



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

**Análisis del Comportamiento del Ozono mediante Modelos
Dinámicos de Series Temporales**

**Autor: Adrián Espino Esteban
Tutora: María Pilar Rodríguez del Tío
Junio 2025**

Agradecimientos

A mis padres y familia por su apoyo constante en los momentos más duros de esta etapa. También a quienes, en la distancia, han estado a mi lado, acompañándome en este camino.

Agradezco especialmente a mi tutora Pilar por su dedicación y compromiso con la enseñanza, y también por haberme guiado a lo largo de este trabajo.

Índice General

1. Introducción	13
2. Marco Teórico	14
2.1. Metodología de Box-Jenkins	14
2.1.1. Serie Estacionaria	14
2.1.2. Modelo Autorregresivo (AR)	14
2.1.3. Modelo de Media Móvil (MA)	15
2.1.4. Modelo SARIMA	16
2.2. Multiple Seasonal and Trend Decomposition using Loess (MSTL)	16
2.3. Función de Transferencia: Modelo dinámico entre dos series estacionarias	17
2.3.1. Representación General	17
2.3.2. Función de Transferencia como cociente de polinomios	18
2.3.3. Función de correlación cruzada muestral	19
2.4. El agente contaminante Ozono (O_3)	19
3. Datos	22
3.1. Origen y Descripción de los datos	22
3.2. Limpieza y Preprocesamiento	23
3.2.1. Filtro de Kalman	24
3.2.2. Imputación de Valores Missing	25
4. Análisis Descriptivo	27
4.1. Serie diaria de Ozono (O_3)	27
4.1.1. Descomposición MSTL	29
4.1.2. Descomposición Clásica a partir de medias móviles	30
4.2. Serie diaria de Temperatura Media	31
4.2.1. Análisis de la tendencia lineal en la variable temperatura	31
4.3. Relación entre la temperatura media y la concentración de ozono	33
4.3.1. Comparación de índices diarios de Ozono y Temperatura	34
4.4. Serie diaria Dióxido de Nitrógeno (NO_2)	35
4.4.1. Relación entre el NO_2 y el O_3	36
4.5. Serie diaria Precipitación	36
5. Análisis univariante	38
5.1. Modelo ARIMA para datos diarios del contaminante O_3	38
5.2. Modelo ARIMA para datos diarios de Temperatura Media	44
5.3. Modelo ARIMA para datos diarios de Log Precipitación	47
5.4. Modelo ARIMA para datos diarios de NO_2	49
6. Modelos dinámicos para la predicción del ozono	51
6.1. Modelo dinámico con la temperatura como predictor	51
6.2. Modelo dinámico con precipitación como predictor	53
6.3. Modelo dinámico con NO_2 como predictor	56
6.4. Comparación de modelos	59

7. Conclusiones, limitaciones y líneas futuras	60
Bibliografía	63
Anexo I.	66
Anexo II.	73

Índice de figuras

1.	Quema de combustibles fósiles (Fuente: [9])	20
2.	Dataset Inicial Ozono	23
3.	Dataset Final Ozono	26
4.	Dataset Final	26
5.	Serie Original O ₃	27
6.	Box-Plot Mensual O ₃	27
7.	Bar Chart Concentración Anual O ₃	28
8.	Descomposición MSTL variable Ozono	29
9.	Descomposición MSTL estacionalidad semanal	29
10.	Descomposición Estacional periodo 365 (Ozono)	30
11.	Descomposición Estacional periodo 7 (Ozono)	30
12.	Serie Original Temperatura media	31
13.	Descomposición Clásica variable Temperatura media	31
14.	Recta ajustada a la tendencia 1974/2006	32
15.	Recta ajustada a la tendencia 2007/2024	32
16.	Recta ajustada a la tendencia 1974/2024	33
17.	Gráfico de Dispersión O ₃ y Temperatura	33
18.	Descomposición Estacional Conjunta	34
19.	Serie Original Dióxido de Nitrógeno (NO ₂)	35
20.	Distribución Original Dióxido de Nitrógeno (NO ₂)	35
21.	Gráfico de Dispersión NO ₂ frente a O ₃	36
22.	Serie temporal original de la precipitación	36
23.	Distribución de la precipitación l/m ²	36
24.	Evolución de la varianza móvil en la serie de precipitación	37
25.	ACF y PACF de la serie original del ozono	38
26.	ACF y PACF Serie Diferenciada con 1500 lags	39
27.	ACF y PACF Serie Diferenciada	39
28.	Datos Reales vs. Predicciones Nuevo Año	42
29.	Predicción para O ₃ 2024. Modelo 2	43
30.	ACF y PACF de la serie original de la temperatura	44
31.	ACF y PACF de la serie diferenciada hasta el retardo 900	44
32.	ACF y PACF de la serie diferenciada hasta el retardo 50	45
33.	Predicción para Temperatura 2024. Modelo 2	46
34.	ACF y Periodograma Log Precipitación	47
35.	ACF y PACF con 50 lags	47
36.	Ajuste del Modelo Log Precipitación	48
37.	ACF residual modelo ARIMA Log Precipitación	48
38.	ACF y Periodograma NO ₂	49
39.	ACF y PACF NO ₂ Diferenciado 1 y 365	49
40.	Ajuste del Modelo para NO ₂	50
41.	ACF residual modelo ARIMA NO ₂	50
42.	CCF entre O ₃ y Temperatura	51
43.	Ajuste Modelo Función de Transferencia para Ozono. [Código SAS: Anexo II]	52
44.	Predicción Año 2024 Función de Transferencia	53

45.	Correlaciones cruzadas entre $O_3(365)$ y <i>Log Precipitación</i>	54
46.	Parámetros ajustado para el modelo con Log Precipitación	55
47.	Predicciones 2024 Modelo dinámico con Precipitación	56
48.	CCF de O_3 y NO_2	57
49.	Parámetros ajustados para Función Transferencia con NO_2	57
50.	Análisis Residual Función Transferencia NO_2	58
51.	Predicciones 2024 Modelo dinámico con NO_2	58
52.	Comparación Modelos O_3	59
53.	PACF Serie O_3 Zoom Retardo 7	66
54.	Matriz Correlación Ozono Modelo 1	66
55.	Matriz Correlación Ozono Modelo 2	66
56.	Matriz Correlación Ozono Modelo 3	66
57.	Matriz Correlación Ozono Modelo 4	66
58.	Ajuste Ozono modelo 1	67
59.	Ajuste Ozono modelo 2	67
60.	Ajuste Ozono modelo 3	67
61.	Ajuste Ozono modelo 4	67
62.	ACF Residual Modelo ozono	68
63.	Predicción 2024 Serie O_3 últimas 30 obvs.	68
64.	Distribución Temperatura	69
65.	ACF Residuales Modelo ARIMA Temperatura Media	69
66.	Ajuste Temperatura Modelo 2	70
67.	Ajuste Temperatura Modelo 3	70
68.	Correlaciones Función de Transferencia Temperatura	71
69.	Correlaciones Función de Transferencia Precipitación	71
70.	Correlaciones Función de Transferencia NO_2	72
71.	Residuales Modelo Función de Transferencia	72

Índice de cuadros

1.	Ejemplo tabla contaminantes	22
2.	Ejemplo tabla O3	23
3.	Resultados del test de homocedasticidad de Breusch-Pagan	37
4.	Asimetría y kurtosis	37
5.	Tabla Ordenadas Periodograma Ozono	40
6.	Tabla Desviación Estándar del Ozono	40
7.	P-valores del Test de Ljung-Box para el Ozono	41
8.	Tabla de medidas de ajuste y selección para los modelos ARIMA Ozono	42
9.	Tabla Desviación Estándar de Temperatura	45
10.	P-valores del Test de Ljung-Box de Temperatura	46
11.	Tabla de medidas de ajuste y selección para los modelos ARIMA Temperatura	46
12.	Tabla Desviación Estándar de Log Precipitación	47
13.	Tabla Desviación Estándar del NO_2	49
14.	Comparación de modelos con medidas AIC y SSEp	59

Resumen

Este Trabajo de Fin de Grado tiene como objetivo analizar la influencia de diferentes variables meteorológicas y factores atmosféricos contaminantes en la modelización de la concentración de ozono troposférico. Para ello, se emplean modelos estadísticos clásicos como los modelos de Box-Jenkins (ARIMA) y modelos dinámicos de función de transferencia, utilizando datos diarios.

El estudio se centra en evaluar la relación entre algunas variables atmosféricas y determinar su impacto en la explicación y predicción de los niveles de ozono. La comparación entre modelos se realiza en términos de capacidad predictiva, con el fin de seleccionar el modelo más adecuado entre los ajustados.

El trabajo se apoya en herramientas estadísticas y de análisis de datos como SAS (de forma principal), así como R y Python para el tratamiento, modelización y visualización de los datos.

Palabras clave: Ozono, ARIMA, Función de Transferencia, SAS, Series Temporales, Contaminación atmosférica, Climatología, Predicción.

Abstract

This Final Degree Project aims to analyze the influence of various meteorological variables and atmospheric pollutant factors on the modeling of tropospheric ozone concentration. To this end, classical statistical models are employed, such as Box-Jenkins models (ARIMA) and dynamic transfer function models, using daily data.

The study focuses on evaluating the relationship between certain atmospheric variables and determining their impact on explaining and forecasting ozone levels. Model comparison is carried out in terms of predictive performance, with the goal of selecting the most suitable model among those fitted.

The work relies on statistical and data analysis tools such as SAS (primarily), as well as R and Python, for data processing, modeling, and visualization.

Keywords: Ozone, ARIMA, Transfer Function, SAS, Time Series, Air Pollution, Climatology, Forecasting.

1 Introducción

La necesidad de entender cómo evolucionan ciertos fenómenos a lo largo del tiempo ha hecho que el análisis de series temporales sea una herramienta clave en muchas áreas de estudio. En este trabajo se analiza el nivel de ozono como contaminante atmosférico, junto con otras variables relacionadas con la calidad del aire. Anticipar los niveles de contaminantes como el ozono es muy útil desde un punto de vista científico, con intereses para la salud pública y el cuidado del medio ambiente. Este estudio, por tanto, se centra en el modelado y la predicción de este contaminante atmosférico, utilizando técnicas estadísticas para lograr predecir su comportamiento futuro.

Uno de los enfoques principales del estudio es el uso de modelos Box-Jenkins, concretamente los modelos ARIMA con componentes estacionales. Estos modelos permiten identificar y modelar patrones estacionales en las series temporales, incluso en los casos en que se presentan múltiples estacionalidades. Veremos como algunas de las variables en estudio contendrán patrones estacionales dentro de otros, dando lugar a series con más de una estacionalidad.

Por otro lado, también se ajustarán modelos dinámicos de función de transferencia, que permiten introducir en el análisis otras variables que influyen e interactúan con los niveles de ozono troposférico. Esto es relevante en el contexto del estudio, donde otros factores climáticos y contaminantes interactúan de forma compleja. Incorporar esta información nos permitirá mejorar la capacidad predictiva de los modelos.

El análisis se ha realizado con herramientas del entorno de Python, R y SAS, especialmente el último, aprovechando su potencia para el tratamiento de datos y la construcción de este tipo de modelos. A través de este trabajo, se pretende evaluar la utilidad práctica de los modelos SARIMA y de los modelos de función de transferencia para describir y anticipar el comportamiento del ozono, comparando su capacidad predictiva.

2 Marco Teórico

2.1 Metodología de Box-Jenkins

Los modelos de Box-Jenkins, propuestos por **George Box y Gwilym Jenkins en 1970**, constituyen un enfoque metodológico para el análisis y modelado de series temporales univariantes. La metodología Box-Jenkins se caracteriza por su enfoque iterativo y se basa en tres etapas fundamentales: identificación del modelo, estimación de parámetros y validación del modelo.

Box-Jenkins propone modelos **autorregresivos (AR)** y de **media móvil (MA)**, en los que los valores futuros de una serie se expresan como combinaciones lineales de sus propios valores pasados y de errores aleatorios. El enfoque Box-Jenkins se aplica a series temporales estacionarias.

2.1.1 Serie Estacionaria

Cuando una serie $\{X_t\}$ es *débilmente estacionaria*, los coeficientes del modelo ARMA

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

se mantienen invariantes en el tiempo, y satisface simultáneamente las siguientes condiciones:

1. Media constante:

$$E[X_t] = \mu, \quad \forall t. \quad (1)$$

2. Varianza finita e invariante:

$$\text{Var}[X_t] = \sigma^2 < \infty, \quad \forall t. \quad (2)$$

3. Autocovarianza independiente del tiempo:

$$\gamma_k = \text{Cov}(X_t, X_{t-k}), \quad \forall t, k. \quad (3)$$

Estas propiedades garantizan que los coeficientes de un modelo (tanto autoregresivos ϕ_i como los de media móvil θ_j) permanezcan constantes en el tiempo, y que el efecto aleatorio ε_t decaiga.

Cuando la serie tiene tendencia, se aplica la diferenciación de orden d al operador *Backward*

$$Y_t = (1 - B)^d X_t, \quad (4)$$

para eliminar tendencia hasta conseguir estacionariedad. El **operador Backward** actúa sobre la serie temporal retardándola: ($BX_t = X_{t-1}$).

2.1.2 Modelo Autorregresivo (AR)

Un *modelo autorregresivo (AR)* es un modelo para series temporales que describe el valor actual de una variable como una combinación lineal de sus valores pasados y un error aleatorio. Formalmente, un modelo autorregresivo de orden p , denominado $\text{AR}(p)$, se expresa como:

$$X_t - \delta = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t, \quad (5)$$

donde X_t es la serie en el instante t , ϕ_1, \dots, ϕ_p son los coeficientes autorregresivos, y ε_t es un proceso de ruido blanco, con distribución $\varepsilon_t \sim N(0, \sigma^2)$.

El parámetro δ es la constante de la serie temporal, que puede ser significativa o no. Este parámetro viene dado por: $\delta = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$, con $\mu = E(X_t)$.

En la práctica, la función de autocorrelación parcial (PACF) de un $AR(p)$ se anula a partir del retardo $p + 1$, lo que facilita la identificación del orden del modelo. Los parámetros del modelo ϕ_i se estiman mediante métodos como mínimos cuadrados condicionados a valores faltantes o sin condicionar, estimación de máxima verosimilitud o tomando como valores iniciales en la implementación de estos métodos por procedimientos iterativos los obtenidos por aplicación de las ecuaciones de Yule-Walker.

Cuando las series temporales presentan patrones estacionales, se extiende el modelo autorregresivo para capturar dicha estructura. El modelo autorregresivo *estacional* puro restringe las observaciones que intervienen en el modelo a valores pasados separados por un número fijo de períodos, denominado componente estacional s . Un modelo autorregresivo estacional puro de orden P con periodicidad s se denota como $AR_s(P)$ y se expresa como:

$$X_t = \Phi_1 X_{t-s} + \Phi_2 X_{t-2s} + \dots + \Phi_P X_{t-Ps} + \varepsilon_t, \quad (6)$$

donde los coeficientes Φ_1, \dots, Φ_P capturan la dependencia estacional. Este modelo puede combinarse con componentes no estacionales, dando lugar al modelo general SARIMA.

2.1.3 Modelo de Media Móvil (MA)

El *modelo de media móvil* (MA) es un modelo para series temporales que describe el valor actual de una variable como una combinación lineal basada en promedios ponderados de errores pasados. Un modelo de media móvil de orden q , denotado $MA(q)$, se define como:

$$X_t - \delta = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (7)$$

donde $\theta_1, \dots, \theta_q$ son los coeficientes del modelo y $\varepsilon_t \sim N(0, \sigma^2)$ es un proceso de ruido blanco.

El modelo $MA(q)$ es siempre *estacionario*, a diferencia del modelo $AR(p)$ que su condición es que es *invertible*.

La *función de autocorrelación* (ACF) de un modelo $MA(q)$ se puede anular para retardos mayores que q , indicando a simple vista el orden del modelo.

Al igual que el modelo AR, los modelos de Media Móvil también pueden restringirse a errores en retardos múltiples del período estacional s . Un modelo puro de media móvil estacional de orden Q y periodicidad s , lo denotamos $MA_s(Q)$.

2.1.4 Modelo SARIMA

Los modelos autorregresivos y de media móvil pueden combinarse para capturar dependencias y generar así modelos más completos. Esta combinación da lugar al *Modelo ARMA*. Un modelo autorregresivo de media móvil de orden (p, q) , denotado como **ARMA** (p, q) , tiene p términos autorregresivos y q términos de media móvil:

o, de forma compacta, utilizando los operadores Backward:

$$\Phi(B)X_t = \Theta(B)\varepsilon_t, \quad (8)$$

Este modelo es adecuado solo para series *estacionarias*.

Cuando la serie no es estacionaria pero puede convertirse en estacionaria aplicando diferenciación, entonces se usa el *Modelo ARIMA*. Un modelo **ARIMA** (p, d, q) introduce un operador de diferenciación de orden d para convertir la serie en estacionaria:

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\varepsilon_t. \quad (9)$$

Además, muchas series tienen tanto no estacionariedad como estacionalidad. Para modelar este tipo de casos se usa el modelo *SARIMA* (Seasonal ARIMA), que extiende el ARIMA incorporando componentes estacionales autorregresivos, de media móvil y de diferenciación.

