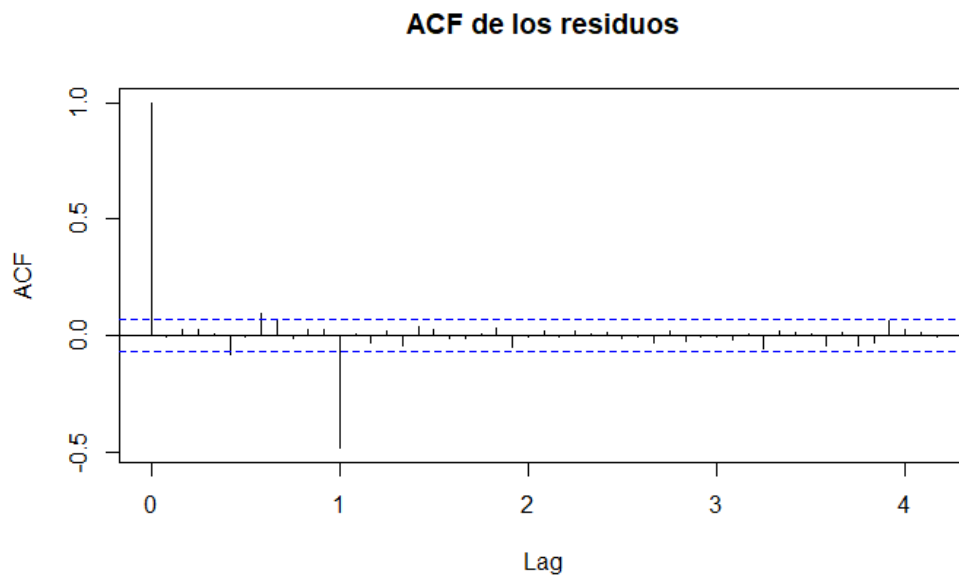


Figura 10: ACF de los residuos del modelo ARMA(1,0)

3.4. Ajuste del modelo SARIMA

Ajusta un modelo SARIMA a la serie original que unifique en un modelo global los resultados de los apartados anteriores. El modelo debe ser el mismo que has obtenido en el paso anterior, partiendo de la serie estacionaria.

Recapitulando, en el apartado anterior se optó por el modelo ARMA(1,0). Sin embargo, al analizar la autocorrelación de sus residuos, se detectó una estacionalidad en el rezago 12, lo que indica un patrón que se repite cada 12 periodos. Por esta razón, se decidió incorporar componentes estacionales mediante modelos SARIMA. Se compararon dos alternativas: SARIMA(1,0,0)(0,0,1)[12] y SARIMA(1,0,1)(0,0,1)[12].

El modelo SARIMA(1,0,0)(0,0,1)[12], que conserva la estructura del ARMA inicial, presentó un AIC más alto (5749.18) y un p-valor muy bajo en la prueba de Ljung-Box ($p < 0.001$). Esta prueba evalúa la presencia de autocorrelación significativa en los residuos; la hipótesis nula establece que no existe autocorrelación (lo cual es deseable en un buen modelo), y la alternativa plantea que sí la hay. Como el p-valor fue menor a 0.05, se rechazó la hipótesis nula, lo que sugiere que el modelo no logra eliminar la dependencia temporal en los residuos.

Por el contrario, el modelo SARIMA(1,0,1)(0,0,1)[12] mostró un AIC menor (5734.45), un mejor ajuste de los residuos ($p = 0.0168$ en la prueba de Ljung-Box) y mantuvo la normalidad de los residuos ($p = 0.2883$).

en la prueba de Shapiro-Wilk). Además, todos los coeficientes estimados, incluido el término MA, fueron estadísticamente significativos, aportando valor explicativo al modelo.

En consecuencia, aunque el modelo ARMA(1,0) fue adecuado para una serie sin estacionalidad, la presencia de un patrón estacional justificó la incorporación de componentes adicionales. El modelo SARI-MA(1,0,1)(0,0,1)[12] logró capturar de forma más precisa la estructura temporal de los datos y mejorar el ajuste general del modelo. Por tanto, se concluyó que esta fue la opción más sólida y eficiente para modelar la serie temporal analizada.

```
Call:
arima(x = dif12ts_malaga, order = c(1, 0, 0), seasonal = list(order = c(0, 0, 1), period = 12))

Coefficients:
      ar1      sma1  intercept
    0.4053  -0.8909  -0.0970
s.e.  0.0363  0.0324  0.0852

sigma^2 estimated as 113.4:  log likelihood = -2870.59,  aic = 5749.18

Training set error measures:
      ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set 0.1289546 10.64925  8.336604 NaN   Inf  0.6021767 -0.04312407

Box-Ljung test

data:  residuos
X-squared = 69.534, df = 20, p-value = 2.17e-07

Shapiro-wilk normality test

data:  residuos
W = 0.99728, p-value = 0.2436
```