El modelo se denota como **SARIMA** $(p, d, q) \times (P, D, Q)_s$, donde (p, d, q) son los órdenes no estacionales ya mencionados y (P, D, Q) son los mismos órdenes pero estacionales. El modelo completo se expresa como:

$$\Phi(B)\Phi_s(B^s)(1 - B)^d(1 - B^s)^D X_t = \Theta(B)\Theta_s(B^s)\varepsilon_t, \quad (10)$$

donde $\Phi_s(L^s)$ y $\Theta_s(L^s)$ son los polinomios autorregresivos y de media móvil estacionales, respectivamente. Este esquema permite modelar tendencias complejas de series temporales.

2.2 Multiple Seasonal and Trend Decomposition using Loess (MSTL)

La descomposición de series temporales es una técnica que permite separar las componentes tendencia y estacionalidad del ruido blanco. El método **MSTL (Descomposición para patrón Estacional Múltiple y Tendencia usando Loess)** es una evolución de las técnicas de descomposición tradicionales, diseñado para abordar series con- esquema estacional de varios periodos.

MSTL se basa en una extensión del conocido método STL (Seasonal-Trend decomposition using Loess), desarrollado inicialmente por Cleveland et al. (1990). Mientras STL permite descomponer una serie temporal en componentes de tendencia, estacionalidad simple y residuales, MSTL amplía esta capacidad para manejar múltiples estacionalidades, lo cual es especialmente útil en contextos como datos diarios u horarios.

La estructura del modelo se representa de la siguiente forma:

$$Y_t = T_t + \sum_{j=1}^J S_{j,t} + R_t \quad (11)$$

donde:

- Y_t : valor observado en tiempo t
- T_t : componente de tendencia
- $S_{j,t}$: j -ésimo componente estacional (con J número de estacionalidades)
- R_t : componente aleatorio (Residuos)

El procedimiento de descomposición en MSTL se basa en regresión local, concretamente en el método **LOESS (Locally Estimated Scatterplot Smoothing)**, para estimar suavemente los distintos componentes de la serie. Esto permite capturar cambios no lineales en la tendencia y la estacionalidad. Este enfoque es altamente flexible y permite personalizar los parámetros de suavizado y la cantidad de estacionalidades a considerar, siendo robusto ante valores atípicos y adaptable a estructuras complejas.

2.3 Función de Transferencia: Modelo dinámico entre dos series estacionarias

Los modelos de función de transferencia constituyen una herramienta fundamental para el análisis dinámico entre series temporales. Su objetivo principal es modelar la relación entre una variable de salida (output) y una o más variables de entrada (inputs), bajo el supuesto de que las variables de entrada afecten a la salida. Esta estructura permite estudiar cómo la evolución de ciertos factores externos influye sobre una variable de interés.

A diferencia del modelo lineal clásico, que establece relaciones instantáneas y asume errores independientes.

La función de Transferencia es útil para la predicción. En contextos donde las variables explicativas son observables y controlables, los modelos de función de transferencia pueden mejorar significativamente la calidad de las predicciones respecto a los modelos univariantes. A continuación, veremos cómo se construyen los modelos de función de transferencia.

2.3.1 Representación General

Una representación inicial de la función de transferencia en una relación unidireccional entre una variable explicativa x_t y una variable respuesta y_t puede expresarse mediante el siguiente modelo:

$$y_t = v_0x_t + v_1x_{t-1} + v_2x_{t-2} + \dots + n_t \quad (12)$$

donde los coeficientes v_i constituyen la *función de transferencia*, la cual describe cómo la serie x_t influye dinámicamente sobre y_t . El término n_t representa el *proceso de perturbación*, que es también

un proceso estacionario y recoge el efecto de otras variables no incluidas en el modelo o de ruidos aleatorios.

La ecuación también puede escribirse mediante el operador de retardo Backward B , tal que $Bx_t = x_{t-1}$:

$$y_t = v(B)x_t + n_t \quad (13)$$

donde la función de transferencia $v(B)$ se define como:

$$v(B) = v_0 + v_1B + v_2B^2 + \dots \quad (14)$$

La *ganancia* de la función de transferencia, que representa el efecto a largo plazo de un cambio permanente en x_t sobre y_t , se define como:

$$g = v(1) = \sum_{i=0}^{\infty} v_i \quad (15)$$

Por otra parte, el proceso de perturbación n_t se modela mediante un modelo ARMA estacionario. El conjunto formado por las ecuaciones (13) y este modelo ARMA constituyen un *modelo de función de transferencia*, utilizado para modelar relaciones entre series temporales.

2.3.2 Función de Transferencia como cociente de polinomios

La expresión obtenida en (12) no es útil para estimar en la práctica, ya que si x_t contiene alta autocorrelación será difícil separar sus efectos por el problema de multicolinealidad. Además, será muy probable que haya que estimar demasiados parámetros. Una forma alternativa de escribir la función de transferencia es aproximar una función infinita como cociente de polinomios, de la siguiente manera:

$$v(B) = \frac{\omega_m(B)}{\delta_a(B)} B^b, \quad (16)$$

donde:

$$\omega(B) = \omega_0 + \omega_1B + \dots + \omega_mB^m, \quad (17)$$

$$\delta(B) = 1 - \delta_1B - \dots - \delta_sB^s, \quad (18)$$

y b como retardo inicial. Aquí, $\omega(B)$ actúa como componente de media móvil y $\delta(B)$ como componente autorregresivo. Para que el sistema sea estable, las raíces de $\delta(B)$ deben estar fuera del círculo unitario.

Una ventaja de esta representación es que permite separar el efecto inicial del patrón de la respuesta, facilitando así la interpretación dinámica del modelo. Además, en la práctica, los grados m y a de los polinomios suelen mantenerse bajos, lo cual simplifica la estimación y reduce el riesgo de sobreajuste.

La forma general que se da al modelo de función de transferencia:

$$Y_t = \delta + \frac{\omega_0 + \omega_1 B}{1 - \delta_1 B} X_{t-b} + N_t \quad (19)$$

donde:

- Y_t es la serie respuesta que se desea modelar
- X_t es la serie explicativa, mientras que b representa el retardo mínimo de impacto: indica cuántos periodos después de un cambio en X_t se empieza a reflejar su efecto en Y_t
- **Numerador** $\omega_0 + \omega_1 B$: Es un polinomio en el operador retardo. Describe el efecto inmediato y de retardo del predictor X_t sobre la salida Y_t .
- **Denominador** $1 - \delta_1 B$: Es otro polinomio en B , que representa la *dinámica de decaimiento o memoria* del sistema.
- $N_t \sim \text{ARIMA}(p, d, q)$: Es el *componente de ruido* que se modela como un proceso ARIMA para capturar la autocorrelación residual no explicada por la entrada. Es el modelo univariante de la variable respuesta.
- δ : constante del modelo.
- B : operador de retardo Backward.

2.3.3 Función de correlación cruzada muestral

La función de correlación cruzada es una herramienta fundamental en la identificación de modelos de función de transferencia. Dada una muestra de tamaño n , la función de correlación cruzada muestral se estima mediante:

$$\hat{\rho}_{xy}(j) = \frac{\hat{\gamma}_{xy}(j)}{S_x S_y}, \quad (20)$$

donde $\hat{\gamma}_{xy}(j)$ es la covarianza muestral entre x_t e y_{t+j} , y S_x, S_y son las desviaciones típicas muestrales de las respectivas series.

La función de autocorrelación muestral cruzada (CCF) permite detectar si existe una relación significativa entre una serie explicativa x_t y una serie respuesta y_t , así como determinar en qué retardos esa relación es más fuerte, antes de ajustar un modelo de función de transferencia.

Cuando la CCF refleja la relación tras la estimación del modelo, esta debe ser indistinguible de 0, indicando que la relación dinámica ha sido completamente capturada por el modelo y que no queda dependencia estructural no explicada.

2.4 El agente contaminante Ozono (O_3)

El ozono (O_3) es un gas que puede actuar como contaminante cuando se encuentra en la troposfera (capa baja de la atmósfera). Se forma por reacciones químicas entre óxidos de nitrógeno (NO_x) y

compuestos orgánicos volátiles (COVs) en presencia de luz solar. Como contaminante, el ozono puede afectar la calidad del aire, la salud humana y el medio ambiente, influyendo en fenómenos como el calentamiento urbano y la variación de la temperatura. En estudios ambientales, se analiza su impacto en la temperatura y su relación con otros factores climáticos.

En Valladolid, se han registrado episodios de alta concentración de ozono que han llevado a la activación de medidas preventivas. Por ejemplo, en agosto de 2024, el Ayuntamiento activó la 'Situación 1, preventiva' tras superar durante tres días consecutivos el valor de 100 microgramos por metro cúbico en las estaciones de medición de la Red de Control de la Contaminación Atmosférica. Estos episodios suelen estar asociados a condiciones meteorológicas específicas, como olas de calor, que favorecen la formación de ozono troposférico [7]

En este contexto, comprender el comportamiento del ozono troposférico y su interacción con variables climáticas resulta útil para anticipar episodios críticos, implementar políticas ambientales eficaces y proteger la salud pública frente a los efectos adversos de este contaminante.



Figura 1: Quema de combustibles fósiles (Fuente: [9])

Efecto Fin de Semana (EFS)

El ozono troposférico es un contaminante secundario del aire, cuyas concentraciones se incrementan por las reacciones fotoquímicas de los óxidos de nitrógeno NO_x y COV en presencia de radiación solar:



EL NO_2 con radiación solar se divide dando NO y O . Con ese átomo de O liberado, junto con O_2 se formará el ozono. En ausencia de COV, el NO da lugar a la escisión del O_3 formando NO_2 y O_2 . Esto es denominado titulación, y es el proceso mediante el cuál el NO provoca la destrucción de ozono.

Cuando existe COV en la atmósfera, estos pueden oxidarse a radicales peroxilo (RO_2), los cuales reaccionan con el NO , promoviendo la formación de NO_2 sin consumir ozono:



Durante los fines de semana, la menor emisión de NO implica una menor titulación de ozono (ecuación 23), lo que permite que las concentraciones de O_3 aumenten. Esto es especialmente relevante en entornos donde el régimen químico es limitado por COV (VOC-limited). Cuando hay una gran cantidad de (NO_x) pero una cantidad relativamente baja de COV (relación COV/ NO_x baja), la formación de ozono se encuentra limitada por la disponibilidad de COV. A esto se le denomina régimen limitado por COV. En estas condiciones, una reducción en las emisiones de NO_x (fines de semana) puede disminuir la titulación del ozono (la reacción que lo destruye), permitiendo que se acumulen concentraciones más altas de O_3 . Por tanto, paradójicamente, menos emisiones de NO pueden conducir a un mayor nivel de ozono en este tipo de **entornos urbanos** [17].

Este fenómeno, denominado **efecto fin de semana**, refuerza la idea de un componente semanal en el análisis de la concentración de ozono.

Sin embargo, tras comprender el *efecto fin de semana*, podemos pensar cómo es posible que durante la pandemia del COVID-19, como hemos visto anteriormente, el nivel de emisiones O_3 en el aire se redujese. Un artículo que aborda específicamente este fenómeno [18], nos ayuda a comprender por qué ocurre.

En el contexto de la pandemia se observó una reducción generalizada del ozono troposférico. Este comportamiento aparentemente contradictorio con la anterior idea planteada, se explica por la reducción simultánea tanto de NO_x (NO_2 principalmente) como de compuestos orgánicos volátiles (COVs) que son, como ya mencionamos, precursores fundamentales del ozono. Al disminuir ambas categorías de contaminantes, se redujo también la capacidad de la atmósfera para generar ozono mediante reacciones fotoquímicas. Además, en zonas donde el régimen químico es controlado por NO_x (donde un exceso de NO_x favorece la formación de ozono) la fuerte disminución de estas emisiones hizo menos eficiente la producción de O_3 . Por tanto, el confinamiento y el parón industrial indujeron condiciones más limpias en la atmósfera.

3 Datos

3.1 Origen y Descripción de los datos

Los datos utilizados a lo largo de este trabajo están relacionados con información climatológica y niveles de contaminación en la ciudad de Valladolid. En particular, abordaremos la relación entre la temperatura media, NO_2 y precipitación con el ozono O_3 . Para realizar este estudio se van a tomar datos de dos fuentes distintas. El ozono y NO_2 han sido extraídos de la Red de Control de la Contaminación Atmosférica del Ayuntamiento de Valladolid [14].

La estación donde han sido recogidos los datos es la estación meteorológica Puente Poniente - M^a Luisa Sánchez. De aquí, se ha tomado una muestra de datos diarios comprendida entre el 1 de Enero de 2010 al 31 de Diciembre de 2024. Las mediciones de la concentración de ozono y NO están tomadas en $\mu g/m^3$.

Los datos climatológicos han sido obtenidos del Repositorio de datos abiertos de Aemet [15]. La estación indicada en este caso es Valladolid - Valladolid.

El valor de medición que usaremos en este estudio será el valor de la Temperatura Media tomado diariamente en la estación mencionada. Este valor medio se obtiene de hacer la media diaria entre la Temperatura Máxima y Mínima registrada en dicho medidor.

El periodo extraído en este caso será entre el 1 de Enero de 1974 al 31 de Diciembre de 2024. Estos datos serán usados, también, para estudiar la tendencia de la temperatura a lo largo del tiempo, con un amplio periodo para sacar conclusiones robustas. Por otro lado, un periodo menor, y de igual longitud que el Ozono, que se extenderá del 1 de Enero de 2010 al 31 de Diciembre de 2024. Este segundo periodo más acotado será usado para ajustar el modelo que explique la variable del Ozono y las variaciones que sufre este en los últimos años. Para los datos de Precipitación, el periodo escogido será el mismo.

A continuación, se da un ejemplo del formato que comprenden ambas fuentes de datos que vamos a utilizar.

Fecha	Periodo	$O_3(\mu g/m^3)$	F_O3	$NO_2(\mu g/m^3)$	F_NO2
11/03/2021	11	38	V	7	V

Cuadro 1: Ejemplo tabla contaminantes

Como podemos ver, la fuente de datos correspondiente a los agentes contaminantes nos proporciona la **Fecha** en la que fue tomado el dato, un indicador **Periodo** que corresponde con el día del mes, $O_3(\mu g/m^3)$ y $NO_2(\mu g/m^3)$ con los valores contaminantes de dicho día y **F_O3** indicando si el valor de O_3 es correcto o nos encontramos ante un missing.

La Tabla 2 contiene un resumen de la tabla obtenida de Aemet, que contiene varias variables meteorológicas, y entre las cuales hemos elegido, como ya se ha mencionado, la temperatura media y la precipitación como medidas para estudiar su relación con los contaminantes registrados en la ciudad. Esta tabla contiene la Fecha en las que fueron tomadas las medidas y el valor de la misma.

Para realizar este estudio, se ha optado por ajustar un modelo que explique la variación del ozono (O_3)

Fecha	temperatura_media	Temperatura Máxima	Temperatura Mínima	Presión Máxima
11/03/2016	4,7	9,4	0	4

Presión Mínima	Racha Máxima	Velocidad Media	Horas de Sol	Precipitación
4	4	4	4	4

Cuadro 2: Ejemplo tabla O3

en la ciudad de Valladolid, considerando las demás variables como factores influyentes. Por este motivo, se usan estas variables en adelante y buscaremos crear un único dataset con ellas, obviando el resto de datos extraídos.

3.2 Limpieza y Preprocesamiento

En principio, se tienen dos conjuntos de datos en los cuáles hay variables que no usaremos en el modelo, así como valores missing dentro de las variables en estudio. Estos problemas surgidos con el Ozono, también han surgido con el NO_2 y se han resuelto de la misma manera que vamos a ver. Estos problemas que han surgido durante el preprocesamiento de datos, son lo que habrá que mitigar en esta sección.

	FECHA	O3 ($\mu\text{g}/\text{m}^3$)	F_03
1493	2014-02-02 00:00:00	40	V
1494	2014-02-03 00:00:00	-100000	M
1495	2014-02-04 00:00:00	-100000	M
1496	2014-02-05 00:00:00	-100000	M
1497	2014-02-06 00:00:00	-100000	M
1498	2014-02-07 00:00:00	58	V
1499	2014-02-08 00:00:00	74	V
1500	2014-02-09 00:00:00	61	V

Figura 2: Dataset Inicial Ozono

Como podemos notar en Figura 2 nuestros datos contienen valores missing asignados como "-100000". La variable indicador codifica estas filas como valores faltantes. Por otro lado, en el dataset inicial de ozono se puede apreciar que faltan filas correspondientes a ciertas fechas, por lo que esas filas inexistentes también hay que recuperarlas.

En primer lugar, se han tenido que unir los dataset anuales desde el 2010 hasta el 2024, ya que la fuente de datos [14] solo permite obtener las tablas de datos diarios de manera anual.

Ante este conjunto de datos ya con todos los años unidos en un único archivo, lo siguiente es encontrar los días que faltan dentro del periodo indicado para poder ser recuperados y, en principio, ser asignados como NaN's. La búsqueda de estos días se realizó desde R donde se pudieron identificar y

añadir las filas faltantes a partir de la Fecha, detectando los días que no se encuentran en un principio dentro del dataset. Las filas con la clave "Fecha" que no se encontraban en el dataset original y que, por tanto, han tenido que ser añadidos a este son:

01/01/2010, 26/05/2017-01/06/2017, 11/03/2019, 08/04/2019, 16/08/2019, 16/12/2019, 17/12/2019, 20/01/2020, 21/01/2020, 20/04/2020, 15/07/2020, 31/08/2020, 23/11/2020, 08/07/2020, 28/09/2020, 29/03/2021, 30/03/2021, 12/04/2021,26/07/2021, 17/08/2021, 22/11/2021
--

Tras tener ya de forma definitiva las 5479 filas correspondientes a todos los días comprendidos entre 01/01/2010 - 31/12/2024, hay que abordar el problema de los datos faltantes.

3.2.1 Filtro de Kalman

El filtro de Kalman es un algoritmo de estimación recursiva ampliamente utilizado en la predicción en series temporales.

Éste es especialmente útil en la imputación de valores faltantes en series temporales, ya que puede predecir datos ausentes basándose en la evolución dinámica del resto del conjunto. A diferencia de otros métodos de interpolación o regresión, este enfoque permite estimaciones más precisas al modelar explícitamente la estructura de los datos y tener en cuenta la incertidumbre asociada a dicha estimación.

Modelo matemático del Filtro de Kalman

El filtro de Kalman se basa en un modelo de espacio de estados, que representa la evolución de un sistema mediante dos ecuaciones fundamentales:

La ecuación de estado describe cómo evoluciona el sistema a lo largo del tiempo:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \mathbf{w}_k$$

Donde:

- \mathbf{x}_k es el vector de estado en el instante k
- \mathbf{F}_k es la matriz de transición de estado, que define cómo el estado anterior \mathbf{x}_{k-1} influye en el estado actual.
- $\mathbf{B}_k \mathbf{u}_k$ entrada de control.
- $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}_k)$ es un ruido del proceso.

Dado que los estados del sistema no siempre son directamente observables, se modela la relación entre el estado y las mediciones disponibles a través de la ecuación de observación:

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k$$

Donde:

- \mathbf{z}_k es el vector de observaciones en el tiempo k .

- \mathbf{H}_k es la matriz de observación, que define cómo el estado \mathbf{x}_k se traduce en la medición \mathbf{z}_k .
- $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R}_k)$ es el ruido de medición, que sigue una distribución normal con covarianza \mathbf{R}_k .

Cuando una serie temporal presenta valores faltantes, el filtro de Kalman puede estimarlos utilizando su mecanismo de predicción y actualización. La imputación se basa en dos fases principales:

Si un valor de la serie está ausente en el instante k , se realiza una predicción basada únicamente en la ecuación de estado:

Predicción del estado:

$$\hat{\mathbf{x}}_k |_{k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1} |_{k-1} + \mathbf{B}_k \mathbf{u}_k$$

Aquí, $\hat{\mathbf{x}}_k |_{k-1}$ representa la estimación del estado en k basada en información hasta $k - 1$.

Cuando los valores ausentes son cortos, esta predicción suele ser suficiente para obtener una buena estimación. Sin embargo, cuando hay largas secuencias de valores faltantes, la incertidumbre aumenta, y el filtro debe manejar la propagación del error de predicción.

En estos casos, se continúa aplicando la predicción sin corrección durante los periodos sin observaciones. A medida que se extiende la ausencia de datos, la incertidumbre P_k crece, reflejando la acumulación del error de predicción. Cuando vuelven a estar disponibles observaciones, se calcula la ganancia de Kalman que recalibra, ajustando las estimaciones previas y corrigiendo el estado.

3.2.2 Imputación de Valores Missing

En el caso que estamos abordando, los datos son tomados a partir de mediciones en monitores automáticos de absorción ultravioleta, que forman parte de las infraestructuras de las redes de vigilancia de la calidad del aire, constituidas por cabinas donde se integran medidores de otros contaminantes junto a instrumentación meteorológica. En consecuencia, hay que tener en cuenta que estos tipos de sistemas pueden dejar de tomar muestras debido a varias causas como puede ser: Fallos en los sensores, interrupciones en la comunicación, situaciones de mantenimiento o de error humano, etc. Por todo esto, los datos que vamos a manejar contienen zonas en las que los valores missing ocupan varios días consecutivos. A lo largo de esta sección vamos a tratar este problema para poder conseguir un conjunto de datos completo.

Para imputar los valores missing que tiene el conjunto de datos se tienen en consideración dos posibles opciones. Una es el suavizado exponencial Holt-Winters y otra es el filtro de Kalman. Aunque ambos métodos obtienen resultados concordantes, se ha decidido usar el Filtro de Kalman, ya que éste es un mejor método a la hora de imputar largas cadenas de valores missing. Esto es así debido a que tiene en cuenta la evolución temporal de los datos. El Filtro de Kalman, además, usa un enfoque que permite actualizar las estimaciones a medida que se reciben nuevas observaciones.

En primer lugar, se convierte la codificación dada en el dataset inicial para los valores faltantes y se transforma en NaN's. Después, usando en este caso la cobertura dada por la herramienta R para este problema se imputan los datos ausentes inicialmente para completar el conjunto de datos correspondiente al ozono. Finalmente este conjunto de datos tendrá la siguiente estructura:

	FECHA	O3
1493	2014-02-02 00:00:00	40
1494	2014-02-03 00:00:00	49.7212
1495	2014-02-04 00:00:00	52.4394
1496	2014-02-05 00:00:00	55.1576
1497	2014-02-06 00:00:00	57.8758
1498	2014-02-07 00:00:00	58

Figura 3: Dataset Final Ozono

Como podemos ver, los valores NaN's han sido reemplazados usando el Filtro de Kalman por la estimación a partir del resto de valores del conjunto de datos. Por otro lado, se ha decidido prescindir de la columna de control, ya que una vez procesados los datos ya no es necesaria para nuestro caso de estudio.

En cuanto al conjunto de datos obtenido de Aemet ha tenido un procedimiento similar, pero más sencillo. Se ha creado un nuevo conjunto de datos con las variables *Temperatura_media*, *Precipitación* y la variable *Fecha*. Este dataset contaba también con algún valor missing el cual ha sido imputado de la misma manera contada anteriormente.

Por último, se han unido ambos dataset por medio de la Fecha y se ha creado el conjunto de datos que se utilizará a lo largo de este estudio.

	Fecha	O3	temperatura_media	precipitacion	N02
0	2010-01-01 00:00:00	26.2258	5	0.7	27.8656
1	2010-01-02 00:00:00	26	4.6	4.4	28
2	2010-01-03 00:00:00	29	6.7	8.1	29
3	2010-01-04 00:00:00	24	6.6	9.3	28
4	2010-01-05 00:00:00	34	4.3	0.9	15
5	2010-01-06 00:00:00	37	2.85	0	22
6	2010-01-07 00:00:00	38	0.9	0.2	17
7	2010-01-08 00:00:00	53	-0.6	0.1	11
8	2010-01-09 00:00:00	44	-1.1	0.6	16
9	2010-01-10 00:00:00	28	-3.05	1.1	26

Figura 4: Dataset Final

4 Análisis Descriptivo

En esta sección se presenta el análisis descriptivo de las series en estudio, con el objetivo de comprender su comportamiento. Para ello, se llevará a cabo la visualización y exploración de los datos, lo que permitirá identificar patrones, periodos relevantes y la existencia de posibles valores atípicos.

4.1 Serie diaria de Ozono (O_3)

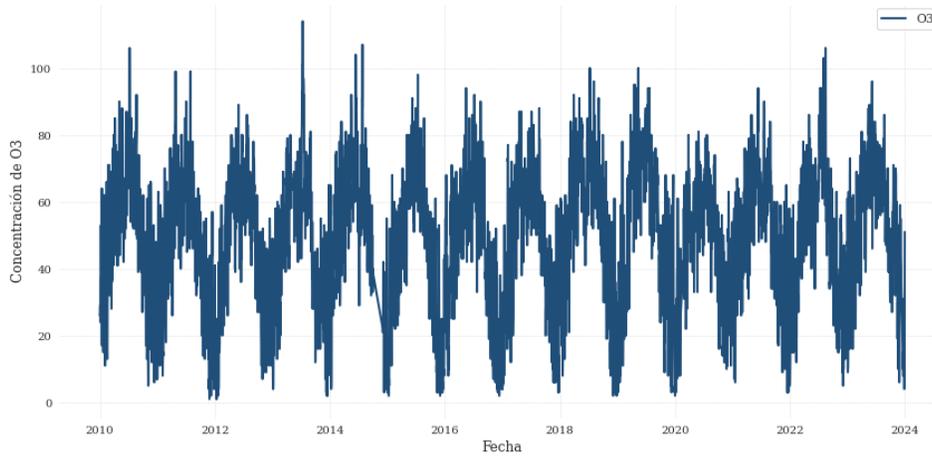


Figura 5: Serie Original O_3

En el gráfico de la Figura 5 observamos los datos extraídos de la estación meteorológica Puente Poniente - M^a Luisa Sánchez que corresponden al Ozono. Esta serie está formada por 5479 observaciones mensuales (2010-2024).

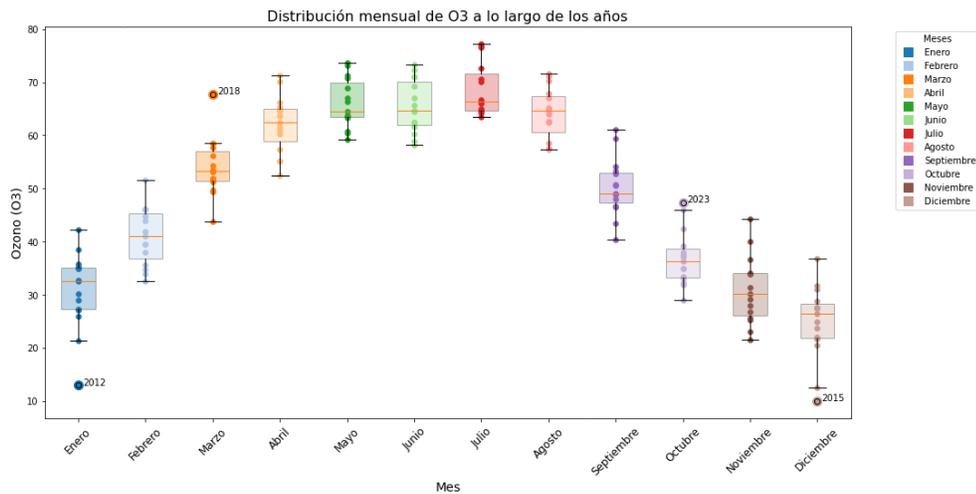


Figura 6: Box-Plot Mensual O_3

El gráfico de la Figura 6 muestra los promedios mensuales de cada año en los niveles de ozono en el aire mediante diagramas de caja, así como la variabilidad entre los meses de distintos años, y de los meses entre sí. Junto a los outliers observados tenemos una etiqueta que indica el año del que se extrae el dato. Los meses de Verano son los que muestran una mayor concentración de Ozono en el aire, mientras que, los meses de Invierno son los meses con menor concentración. Esto nos puede indicar una posible relación entre el aumento de la temperatura con el aumento del ozono, ya que los meses con mayores concentraciones de ozono coincide con los meses más calurosos.

Aunque no hay demasiados outliers, podemos destacar Enero de 2012 como mes que tuvo un concentración especialmente baja. En el lado contrario está Marzo de 2018, con la mayor concentración del gas en el ambiente registrada en nuestro periodo.

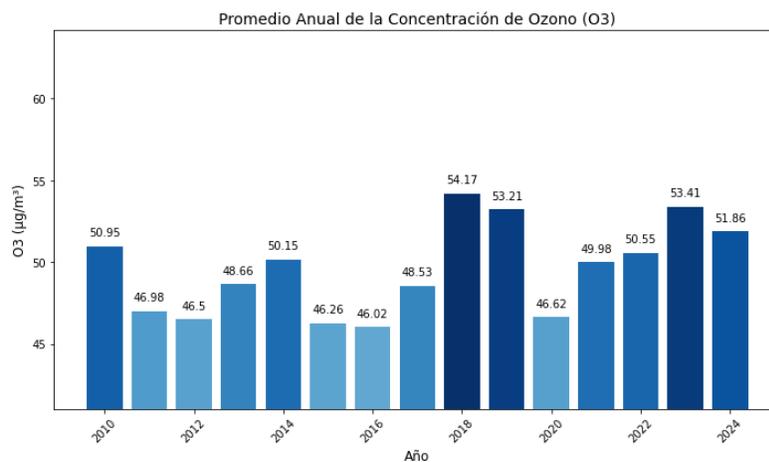


Figura 7: Bar Chart Concentración Anual O3

El gráfico de barras de la Figura 7 muestra el promedio anual de concentración de ozono O3 ($\mu\text{g m}^{-3}$). Se muestra el promedio de manera numérica encima de las barras de cada año. Además, tenemos una escala de azules de más claros a más oscuros dependiendo de si la concentración promedio del contaminante ha sido menor o mayor en dicho año.

Podemos ver que de manera general, la concentración anual ha ido incrementando en los últimos años. Esto puede ser debido a un aumento generalizado de las temperaturas, como también puede deberse a una mayor presencia de otros contaminantes en la atmósfera que tengan un efecto sobre el ozono. Podemos observar que en el año 2020 la concentración de ozono es inferior al promedio registrado en los años contiguos, en concreto de 46.62 (ppb), lo que supone una disminución en un $\approx 12,4\%$ con respecto al año anterior, siendo un decremento considerable.

Esta disminución puede estar dado por la pandemia de la COVID-19 sufrida en dicho año, durante la cual se produjo una paralización parcial de la actividad, emitiéndose menos contaminantes (los cuáles también pueden afectar a la concentración de O3) por factores como el confinamiento, la bajada de la producción industrial, entre otros.

4.1.1 Descomposición MSTL

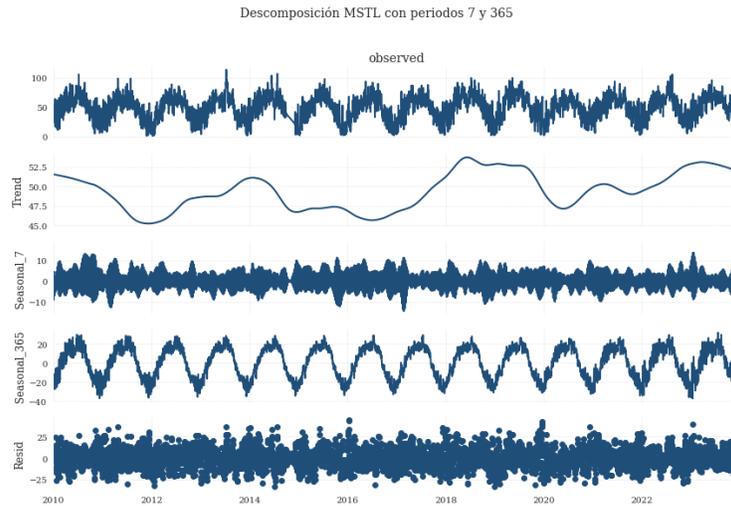


Figura 8: Descomposición MSTL variable Ozono

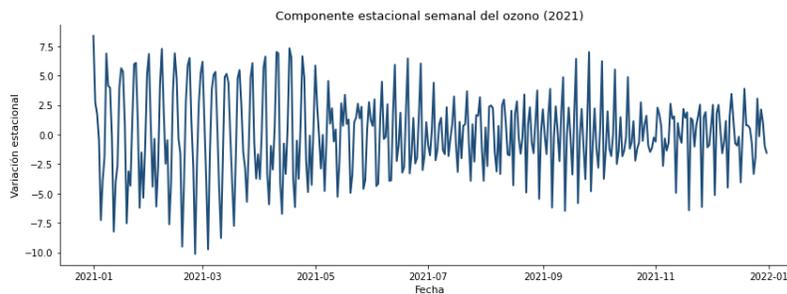


Figura 9: Descomposición MSTL estacionalidad semanal

Continuando con el análisis de la variable Ozono, tenemos la Figura 8 que nos muestra la Descomposición MSTL (Multiple Seasonal and Trend decomposition using Loess). Esta técnica novedosa permite descomponer una serie temporal compleja en tres componentes fundamentales: tendencia, estacionalidad y residuos. El análisis se realiza para comprender mejor la estructura subyacente de los datos, lo cual es fundamental para tareas como la predicción o el modelado. Con ello, veremos como se comporta la tendencia y estacionalidad del O₃. [Marco teórico MSTL: 2.2]

En la primera fila de la Figura 8 tenemos la serie original, ya vista en la Figura 5. En el segundo gráfico, tenemos de manera suavizada las fluctuaciones que revelan la evolución a largo plazo del contaminante. Durante los años comprendidos entre 2016 a 2020 vemos como a existido un aumento considerable, mientras que en 2020 cayeron las emisiones, como ya hemos indicado, para más adelante volver a repuntar.