Figura 11: Pruebas realizadas a SARI-MA(1,0,0)(0,0,1)[12]

```
Call:
arima(x = dif12ts_malaga, order = c(1, 0, 1), seasonal = list(order = c(0, 0, 1), period = 12))

Coefficients:
      ar1      ma1      sma1  intercept
    0.7679  -0.4451  -0.9394  -0.1311
s.e.  0.0689  0.1012  0.0263  0.0777

sigma^2 estimated as 109.9:  log likelihood = -2862.22,  aic = 5734.45

Training set error measures:
      ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set 0.1072614 10.48507  8.241241 NaN   Inf  0.5952883  0.03123255

Box-Ljung test

data:  residuos
X-squared = 35.679, df = 20, p-value = 0.01677

Shapiro-wilk normality test

data:  residuos
W = 0.99743, p-value = 0.2883
```

Figura 12: Pruebas realizadas a SARI-MA(1,0,1)(0,0,1)[12]

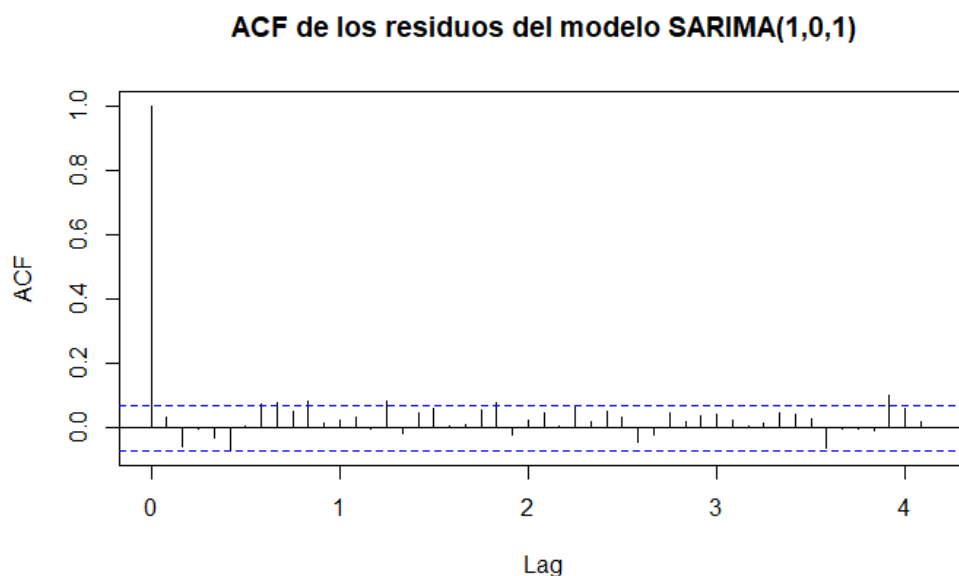


Figura 13: ACF de los residuos del modelo SARIMA(1,0,1)

3.5. Predicción y visualización

Calcula predicciones de la serie original para el próximo año. Calcula también los intervalos de confianza al 95 % para estas predicciones y representa en un mismo gráfico las predicciones junto con sus respectivos intervalos de confianza.

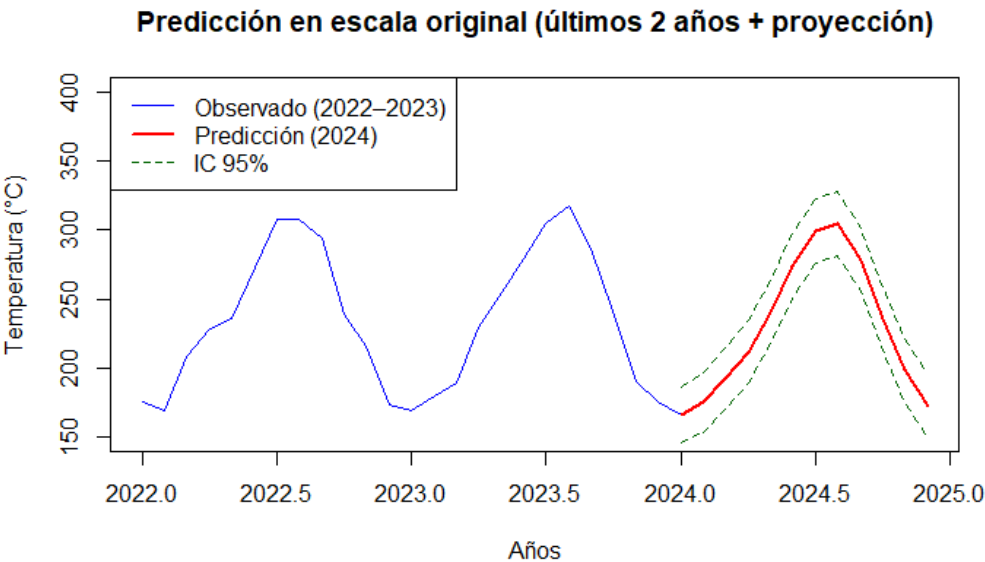
Se presentó la proyección de temperaturas mensuales para 2024 utilizando un modelo SARIMA seleccionado. Para construirlo, se eliminó la estacionalidad anual con una diferenciación de 12 meses, confirmando la

estacionariedad con pruebas como KPSS. Se compararon dos modelos SARIMA basados en el ARMA inicial y se eligió el que mejor ajustó la serie.