En el gráfico de la Figura 9 vemos la estacionalidad semanal de la serie ampliando sobre el año 2021 para identificar mejor el patrón subyacente. Este componente revela patrones que se repiten de forma regular semanalmente, con periodos de mayor amplitud dentro de cada año. En el cuarto gráfico, tenemos la estacionalidad anual de la serie. Ésta presenta un patrón claramente cíclico a lo largo del tiempo. Se

pueden identificar ciclos anuales marcados, indicando una fuerte dependencia estacional. El periodo identificado es de 365 ($s = 365$).

El último gráfico, corresponde a los residuos del modelo MSTL. Los residuos parecen no presentar una estructura clara, lo que es deseable para un buen ajuste. Sin embargo, se puede llegar a observar cierta heterocedasticidad, lo que puede indicar que en algunos periodos puede existir mayor incertidumbre. Esto puede estar provocado también por la existencia de otros periodos por debajo del identificado, que nos ayuden también a explicar la serie.

4.1.2 Descomposición Clásica a partir de medias móviles

Una vez observada la descomposición MSTL, intentaremos responder a una duda que nos surge durante el análisis. ¿Es posible que exista una segunda estacionalidad semanal de este gas?. Para poder resolverla, debemos de realizar una Descomposición Clásica con media móvil semanal, para comprobar si puede existir este efecto en la emisión y presencia de este gas en el aire.

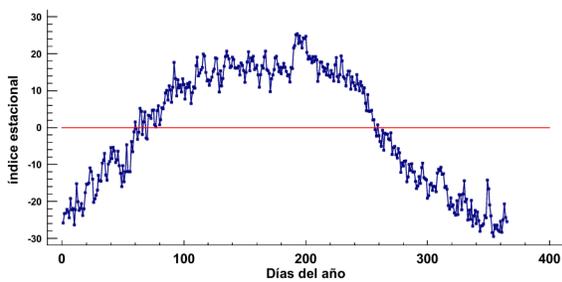


Figura 10: Descomposición Estacional periodo 365 (Ozono)

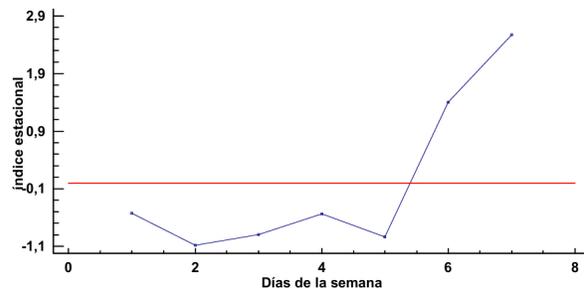


Figura 11: Descomposición Estacional periodo 7 (Ozono)

La Figura 10 contiene los índices estacionales de periodo 365 obtenidos por la descomposición. A la derecha Figura [11] tenemos los índices estacionales de periodo 7 obtenidos por la descomposición clásica. Los índices estacionales para ambas descomposiciones clásicas comienzan el Lunes y terminan en Domingo. Los días del fin de semana (Sábado y Domingo) se tienen concentraciones de O_3 muy superiores al resto de la semana. Aunque este fenómeno es en principio contra-intuitivo, realmente existe y se le ha denominado como el **Efecto Fin de Semana (EFS)**.

El EFS ha sido observado en diversas ciudades a lo largo del mundo. Este efecto se caracteriza por el aumento de las concentraciones de O_3 durante los fines de semana, a pesar de una disminución generalizada en las emisiones de sus precursores: los óxidos de nitrógeno (NO_x) y los compuestos orgánicos volátiles (COV). [Marco teórico EFS: 2.4]

4.2 Serie diaria de Temperatura Media

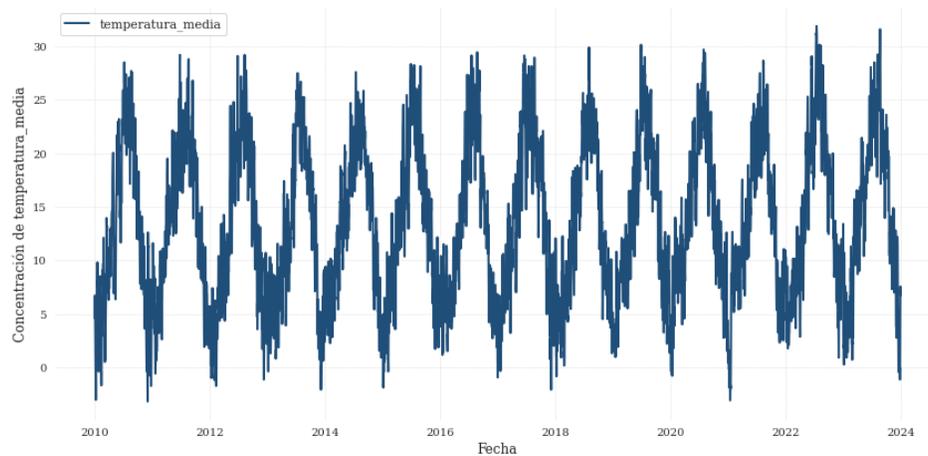


Figura 12: Serie Original Temperatura media

Los datos extraídos del Repositorio de Datos Abiertos de AEMET sobre la evolución de la temperatura media están representados en la Figura 12.

Durante los meses centrales las temperaturas son bastante más elevadas que las temperaturas en los meses de Invierno.

La **descomposición STL** de temperatura nos muestra una descomposición en tres componentes. En la tendencia, podemos ver una curva que, con ciertas variaciones, tiene un comportamiento incremental a largo plazo, lo cual puede ser debido a un ligero aumento general de las temperaturas. En la estacionalidad de esta variable tenemos patrones cíclicos de forma regular a lo largo del tiempo, que se corresponderán a una estacionalidad anual. Por último, los residuos contienen la variabilidad no explicada ni por la tendencia ni por los componentes estacionales. Podemos ver que tienen un comportamiento de ruido blanco, aunque podemos percibir anomalías puntuales.

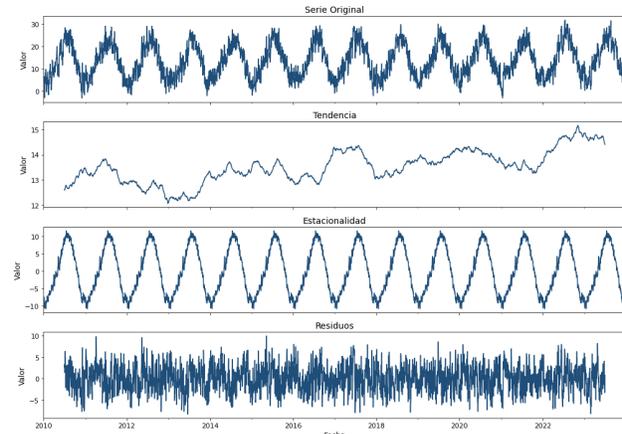


Figura 13: Descomposición Clásica variable Temperatura media

4.2.1 Análisis de la tendencia lineal en la variable temperatura

Para observar si existe una tendencia lineal para la variable temperatura, vamos a llevar a cabo un análisis, en el que ajustaremos distintas rectas a la media móvil de orden 365 para diferentes intervalos en el tiempo. Los datos que se utilizarán han sido obtenidos también del Repositorio de Datos Abiertos de AEMET, pero para un intervalo de tiempo superior al del estudio que estamos abordando. El periodo

total cubre los años de 1974 a 2024.

El objetivo es determinar si existe una tendencia lineal significativa en la temperatura y evaluar si tiene sentido asumir ésta para los datos comprendidos en el periodo en estudio. En primer lugar, tendremos en cuenta la recta ajustada al periodo comprendido entre 1974 a 2006.

• **Periodo 1974 - 2006**

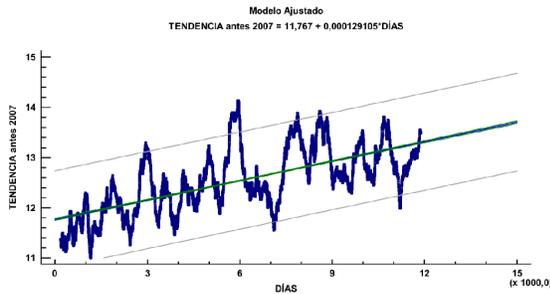


Figura 14: Recta ajustada a la tendencia 1974/2006

- **Recta Ajustada:** $Antes2007 = 11,767 + 0,000129105 \times Dias$

En este primer gráfico tenemos una pendiente de la recta ajustada con un valor de 0.000129105. Aunque se percibe una leve tendencia, ésta no es muy marcada y está comprendida en un periodo largo de tiempo. Esto sugiere que, aunque pudo haber un aumento en la temperatura, su ritmo de crecimiento fue bajo. En el gráfico, también podemos ver una variabilidad notable en los datos de temperatura, con oscilaciones alrededor de la tendencia, reflejando las fluctuaciones de cada año.

• **Periodo 2007 - 2024**

- **Recta Ajustada:** $Despues2007 = 8,56646 + 0,00031499 \times Dias$

En este caso, la pendiente estimada es considerablemente más pronunciada que en el periodo anterior, indicando que el incremento de la temperatura ha sido más acusado durante este periodo. La pendiente obtenida es de 0.00031499. Esta pendiente es aproximadamente el triple que la estimada para el periodo 1974-2006, lo que sugiere un aumento diario de la temperatura tres veces mayor entre 2007-2024.

Para estimar el impacto de este aumento en un año completo, se plantea la siguiente operación sencilla.

$$0,00031499 \times 365 \approx 0,11497^{\circ}C$$

Por tanto, en el periodo estudiado existe un aumento promedio de $0,11497^{\circ}C$ anualmente. Este fenómeno puede estar influenciado por el cambio climático, un aumento general en las emisiones de contaminantes, entre otros sucesos.

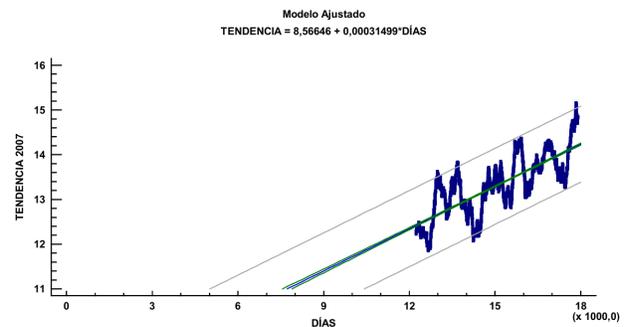


Figura 15: Recta ajustada a la tendencia 2007/2024

• **Periodo 1974 - 2024**

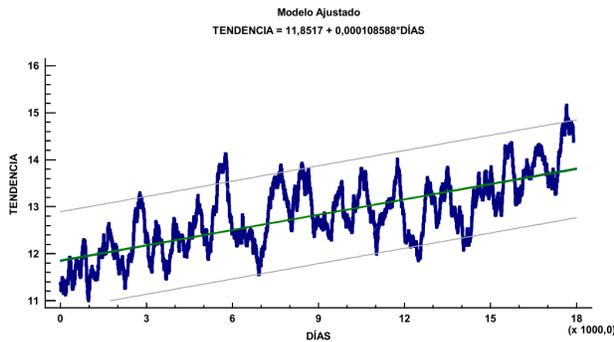


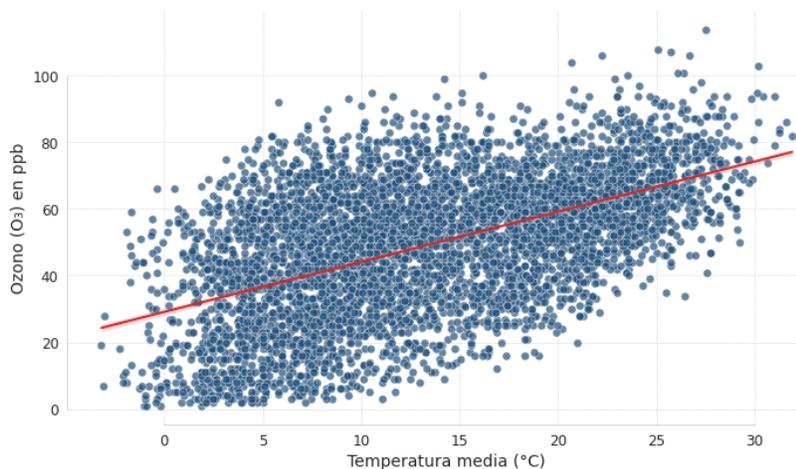
Figura 16: Recta ajustada a la tendencia 1974/2024

- **Recta Ajustada:** $Despues2007 = 11,8517 + 0,000108588 \times Dias$

Por último, la pendiente que ajusta los datos del periodo completo representa una pendiente más suavizada, también debido a la mayor amplitud del periodo seleccionado (pendiente = 0.000108588). Con los datos en el periodo completo vemos que sí existe una tendencia en el aumento de la temperatura y que en promedio, desde 1974 hasta 2024, ha sido de 0,03963°C anualmente.

Como conclusión de este epígrafe diremos que sí existe una tendencia creciente de la temperatura, y que ésta ha sido más acusada en los últimos años. Teniendo en cuenta que esta tendencia va a ser relevante en nuestro estudio, deberemos tenerla en cuenta.

4.3 Relación entre la temperatura media y la concentración de ozono



Recta Regresión Lineal Ajustada:

$$O_3 = 29,23 + 1,5 \times Temp$$

Figura 17: Gráfico de Dispersión O_3 y Temperatura

Para comenzar un análisis conjunto entre el Ozono y la Temperatura media, donde intentaremos comprender mejor la relación entre estas dos variables, tenemos este gráfico de dispersión [Figura 17]. En él, se muestra la relación entre la temperatura media (C) y la concentración de ozono (O_3) en el aire.

La nube de puntos que resulta nos pone de manifiesto la asociación positiva que existe en dicha relación. Es decir, a mayor temperatura media, se asocian mayores concentraciones de gas. Para modelar también

este suceso, se ha ajustado una recta de regresión (Recta roja) a dicha nube de puntos. Este ajuste nos dice que, en un aumento de 1.5°C la Temperatura, el ozono aumenta en promedio una unidad (ppb). Aunque podemos ver que existe cierta dispersión, el modelo lineal recoge bien la variabilidad y tendencia de los datos. Esta variabilidad puede deberse a otros factores atmosféricos o contaminantes no considerados.

También podemos ver que, a mayores valores de la temperatura y el ozono, tenemos una mayor cantidad de valores atípicos (outliers), lo que nos sugiere una menor correlación entre ambas a valores altos de temperatura o, la posible existencia de agentes externos a este ajuste que estén influyendo en la concentración del ozono.

4.3.1 Comparación de índices diarios de Ozono y Temperatura

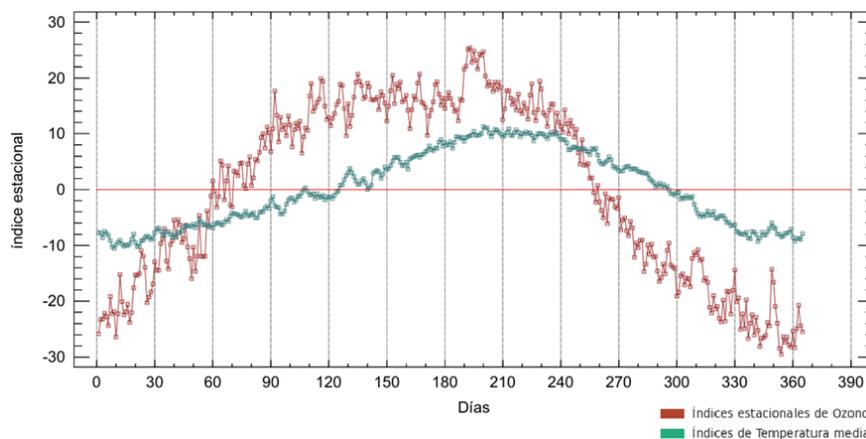


Figura 18: Descomposición Estacional Conjunta

A partir de una representación conjunta de los índices estacionales, comprendemos mejor la relación existente. Ambos conjuntos de índices muestran un incremento durante los meses centrales del año, mientras que en los extremos se suavizan.

Además, vemos como la variación total del ozono (O_3) es más pronunciada que la de la temperatura. Es decir, la diferencia entre los extremos y el tramo central de índices estacionales es mayor para el ozono. Como vemos en la Figura 18 en periodos de baja temperatura, la reducción en los niveles de ozono es más notable que la de la temperatura. Esto se debe a una relación no lineal entre ambas variables, donde el ozono es sensible a cambios en la temperatura.

Observamos dos cortes entre ambos conjuntos de índices estacionales. La primera intersección es más difusa, y ocurre en torno al comienzo del año, en el mes de Febrero. Prácticamente durante este mes entero sucede un cruce entre índices que se extiende hasta el mes de Marzo, momento en el que existen concentraciones de ozono mucho más elevadas. Los índices que incurrn en esta primera intersección tienen valores en un intervalo $[-6; -8]$.

El segundo cruce se efectúa al final del mes de Agosto y el inicio del mes de Septiembre, entre Los índices 249-251, con valores en torno a $[6.5; 8.8]$. Los días del año a los que corresponden estos índices están entre el 31 de Agosto y el 2 de Septiembre. En este momento del año, el descenso del ozono en el aire empieza a ser más acusado que el descenso generalizado de la temperatura.

4.4 Serie diaria Dióxido de Nitrógeno (NO_2)

En esta sección se hará un análisis descriptivo de la variable dióxido de nitrógeno, ya que puede tener interés dentro del estudio que estamos abordando, como hemos visto en el apartado 4.1.2. Estos datos han sido extraídos de la estación meteorológica Puente Poniente - M^a Luisa Sánchez y contienen observaciones diarias comprendidas entre Enero de 2010 a Diciembre 2023.

El dióxido de nitrógeno (NO_2) es un gas de color marrón-rojizo y altamente reactivo que se forma principalmente a partir de la oxidación del monóxido de nitrógeno (NO) en la atmósfera, especialmente durante procesos de combustión a altas temperaturas, como en motores y plantas industriales. Este compuesto es uno de los principales contaminantes atmosféricos y un precursor clave en la titulación del ozono troposférico, afectando gravemente la calidad del aire y a la salud humana. [13]

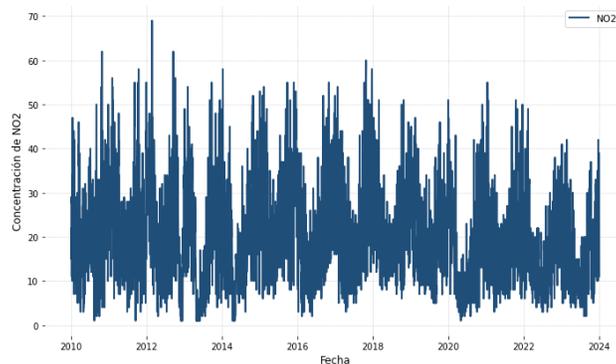


Figura 19: Serie Original Dióxido de Nitrógeno (NO_2)

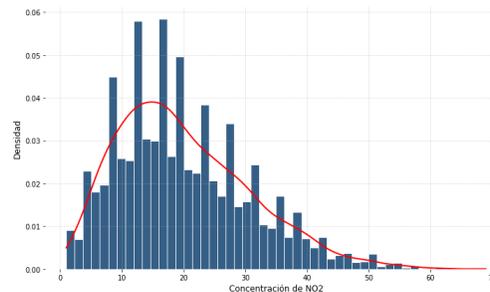
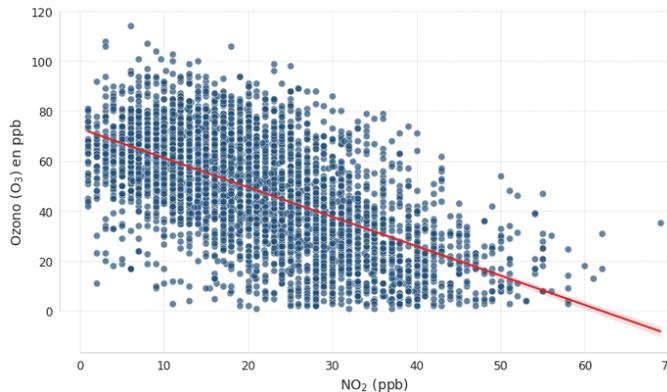


Figura 20: Distribución Original Dióxido de Nitrógeno (NO_2)

En el gráfico de la Figura 20 tenemos, a la izquierda, una representación de la serie temporal del NO_2 . En ella, se aprecia una posible estacionalidad anual, aunque con un patrón algo confuso a simple vista. A la derecha, la distribución de la concentración de los valores de NO_2 . A la vista de esta distribución no se sigue estrictamente una distribución normal. Se observan algunos valores con frecuencias mayores que se alejan significativamente de la media, lo que sugiere la presencia de eventos extremos. Sin embargo, la forma general de la distribución no presenta desviaciones fuertes que justifiquen la aplicación de transformaciones.

4.4.1 Relación entre el NO_2 y el O_3



Existe una relación compleja y negativa entre el ozono troposférico (O_3) y el dióxido de nitrógeno (NO_2). Aunque el NO_2 es un precursor en la formación de ozono mediante reacciones fotoquímicas, en condiciones con altas concentraciones de NO_2 (como en zonas urbanas con tráfico intenso) puede observarse una disminución del ozono. Esto se debe a que el exceso de NO_2 está asociado a un entorno donde el O_3 se forma menos eficientemente o se destruye más rápidamente, lo que genera una correlación negativa entre ambos contaminantes.

Figura 21: Gráfico de Dispersión NO_2 frente a O_3

4.5 Serie diaria Precipitación

Los datos de Precipitación han sido tomados del Repositorio de Datos Abiertos de AEMET para la evaluación de la Precipitación. La unidad de medida en la que se han sido recogidos los datos es en l/m^2 . A continuación, una representación para el periodo en estudio.

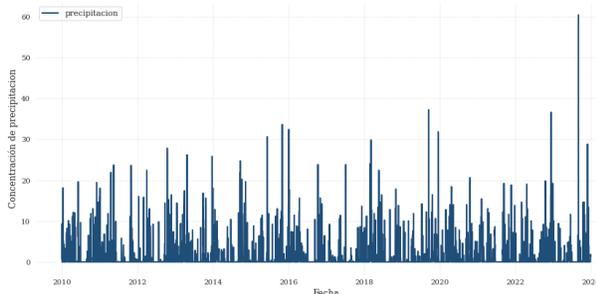


Figura 22: Serie temporal original de la precipitación

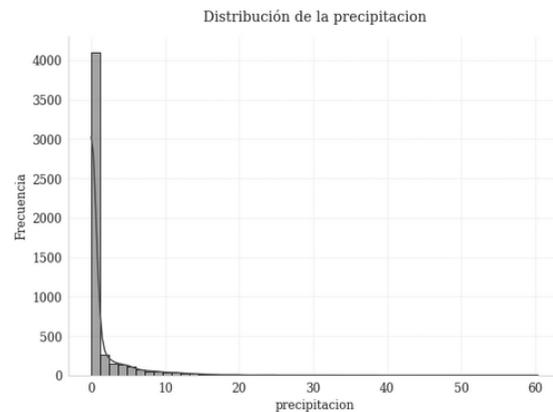


Figura 23: Distribución de la precipitación l/m^2

Existe una distribución muy asimétrica, con una gran concentración de valores en cero $0l/m^2$ y algunos valores extremos positivos. Esta característica que obtenemos al estudiar variables como la precipitación, puede dificultar un papel de predictor en modelos lineales o dinámicos, debido a problemas de no linealidad, heterocedasticidad o influencia desproporcionada de valores atípicos. Para mitigar estos efectos, se va a aplicar la transformación **inflación por ceros**, que es una transformación logarítmica a la que añadimos un offset c (en este caso $c = 1$):

$$\text{Log Precipitación} = \log(\text{precipitación} + 1)$$

En adelante, la variable **LogPrecipitación** hará referencia a la transformación $\log(\text{Precipitación} + 1)$.

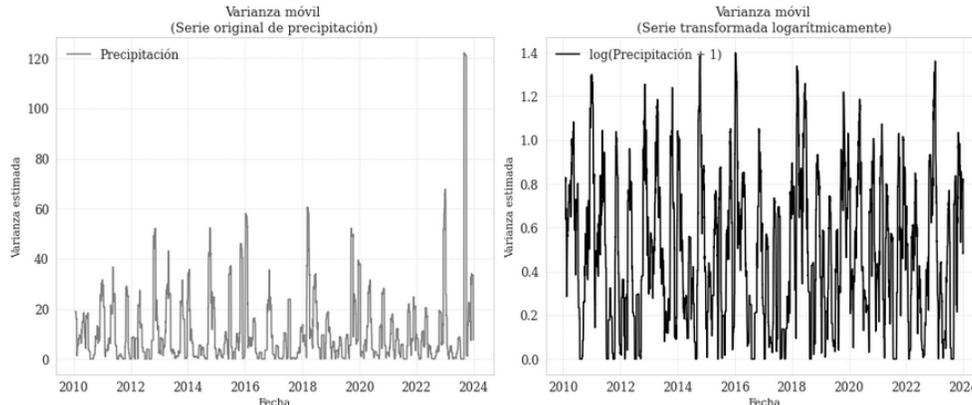


Figura 24: Evolución de la varianza móvil en la serie de precipitación

En el gráfico de la Figura 24 vemos cómo la heterocedasticidad es acusada en la variable sin transformar, además de tener valores más altos y, por tanto, más extremos. Esto evidencia problemas de varianza no constante. Al aplicar la inflación por ceros, se aprecia una estabilización significativa de la varianza. Esta transformación supondrá un efecto suavizante, contribuyendo a la reducción de la varianza y a la homogeneidad de la serie.

Cuadro 3: Resultados del test de homocedasticidad de Breusch-Pagan

Variable	Estadístico	p-valor
Precipitación original	0.7393	0.3899
Precipitación transformada (log + 1)	0.0485	0.8256

Cuadro 4: Asimetría y kurtosis

Variable	Asimetría	Kurtosis
Precipitación original	4.8748	38.2827
Precipitación transformada (log + 1)	1.8984	2.6174

En la **Tabla 3** se muestran los resultados del test de Breusch-Pagan para evaluar la homocedasticidad en la serie original de precipitación y su versión transformada mediante logaritmo con un offset de 1. Vemos como la transformación con logaritmo mejora notablemente el estadístico y aumenta el p-valor, estabilizando la varianza. Por otro lado, la **Tabla 4** presenta las medidas de asimetría y kurtosis. La serie original tiene una asimetría positiva pronunciada (4.87) y una kurtosis extremadamente alta (38.28). Tras la transformación logarítmica, tanto la asimetría como la kurtosis se reducen considerablemente (1.90 y 2.62, respectivamente).

Por tanto, transformar la serie de precipitación mediante inflación de ceros ha permitido estabilizar la varianza y reducir la fuerte asimetría y kurtosis presentes en los datos originales.

5 Análisis univariante

En esta sección se aplicará la metodología de Box-Jenkins [Marco teórico Box-Jenkins: 2.1] para identificar los modelos ARIMA estacionales que modelizan las series diarias de temperatura y ozono (O_3). Para ello, se llevará a cabo un análisis detallado de cada serie, incluyendo la identificación de estacionariedad, la selección de los parámetros adecuados y la validación del modelo resultante. Esta aproximación permitirá comprender la evolución de ambas variables a lo largo del tiempo y realizar posibles predicciones basadas en los patrones observados. Los datos que se van a utilizar para la estimación del modelo son:

- Datos de Enero de 2010 a Diciembre de 2023 para estimar el modelo ARIMA; 4748 observaciones.
- Datos de Enero de 2024 a Diciembre de 2024 (un año) para testear los modelos ARIMA ajustados; 365 observaciones.

5.1 Modelo ARIMA para datos diarios del contaminante O_3

Cuando analizamos una serie temporal, es fundamental determinar si es estacionaria. Para esto, es preciso identificar si se requiere una transformación de Box-Cox que estabilice la serie, en caso de que la varianza no sea constante. A la vista de la figura [5], no parece que sea necesario realizar una transformación a los datos.

Para poder estudiar y ajustar un modelo ARIMA de Box-Jenkins a los datos, tendremos en cuenta la ACF y PACF de la serie, junto al periodograma. La ACF (Función de Autocorrelación) mide la correlación entre la serie temporal y sus retardos a diferentes intervalos de tiempo. Ayuda a identificar patrones y la dependencia temporal en los datos. Por su parte, la PACF (Función de Autocorrelación Parcial) mide la autocorrelación directa entre la serie y su retardo, eliminando el efecto de la serie en los retardos intermedios.

Las figuras [25] nos ayudarán a decidir el orden de diferenciación, para que la serie sea estacionaria.

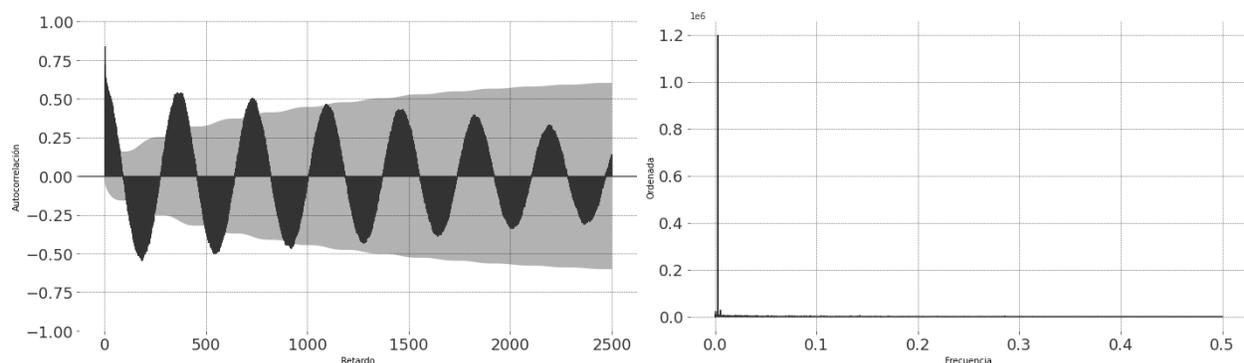


Figura 25: ACF y PACF de la serie original del ozono

En la ACF podemos notar que las autocorrelaciones forman ondas cuyos máximos y mínimos, en retardos múltiplos de 365, decrecen lentamente. Además, el armónico $1/365$ del periodograma es el

más significativo, ocultando la posible contribución del resto de los armónicos a la explicación de la variabilidad de la serie. Esto nos indica que la serie no es estacionaria en su componente estacional, lo que implica la necesidad de aplicar una diferenciación estacional. Dado que el período estacional es de 365 días, se realizará una diferenciación restando los valores del mismo día en el año anterior.

$$\Delta_{365}x_t = x_t - x_{t-365}$$

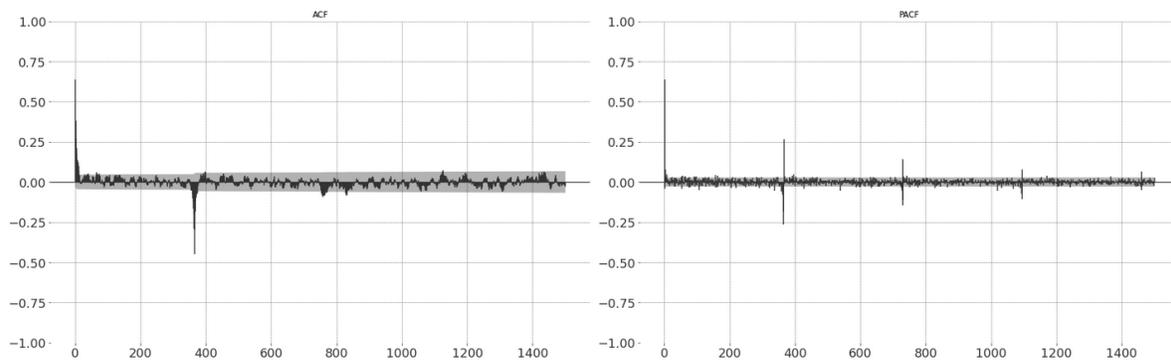


Figura 26: ACF y PACF Serie Diferenciada con 1500 lags

Como se puede ver en la nueva ACF y PACF de la serie diferenciada de manera estacional, se han eliminado las oscilaciones, y ahora contamos con un decrecimiento exponencial de las correlaciones parciales en los retardos múltiplos del periodo.

Haremos Zoom sobre los primeros retardos de dichas ACF y PACF para observar el comportamiento que tiene en ellos las autocorrelaciones de la serie diferenciada estacionalmente y estudiar si fuera necesario una diferenciación regular.

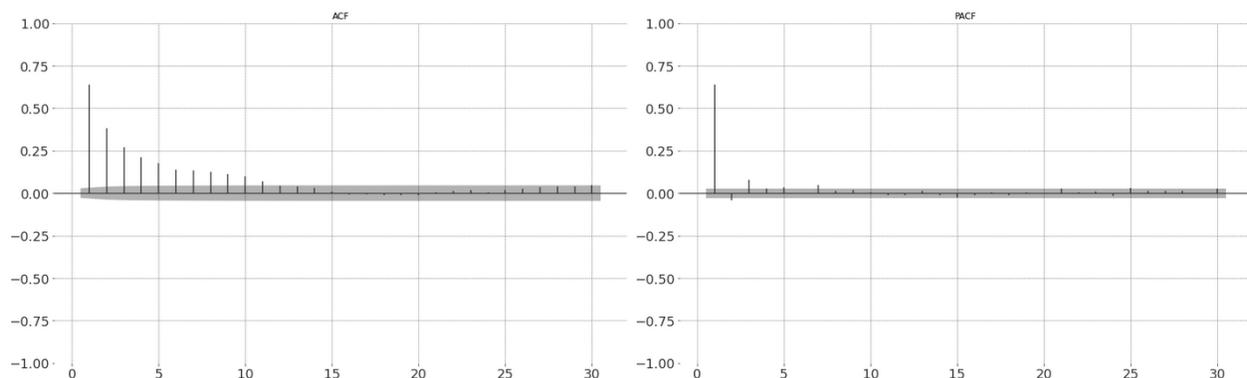


Figura 27: ACF y PACF Serie Diferenciada

Por la parte regular tenemos un decrecimiento exponencial de la ACF por lo que no sería necesario

realizar una diferenciación regular.