La predicción se realizó con la función `forecast()` en R, y luego se reconstruyó la serie en su escala original sumando las predicciones a los valores del mismo mes de 2023 para revertir la diferenciación estacional.

El resultado es una proyección mensual en la escala original de temperatura, con intervalos de confianza al 95 %, que permiten evaluar la variabilidad probable de las predicciones.

A continuación, se muestra el gráfico con la evolución estimada para 2024, junto con los valores proyectados y sus respectivos intervalos de confianza.



Mes_Año <chr>	Prediccion <dbl>	IC_95_inferior <dbl>	IC_95_superior <dbl>
2024-01	165.95	145.40	186.50
2024-02	175.09	153.49	196.68
2024-03	192.03	169.84	214.22
2024-04	211.81	189.28	234.34
2024-05	240.31	217.58	263.04
2024-06	273.80	250.95	296.65
2024-07	299.28	276.37	322.20
2024-08	304.70	281.74	327.65
2024-09	278.59	255.61	301.57
2024-10	236.06	213.07	259.06
2024-11	198.34	175.34	221.34
2024-12	172.36	149.35	195.36

Figura 14: Detalle de las predicciones e intervalo de confianza

3.6. Búsqueda de un modelo SARIMA alternativo

Analiza si existe (puede que no) otro modelo SARIMA que consideres más adecuado que el obtenido en el apartado anterior. Justifica en qué te basas para tu decisión y, si encuentras un modelo mejor, ajústalo.

Se compararon varios modelos SARIMA manteniendo fijos los parámetros estacionales en (0, 0, 1), ya que esta configuración capturó bien la estacionalidad y evitó sobreajuste y coeficientes no significativos. Otras configuraciones estacionales probadas presentaron coeficientes con p-valores altos, indicando poca relevancia, por lo que se mantuvo (0, 0, 1).

Respecto a los parámetros no estacionales, se evaluaron tres modelos diferentes, todos con componente estacional (0, 0, 1)[12]:

- modelo_sarima_1: (1, 0, 2)(0, 0, 1)[12]
- modelo_sarima_2: (2, 0, 1)(0, 0, 1)[12]
- modelo_sarima_3: (2, 0, 2)(0, 0, 1)[12]

El modelo SARIMA(1, 0, 0)(0, 0, 1)[12] fue descartado debido a que presentó un p-valor muy bajo en la prueba de Box-Ljung, indicando residuos con autocorrelación.

Se compararon los modelos con criterios como AIC, BIC, prueba de Box-Ljung para independencia de residuos, test de Shapiro-Wilk para normalidad y significancia de coeficientes (p-valor <0.05).

El modelo con mejor desempeño fue modelo_sarima_2, con un AIC de 5710.13 y BIC de 5737.90, los valores más bajos. Todos los modelos, excepto el modelo base, superaron la prueba de Box-Ljung con p-valores mayores a 0.05, indicando residuos independientes. La prueba de Shapiro-Wilk confirmó normalidad de residuos en todos los casos.

En resumen, aunque *modelo_sarima_2* tuvo el mejor desempeño general, el modelo base sigue siendo una opción válida por su simplicidad y buen ajuste, en línea con el modelo ARMA inicialmente propuesto en este informe. Además, ambos modelos dieron resultados muy similares en sus predicciones.

	df <dbl>	AIC <dbl>
modelo_sarima	5	5734.449
modelo_sarima_1	6	5714.185
modelo_sarima_2	6	5710.134
modelo_sarima_3	7	5711.694

Figura 15: Resultados AIC de los modelos SARIMA

	df <dbl>	BIC <dbl>
modelo_sarima	5	5757.589
modelo_sarima_1	6	5741.953
modelo_sarima_2	6	5737.903
modelo_sarima_3	7	5744.091

Figura 16: Resultados BIC de los modelos SARIMA

Mes_Año <chr>	Prediccion <dbl>
2024-01	169.84
2024-02	178.62
2024-03	195.68
2024-04	215.53
2024-05	244.80
2024-06	278.49
2024-07	304.19
2024-08	308.78
2024-09	282.69
2024-10	240.81

Figura 17: *modelo_sarima_2*

Mes_Año <chr>	Prediccion <dbl>
2024-01	165.95
2024-02	175.09
2024-03	192.03
2024-04	211.81
2024-05	240.31
2024-06	273.80
2024-07	299.28
2024-08	304.70
2024-09	278.59
2024-10	236.06

Figura 18: *modelo_sarima*