	Periodo 365	Periodo 7
Ordenada	1,28393E6	4609,86

Cuadro 5: Tabla Ordenadas Periodograma Ozono

También se ha tenido en cuenta la tabla del periodograma, en la cuál se observan las ordenadas en cada periodo. En la Tabla 3, podemos ver cuál es la ordenada del periodo 365 y del periodo 7. Por su parte, el periodo 365 cuenta con una ordenada de 1, 28393E6 previa diferenciación. El periodo 7 tiene una primera ordenada de valor 4609, 86, y su segunda ordenada de periodo con 3331, 86. Esto confirma nuestra sospecha de que el periodo 7 también cobra importancia en el estudio del Ozono. De este modo, comenzaremos ajustando un modelo ARIMA con doble estacionalidad (7 y 365).

	Sin dif.	dif. 365	dif. 7 y 365
Desv. Estándar	20,43423	18,99253	25.01809

Cuadro 6: Tabla Desviación Estándar del Ozono

Teniendo en cuenta la desviación estándar en los distintos niveles de diferenciación, concluimos que la serie diferenciada una vez estacionalmente es la mejor opción, ofreciéndonos también un descenso en la desviación típica de la serie respecto a la que no lo está. Por otro lado, la serie con una diferenciación estacional de periodo 7 y otra de 365, aumenta la desviación típica, por lo que no parece conveniente. Finalmente, los índices de diferenciación serán $d = 0$, $D_7 = 0$ y $D_{365} = 1$.

Una vez tenemos claros los órdenes de diferenciación de la serie, es momento de tener en cuenta la ACF y PACF que han resultado para determinar los parámetros del modelo ARIMA. Por la parte regular, en la [Figura 27] vemos que a partir del primer retardo la ACF decrece exponencialmente. En la PACF se puede considerar al menos los tres primeros retardos salen de la banda de confianza. Por esto mismo, la parte regular podría comportarse como un $AR(3)$. Por la parte estacional tenemos un retardo del periodo significativo en la ACF, mientras que la PACF desciende exponencialmente en cada uno de los retardos del periodo 365. Un modelo multiplicativo válido sería el modelo de media móvil $MA(1)$.

Como vemos en la imagen del Anexo [**Figura Anexo I**] el séptimo retardo sobresale de la banda de confianza (junto con algún retardo posterior múltiplo de 7), por lo que deberá de tenerse en cuenta en el modelo ARIMA univariante que se ajuste.

En el proceso de ajuste del modelo ARIMA, se ha observado que algunos parámetros estimados no resultan estadísticamente significativos, presentando p-valores elevados. La presencia de patrones estacionales de distinta magnitud puede generar colinealidad, afectando las estimaciones. El ruido en la serie puede dificultar la identificación de relaciones significativas.

Debido a esta situación, es necesario ajustar varios modelos con algunas variaciones para garantizar que la estimación de los coeficientes sean significativos. Por tanto, los polinomios Backward de los modelos SARIMA estimados en esta primera etapa son:

- **Modelo 1:** $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_7 B^7)(1 - B^{365})X_t - \delta = (1 - \Theta_{365} B^{365})a_t$
- **Modelo 2:** $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B^{365})X_t - \delta = (1 - \Theta_7 B^7)(1 - \Theta_{365} B^{365})a_t$
- **Modelo 3:** $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_7 B^7)(1 - B^{365})X_t - \delta = (1 - \theta_{13} B^{13} - \theta_{365} B^{365})a_t$

- **Modelo 4:** $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_7 B^7 - \phi_{365} B^{365})(1 - B^{365})X_t - \delta$
 $= (1 - \theta_{14} B^{14} - \theta_{21} B^{21} - \theta_{28} B^{28} - \theta_{365} B^{365} - \theta_{730} B^{730})a_t$

[Referencia al Ajuste de Modelos Anexo I]

A la vista de los ajustes de los modelos planteados, los cuatro cuentan con parámetros significativos. Todos llevan a cabo un buen ajuste del modelo. Como los AIC de cada uno de los modelos son bastante similares, no nos podemos permitir descartar ninguno de ellos en esta primera etapa.

Por su parte, las matrices de correlación de cada ajuste [**Figura Anexo I**] no contienen coeficientes suficientemente altos como para rechazar el modelo, por lo que los modelos considerados no contienen sobreajuste.

Ahora, debemos de tener en cuenta si los residuales de cada uno de los modelos se comportan más o menos como ruido blanco, de manera que permita al modelo ser válido. [**Figura Anexo I**]

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
p-valor	0.3579	0.1665	0.3985	0.4555

Cuadro 7: P-valores del Test de Ljung-Box para el Ozono

Dentro de la ACF de residuales de cada uno de los modelos no parece que haya ningún retardo con una especial significancia. En la tabla [7] se tienen los p-valores para el test de **Ljung-Box** que nos indica si los residuales se comportan como un ruido blanco. La hipótesis nula que subyace en el test de Ljung-Box es;

H_0 : No hay autocorrelación significativa en la serie temporal.

H_1 : Existe al menos una autocorrelación distinta de cero hasta el retardo h .

A la vista de estos p-valores, vemos que no se rechaza la hipótesis nula del test, por lo que podemos asumir que los residuos de los cuatro modelos se comportan como un ruido blanco.

Para comparar la capacidad de predicción de los diferentes modelos, se usará la **suma de cuadrados del error (SSE)**. Para ello, se emplearán las predicciones correspondientes a un año (Enero 2024 a Diciembre 2024), el cuál ha sido reservado para analizar la capacidad de predicción.

El siguiente gráfico tiene la representación del ajuste de cada uno de los cuatro modelos al año 2024, junto a sus bandas de confianza. En [**Anexo I**] tenemos una Figura con la ampliación sobre las últimas 30 observaciones predichas, para poder ver mejor la predicción. Los puntos representan los datos reales de concentración de O_3 para dicho año.

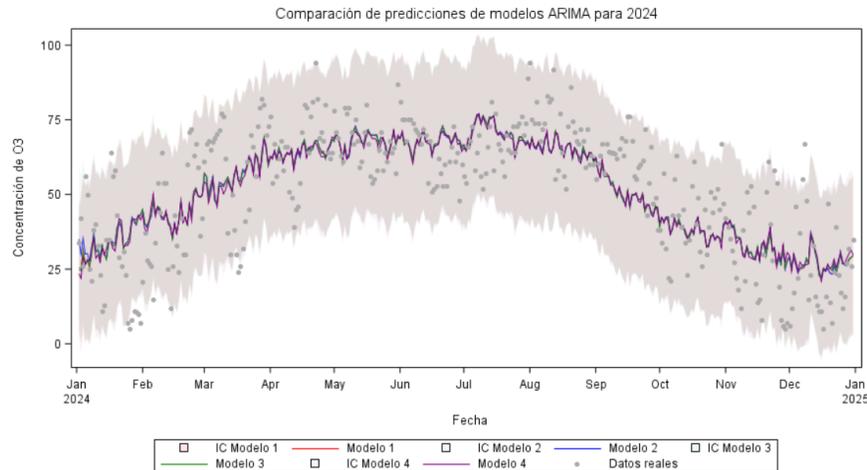


Figura 28: Datos Reales vs. Predicciones Nuevo Año

La representación de la predicción de nuevos datos no nos permite detectar a simple vista cuál de los cuatro modelos ajusta mejor. Por ello, debemos obtener el SSEp de cada modelo. En la siguiente tabla tenemos un resumen de tres criterios de evaluación del ajuste, para poder comparar los diferentes modelos:

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
AIC Hasta 2023	36742.81	36751.8	36737.64	36733.43
SSEp	73295.5131	73168.0192	74052.2271	74342.0685
SBC Hasta 2024	39581.45	39588.5	39586.28	39607.73

Cuadro 8: Tabla de medidas de ajuste y selección para los modelos ARIMA Ozono

Podemos sacar varias conclusiones sobre las anteriores medidas de evaluación. En primer lugar, vemos que el **modelo 4** presenta el **mejor valor de AIC (36733.43)**, sugiriendo un buen equilibrio entre el ajuste y la complejidad del modelo 4 hasta 2023. Por otro lado, el **modelo 2** destaca por tener la **menor Suma de Errores Cuadráticos de predicción (SSEp = 73168.02)**, lo que indica que es el mejor modelo en cuanto a su capacidad de predicción. Aunque tenemos que escoger uno, todos los modelos tienen valores de ambos criterios de evaluación bastante similares.

Por último, se ha obtenido el **SBC Hasta 2024**, que incorpora los datos hasta Diciembre de 2024 para sacar esta medida de ajuste. Como podemos ver, el **modelo 4** pasa a ser el **peor (SBC = 39607.73)**, lo que refuerza que su valor de AIC podría deberse a que se ha dado un **sobreajuste** en este modelo. Esta medida también pone de manifiesto como los datos del año 2024 tienen un comportamiento especial respecto al resto de la muestra.

Por tanto, al considerar tanto la capacidad predictiva como la sencillez del modelo, determinamos que el mejor modelo para la predicción de datos de O_3 será el **Modelo 2: $ARIMA(3, 0, 0)(0, 0, 1)_7(0, 1, 1)_{365}$** .

$$(1 - 0,65445B + 0,10621B^2 - 0,06987B^3)(1 - B^{365})X_t - 0,1112 = (1 + 0,04723B^7)(1 - 0,96604B^{365})a_t$$

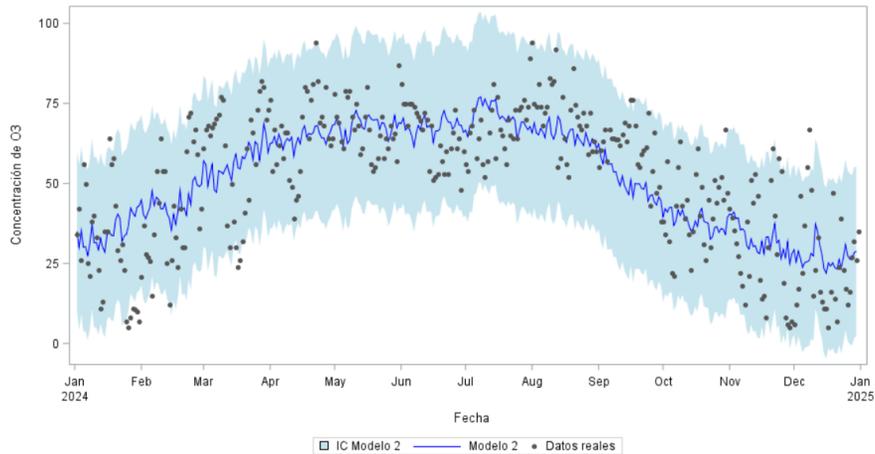


Figura 29: Predicción para O3 2024. Modelo 2

La predicción de los nuevos datos con el **modelo 2** está representada en el gráfico de la Figura [29]. En él, se ve la predicción con dicho modelo, los valores reales para el año 2024 y los intervalos de confianza de la predicción. Prácticamente todos los datos se mantienen dentro de este IC del 95 %, por lo que estamos ante una buena predicción con el modelo que hemos elegido.

Además, hay que destacar que el modelo ajustado contiene la constante del modelo significativa, indicando que el ozono tiene una parte de tendencia determinista.

Del modelo SARIMA $(3, 0, 0)(0, 1, 1)_{365}$ cabe destacar que, al tener una diferenciación estacional e incluir una constante significativa en el modelo, esto implica que la temperatura tendría una parte de tendencia determinista de la forma $T_t = a + bt$, (como incremento diario) donde la pendiente del incremento viene dada por:

$$\text{Constante de estimación} = 0,1112$$

$$b = \frac{\text{Media de la serie diferenciada}}{365} = \frac{0,29109}{365} = \frac{\text{Constante del modelo}}{(1 - \phi_1 - \phi_2 - \phi_3) \cdot 365} = 7,975 \cdot 10^{-4}$$

Al tener una pendiente positiva, nos indica que el contaminante ozono puede continuar experimentando un aumento anual de **0,29109 (ppb)**. Esto apoya la teoría de la existencia de un aumento generalizado en este tipo de contaminantes [7].

5.2 Modelo ARIMA para datos diarios de Temperatura Media

Para ajustar un modelo univariante a la serie de la variable Temperatura Media, nos fijamos en la [**Figura Anexo I**], viendo que no es necesaria ninguna transformación al conjunto de datos.

Para ajustar el modelo ARIMA, veremos en primer lugar, cuál es la diferenciación necesaria que se debe de realizar a la serie para que esta sea estacionaria. De nuevo, tomaremos la ACF y periodograma de la serie para determinar a qué niveles diferenciaremos, tanto regular d como estacional D .

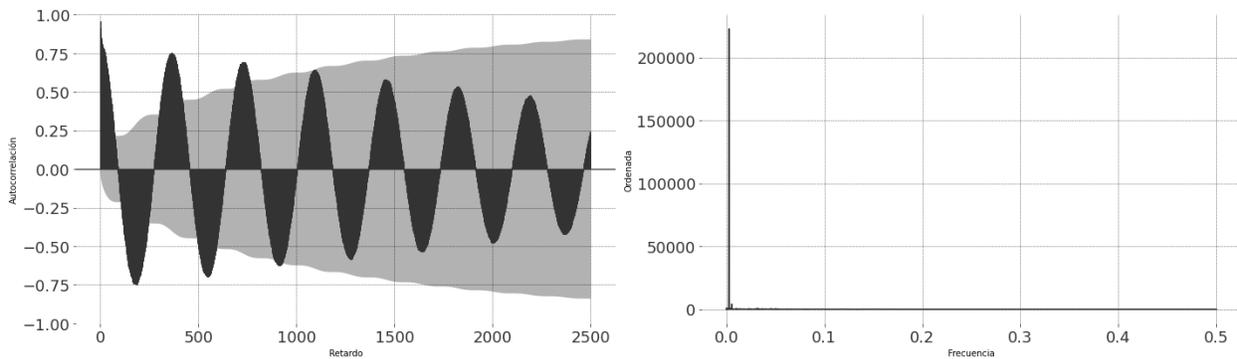


Figura 30: ACF y PACF de la serie original de la temperatura

La ACF de la Temperatura nos indica que las autocorrelaciones de periodicidad anual decrecen lentamente, mientras que el periodograma contiene el armónico del periodo (1/365) muy significativo. Por tanto, será necesario diferenciar de manera estacional para que la serie sea estacionaria.

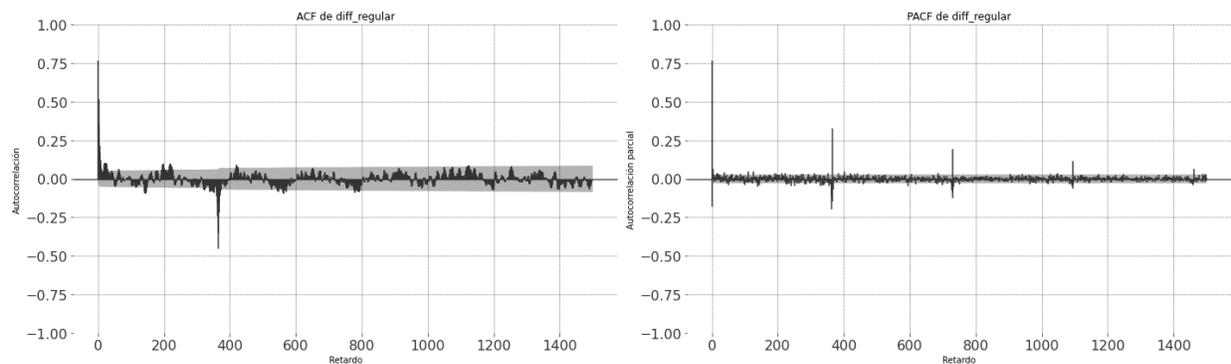


Figura 31: ACF y PACF de la serie diferenciada hasta el retardo 900

Con esta diferenciación estacional, vemos como tanto la ACF como la PACF no presentan decrecimiento lento por la parte estacional. Además, como vimos en el Análisis Descriptivo de la serie, existe una tendencia lineal que debemos de corregir a la hora de ajustar este modelo. Por tanto, ahora tomaremos también una diferenciación regular de la serie.

La desviación típica que se obtiene para cada una de las posibles diferenciaciones apoya la decisión que hemos tomado. Con esto, nos quedaremos con los órdenes de diferenciación $D_{365} = 1$, $d = 1$, con una desviación estándar de 3,10575.

	Sin dif.	dif. 365	dif. 1 y 365
Desv. Estándar	7,350645	4,470259	3,10575

Cuadro 9: Tabla Desviación Estándar de Temperatura

A continuación, presentaremos tanto la ACF y PACF de la serie diferenciada, ampliando sobre los primeros retardos.

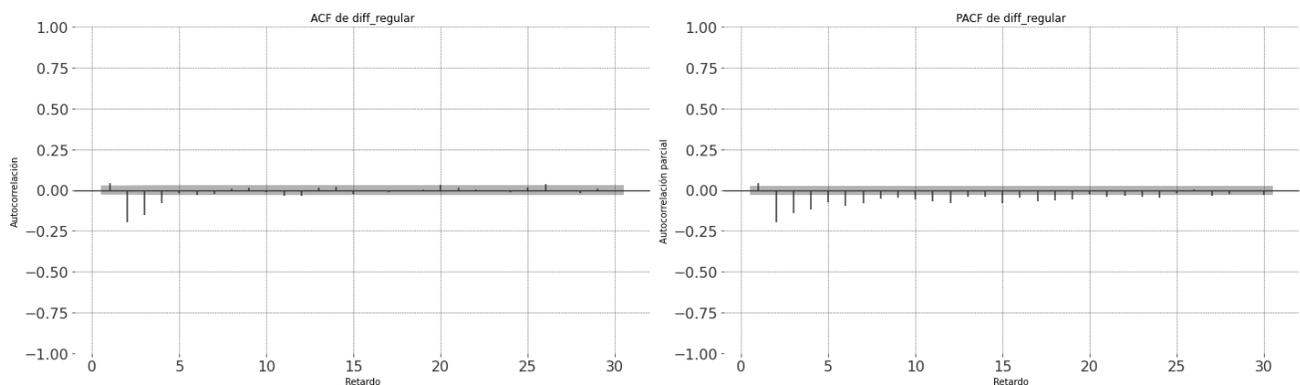


Figura 32: ACF y PACF de la serie diferenciada hasta el retardo 50

Una vez tenemos el orden de diferenciación que aplicamos a la serie, buscaremos cuáles son los valores óptimos para el orden de los polinomios autorregresivos y de media móvil del modelo ARIMA estacional.

Se han considerado aceptables los siguientes modelos para la serie:

- **Modelo 1:** $(1 - \phi_1 B)(1 - B^{365})(1 - B^1)X_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_3 B^3 - \theta_{365} B^{365})a_t$
- **Modelo 2:** $(1 - \phi_1 B)(1 - B^{365})(1 - B^1)X_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_3 B^3 - \theta_{365} B^{365} - \theta_{730} B^{730})a_t$
- **Modelo 3:** $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B^{365})(1 - B^1)X_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_{365} B^{365} - \theta_{730} B^{730})a_t$

[Referencia al Ajuste de Modelos Anexo I]

Los tres modelos contienen todos sus parámetros significativos, a excepción del primer parámetro del *Modelo 2*, con un nivel de significación de 0.11; no obstante, se ha decidido considerar el modelo.

Teniendo en cuenta el criterio AIC, vemos como el *Modelo 3* tiene un valor más elevado que los otros dos modelos. Esta diferencia, aunque no es despreciable, no parece suficiente como para desestimar este modelo.

	Modelo 1	Modelo 2	Modelo 4
p-valor	0.4037	0.2337	<.0001

Cuadro 10: P-valores del Test de Ljung-Box de Temperatura

[Referencia a las ACF de Residuales Anexo I]

Los dos primeros modelos tienen unos residuos que podrían considerarse ruido blanco. Por otro lado, en el *Modelo 3* vemos que sí existe correlación entre los residuos, que resultan de un modelo que no ha ajustado bien a los datos. Apoyándonos en los Test de L-Jung Box [Tabla 7], desecharemos el Modelo 4.

Por último, compararemos el criterio de ajuste AIC, junto con el error predictivo **SSEp** de los dos modelos con los que nos hemos quedado, para determinar cuál es el que mejor ajusta la serie.

	Modelo 1	Modelo 2
AIC	20858.93	20881.38
SSEp	4860.33	5018.84

Cuadro 11: Tabla de medidas de ajuste y selección para los modelos ARIMA Temperatura

A partir de estos criterios, queda escoger entre los Modelos 1 y 2. Vemos como el *Modelo 1* cuenta con mejor criterio AIC, además de un SSEp mucho mejor que el Modelo 2. Por tanto, nos quedamos con el **Modelo 1**. Este modelo es un modelo no multiplicativo con un polinomio Backward que se escribe

$$(1 - 0,66987B)(1 - B^{365})(1 - B^1)X_t = (1 - 0,74697B - 0,24921B^2)(1 - 0,02657B^3 - 0,97343B^{365})a_t$$

Para el ajuste de este modelo univariante de Box-Jenkins a la serie de temperatura media, se ha utilizado los datos hasta el año 2023. Aunque el modelo ajustado con los datos hasta 2023 ha presentado una estructura algo más compleja que el obtenido al incluir todo el periodo hasta 2024, los resultados muestran que ofrece un buen desempeño predictivo. Esto sugiere que, pese a su mayor complejidad, el modelo es capaz de capturar adecuadamente la dinámica de la serie y generalizar bien sobre datos no utilizados en la estimación.

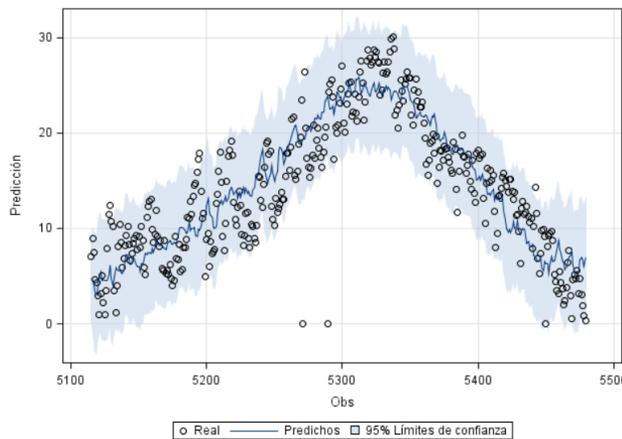


Figura 33: Predicción para Temperatura 2024. Modelo 2

Este gráfico muestra la predicción de las siguientes 365 muestras (año 2024). En negro están representados los datos reales, mientras que la línea azul representa el modelo ajustado. Las bandas azules corresponden a un intervalo de confianza del 95%. Prácticamente todos los datos, salvo algunos outliers, quedan dentro del intervalo, lo que indica que el modelo seleccionado proporciona una buena predicción para la serie de temperatura.

5.3 Modelo ARIMA para datos diarios de Log Precipitación

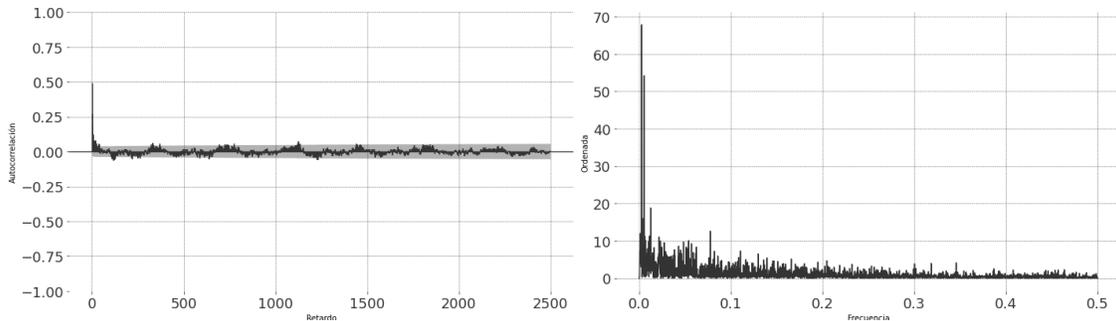


Figura 34: ACF y Periodograma Log Precipitación

Para comenzar el modelo ARIMA veremos si es necesaria alguna diferenciación sobre la serie. En los gráficos de la ACF y periodograma de la serie, vemos como los armónicos del periodo anual pueden resultar significativos, mientras que en la ACF parece no llegar a rozar la significancia el periodo. Por tanto, vamos a tener en cuenta como afecta la diferenciación estacional a la desviación típica de la serie.

	Sin dif.	dif. 365
Dev. Estándar	0,745851	1,051057

Cuadro 12: Tabla Desviación Estándar de Log Precipitación

Ni la desviación típica ni la ACF nos muestran que sea necesaria la diferenciación en la log precipitación. La ACF y PACF de la serie original son:

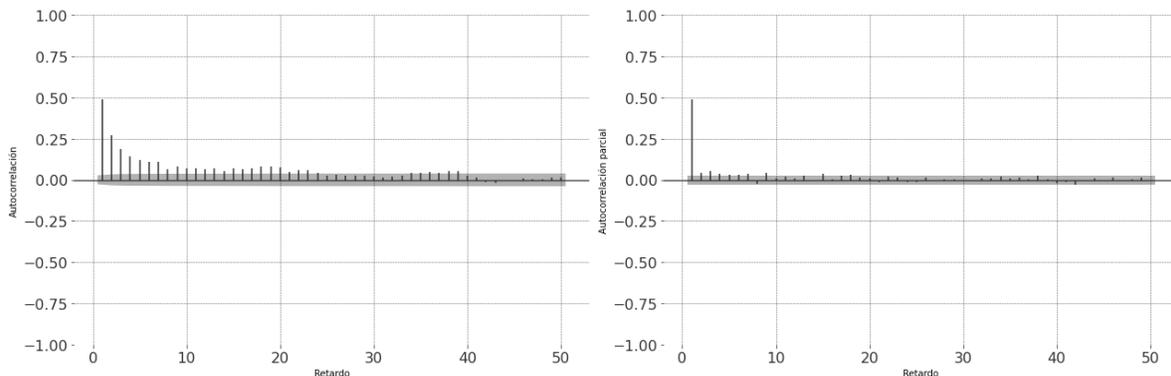


Figura 35: ACF y PACF con 50 lags

La estacionalidad anual, como hemos visto, es importante para explicar la precipitación. La estimación del modelo y sus residuos;

Estimación de probabilidad máxima					
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo
<i>MU</i>	0.39929	0.02807	14.23	<.0001	0
<i>MA1,1</i>	0.91091	0.02265	40.22	<.0001	1
<i>MA2,1</i>	0.67021	0.33147	2.02	0.0432	365
<i>AR1,1</i>	1.36807	0.02804	48.78	<.0001	1
<i>AR1,2</i>	-0.39540	0.02062	-19.18	<.0001	2
<i>AR2,1</i>	0.64946	0.33835	1.92	0.0549	365
Constante Estimación		0.003824			
Varianza Estimación		0.418752			
Error Std Estimación		0.647111			
AIC		10066.17			
SBC		10105.4			
Número de residuales		5113			

Figura 36: Ajuste del Modelo Log Precipitación

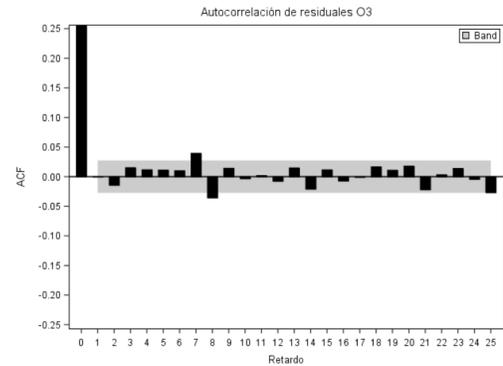


Figura 37: ACF residual modelo ARIMA Log Precipitación

Los residuos no son significativamente diferentes de ruido blanco a simple vista, a excepción de las autocorrelaciones en los retardos 7 y 8, relacionadas con el periodo semanal. A partir del test de Ljung-Box sobre los residuos del modelo, tenemos un p-valor de 0.2017 para un retardo de 36. Aplicando el test sobre un retardo de periodo 12, nos encontramos con un p-valor de 0.0186, fuera del 5 % de significancia. Esto indica que podría haber una periodicidad semanal subyacente en esta serie, reflejándose en estos retardos.

El modelo ARIMA univariante en forma de polinomio Backward es;

$$(1 - 1,368B + 0,395B^2)(1 - 0,649B^{365})X_t - 0,003824 = (1 - 0,911B^1)(1 - 0,67B^{365})\varepsilon_t \quad (25)$$

Para los criterios de medida del ajuste considerados en este estudio, se obtienen los siguientes valores: $AIC = 10066,17$ y $SSE_p = 205,14$.

El modelo ajustado ha demostrado ser válido para los datos analizados. Sin embargo, como se ha mencionado, es posible que exista un **efecto semanal subyacente en las precipitaciones**. Estudios previos han mostrado la presencia de patrones semanales en la precipitación, los cuales se atribuyen a la influencia de contaminantes atmosféricos que pueden afectar la formación y condensación de nubes [19, 20].

Aunque esta relación no está evidenciada por completo en todas las regiones, la existencia de estos patrones en áreas contaminadas de Europa sugiere que fenómenos similares podrían observarse en zonas geográficas como Valladolid, donde la actividad industrial y la contaminación atmosférica podrían influir en la precipitación.

5.4 Modelo ARIMA para datos diarios de NO_2

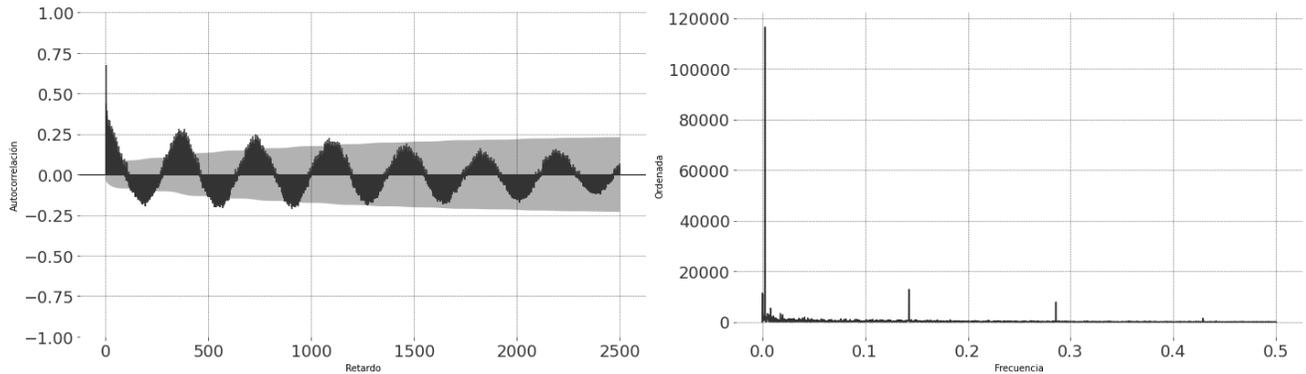


Figura 38: ACF y Periodograma NO_2

A partir de la ACF y el periodograma de la serie original, se observa una clara necesidad de aplicar una diferenciación estacional anual, ya que la ACF presenta oscilaciones en torno a ese periodo y el periodograma muestra armónicos significativos en la frecuencia $1/365$. Tras aplicar esta diferenciación estacional, se ha hecho evidente la presencia de una tendencia determinista (periodograma serie diferenciada), lo cual sugiere que también es necesaria una diferenciación regular. Asimismo, se identifican armónicos de periodo 7 en el periodograma, lo que indica una componente semanal. Para decidir la diferenciación más adecuada, se compara la desviación estándar de varias combinaciones de diferenciaciones:

	dif. 365	dif. 1 y 365	dif. 7 y 365
Desv. Estándar	13,09419	12,13645	16,81592

Cuadro 13: Tabla Desviación Estándar del NO_2

Se opta por aplicar una doble diferenciación (regular y estacional) al comprobar que reduce la desviación estándar y elimina la tendencia. La ACF y PACF de la serie diferenciada se muestran a continuación:

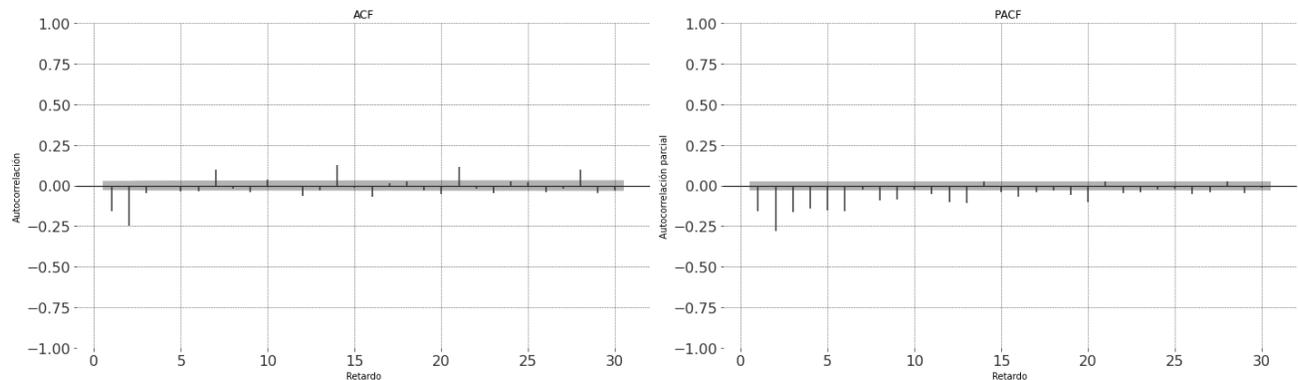


Figura 39: ACF y PACF NO_2 Diferenciado 1 y 365

Analizando la ACF y PACF de la serie diferenciada, así como las funciones de autocorrelación con hasta 900 retardos (donde se observa significancia en los periodos estacionales), se decide ajustar un modelo $ARIMA(2, 1, 1)(1, 1, 1)_{365}(1, 0, 1)_7$. Este modelo incorpora tanto la estacionalidad anual como la semanal, reflejando la doble estacionalidad presente en los datos.

Los parámetros estimados del modelo se muestran a continuación, donde vemos que existe una fuerte influencia del periodo semanal en el modelo. Al lado, el gráfico de la ACF de los residuos, que verifica que estos se comportan como ruido blanco:

Estimación de probabilidad máxima					
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo
MA1,1	-0.23539	0.02872	-8.20	<.0001	1
MA2,1	0.90684	0.0094434	96.03	<.0001	2
MA2,2	0.03794	0.0069504	5.46	<.0001	365
MA3,1	0.99592	0.0019672	506.26	<.0001	7
AR1,1	-0.59289	0.02769	-21.41	<.0001	1
AR1,2	0.33515	0.02674	12.53	<.0001	2
AR2,1	-0.46941	0.01382	-33.97	<.0001	365
AR3,1	0.99993	0.00008001	12498.3	<.0001	7

Figura 40: Ajuste del Modelo para NO_2

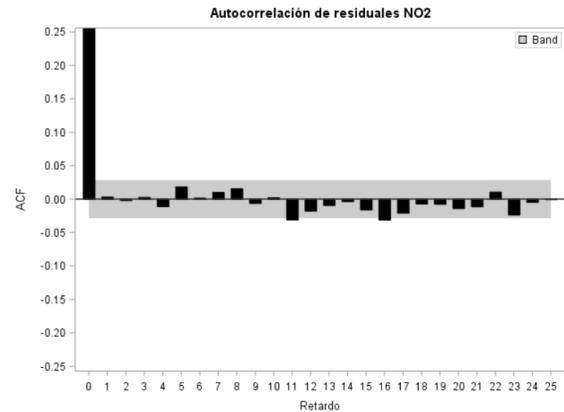


Figura 41: ACF residual modelo ARIMA NO_2

$$\begin{aligned}
 & (1 + 0,593B - 0,335B^2)(1 - 0,999B^7)(1 + 0,469B^{365})(1 - B)(1 - B^{365})Y_t \\
 & = (1 + 0,235B)(1 - 0,907B^2 - 0,038B^{365})(1 - 0,996B^7)\varepsilon_t
 \end{aligned}$$

Finalmente, las medidas de ajuste obtenidas para el modelo son: $AIC = 34320.29$ y $SSEp = 23041.278565$.

6 Modelos dinámicos para la predicción del ozono

6.1 Modelo dinámico con la temperatura como predictor

Como se ha señalado previamente, la concentración de ozono troposférico presenta una relación directa con la temperatura. Por ello, en esta sección se propone el ajuste de un modelo de función de transferencia que permita cuantificar dicho efecto. La variable explicativa considerada será la **temperatura media diaria**, mientras que la serie respuesta será la **concentración de ozono**. El objetivo de este modelo es mejorar el ajuste previamente obtenido mediante un modelo univariante aplicado al ozono, incorporando ahora la influencia de una variable explicativa adicional, con el fin de reducir el error de predicción.

Para construir este modelo en nuestro caso práctico, primero se ha tenido en cuenta la Función de Correlación Cruzada (CCF) entre ambas variables, para determinar la existencia de relación entre ellas. Se aprecia que los primeros retardos presentan correlaciones significativas. Esto sugiere una relación directa entre ambas variables, aunque no necesariamente inmediata. En este contexto, se ha considerado un retardo mínimo de impacto $b = 1$ para el modelo de función de transferencia.

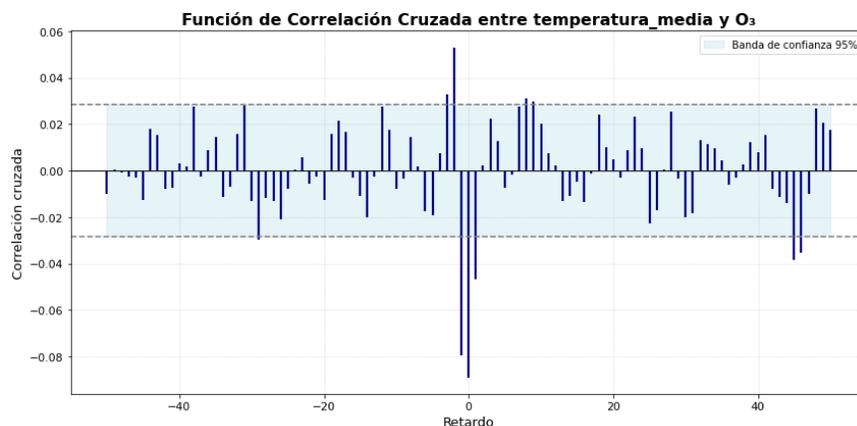


Figura 42: CCF entre O3 y Temperatura

Para confirmar esta elección, se ha procedido a estimar el modelo tanto con $b = 0$ como con $b = 1$, observándose un mejor ajuste con $b = 1$ en términos de estimación y análisis de residuos. Esta elección implica que el efecto de la temperatura sobre el ozono no se manifiesta de forma instantánea, sino que requiere al menos un día para influir en la concentración de ozono. Esta interpretación es coherente con los procesos fotoquímicos implicados en la formación de ozono troposférico, que suelen responder con cierto desfase a las condiciones térmicas.

A continuación, la estimación de los coeficientes del modelo función de transferencia. Todos los términos del modelo son significativos.

Estimación de probabilidad máxima							
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo	Variable	Desplazamiento
MU	0.42588	0.10644	4.00	<.0001	0	O3	0
MA1,1	-0.04715	0.01402	-3.36	0.0008	7	O3	0
MA2,1	0.95052	0.04262	22.30	<.0001	365	O3	0
AR1,1	0.64681	0.01397	46.29	<.0001	1	O3	0
AR1,2	-0.09277	0.01662	-5.58	<.0001	2	O3	0
AR1,3	0.07676	0.01397	5.49	<.0001	3	O3	0
NUM1	-0.16855	0.05465	-3.08	0.0020	0	temperatura_media	0
NUM1,1	0.11373	0.04951	2.30	0.0216	1	temperatura_media	0
DEN1,1	-0.68845	0.15437	-4.46	<.0001	365	temperatura_media	0

Constante Estimación	0.157235
Varianza Estimación	112.1309
Error Std Estimación	10.58919
AIC	33866.64
SBC	33924.1
Número de residuales	4382

Figura 43: Ajuste Modelo Función de Transferencia para Ozono. [Código SAS: Anexo II]

Se plantea la función de transferencia seleccionada. Esta tarea de selección no ha sido sencilla, debido a que el modelo ha planteado diversos problemas de significancia y correlaciones elevadas entre los parámetros. Finalmente, se ha elegido el siguiente modelo, el cual sí contiene constante del modelo significativa, expresado de forma general mediante la ecuación de función de transferencia:

$$(1-B^{365})O_{3,t}-0,157 = \frac{(-0,168 - 0,114B)(1 - B)(1 - B^{365}) \cdot temperatura_{t-1}}{1 + 0,688B^{365}} + \frac{(1 + 0,05B^7)(1 - 0,95B^{365}) a_t}{1 - 0,65B + 0,09B^2 - 0,08B^3}$$

donde $O_{3,t}$ representa la concentración de ozono diaria. El polinomio del numerador describe la forma en que la serie explicativa entra en el modelo (corto plazo), y el denominador recoge la dinámica a largo plazo y estacional del efecto de la temperatura sobre el ozono.

El segundo término del sumando representa el componente de ruido del modelo, que ha sido modelado como el proceso ARIMA de la serie univariante del ozono.

La matriz de correlaciones entre los parámetros del modelo [Anexo I] se considera válida, al no contar con elevadas correlaciones para la gran cantidad de datos que se están usando en la estimación. Se ha optado por un modelo con un número moderado de parámetros, con el fin de evitar un sobreajuste.

[Anexo I] El test de Ljung-Box sobre los residuales tiene un p-valor de 0.0438. Aunque se encuentra al borde de la significancia, tomaremos como válidos los residuos del modelo, ya que es complicado encontrar uno que ajuste los datos.

Teniendo en cuenta distintos criterios de medida del ajuste, obtenemos un AIC de 33866,89 y un SSEp de 72114,09. Ambos suponen una mejora respecto a los valores obtenidos para estas medidas en el

modelo univariante del ozono [Apartado 5.1]. Esta reducción en los criterios de ajuste respalda la incorporación de la temperatura como variable explicativa en el modelo.

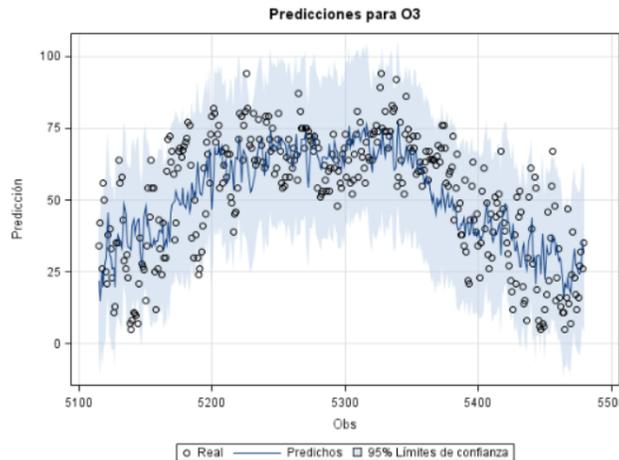


Figura 44: Predicción Año 2024 Función de Transferencia

Este nuevo modelo permite cuantificar el efecto dinámico de la temperatura sobre la concentración de ozono, mejorando el ajuste respecto a un modelo puramente univariante.

Sin embargo, en la función de transferencia ajustada se observa que la temperatura media diaria puede tener un leve efecto negativo sobre los niveles de ozono (O_3). En concreto, la suma de los coeficientes del numerador es $-0,16 - 0,11 = -0,26$, lo que indica que un aumento sostenido en la temperatura se asocia con una ligera disminución del ozono.

Aunque este resultado pueda parecer erróneo, ya que el ozono troposférico suele formarse con altas temperaturas y radiación solar, hay estudios que explican este comportamiento. Por ejemplo, [21] muestra que, en California, cuando se superan ciertos niveles térmicos (por encima de $39\text{ }^\circ\text{C}$), el efecto de la temperatura sobre el ozono puede volverse negativo debido a la descomposición de compuestos como los PAN, claves para la formación del O_3 . También en otro estudio [22] se señala que en algunas zonas del hemisferio norte pueden darse situaciones que desfavorezcan la formación del ozono con temperaturas altas.

Por tanto, aunque nuestros datos no permiten confirmar este hecho, los resultados podrían reflejar procesos en Valladolid que reduzcan el ozono cuando hay una exposición continuada a altas temperaturas, mayor ventilación atmosférica, etc.

Como interpretación final de este modelo diremos que el retardo diario de la temperatura tiene, según este modelo, un efecto negativo en los niveles de ozono diarios. Este efecto se manifiesta especialmente a lo largo del tiempo debido a la presencia de una estructura anual. En concreto, las temperaturas elevadas sostenidas en el tiempo ayudan a la disminución del ozono.

6.2 Modelo dinámico con precipitación como predictor

En este apartado vamos a tratar de mejorar la predicción del ozono añadiendo a la función de transferencia anterior la variable explicativa de precipitación.

Relación entre O₃ y Precipitación

La relación entre la precipitación y los niveles de ozono troposférico O₃ es indirecta pero significativa. El ozono troposférico no se emite directamente a la atmósfera como ya hemos explicado, sino que se forma a través de reacciones fotoquímicas. Esto lo convierte en un contaminante secundario, cuya concentración está altamente influenciada por las condiciones meteorológicas y ambientales.

Cuando se producen lluvias, el aire experimenta una serie de cambios que tienden a reducir las concentraciones de ozono. Por un lado, la precipitación ayuda a eliminar algunos de sus precursores, lo que indirectamente reduce la formación de ozono. Por otro lado, los días lluviosos suelen estar asociados con nubosidad y menor radiación solar, lo que limita la energía para las reacciones fotoquímicas que lo generan. Además, las lluvias suelen venir acompañadas de mayor agitación atmosférica y vientos, lo que favorece la dispersión de contaminantes acumulados.

Este efecto de la lluvia hace que, en muchas regiones, los niveles de ozono tiendan a ser más bajos en días húmedos o durante los meses más lluviosos.

Como veíamos en el análisis descriptivo de la variable precipitación, para conseguir una distribución más uniforme y con mayor homogeneidad de varianza ha sido necesario transformar mediante el método de inflación de ceros.

Función de Transferencia

En el **Apartado 6** se mostró cómo el modelo dinámico de función de transferencia puede ser útil para explicar la evolución del ozono troposférico al incorporar una variable predictora. En esta sección, ajustaremos de nuevo un modelo dinámico teniendo la precipitación como variable explicativa, con el objetivo de comparar la capacidad predictiva entre modelos.

En la Figura 45, se representa la función de correlación cruzada (CCF) entre el O₃(365), y la precipitación log-transformada. Esta gráfica permite examinar la posible relación dinámica entre ambas variables.

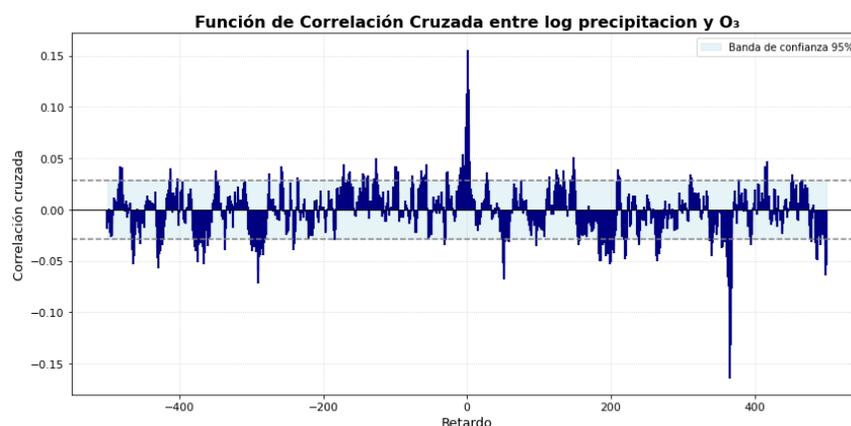


Figura 45: Correlaciones cruzadas entre O₃(365) y Log Precipitación

En la CCF se observa un pico positivo en los retardos cercanos a cero, concretamente entre 2 y 5 días,

lo cual sugiere una asociación entre la precipitación reciente y un incremento en los niveles de ozono en días posteriores. El comportamiento la CCF motiva la consideración del predictor en el modelo.

Asimismo, se aprecia un aumento en del retardo 365. Esta observación refleja una dinámica de tipo estacional de largo plazo, relacionada con patrones anuales de la variabilidad. Por tanto, aunque en el modelo univariante veíamos que la estacionalidad no jugaba un papel significativo, aquí la deberemos de tener en cuenta dentro del vector de entradas de la variable en el modelo dinámico.

El modelo propuesto para esta ampliación de la función de transferencia viene dado por:

$$(1-B^{365})O_{3,t} = \frac{(2,046 - 0,766 B - 1,397 B^2) \cdot \log(\text{precipitación})_{t-1}}{(1 + 0,798 B^{365} + 0,589 B^{730})(1 + 0,33 B^2)} + \frac{(1 + 0,05B^7)(1 - 0,92B^{365}) a_t}{1 - 0,65B + 0,10B^2 - 0,08B^3}$$

La estimación de coeficientes del modelo dinámico;

Estimación de probabilidad máxima							
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo	Variable	Desplazamiento
MA1,1	-0.04618	0.01422	-3.25	0.0012	7	O3	0
MA2,1	0.91558	0.02580	35.48	<.0001	365	O3	0
AR1,1	0.65078	0.01404	46.35	<.0001	1	O3	0
AR1,2	-0.10354	0.01672	-6.19	<.0001	2	O3	0
AR1,3	0.08405	0.01409	5.96	<.0001	3	O3	0
NUM1	2.04571	0.20332	10.06	<.0001	0	log_precipitacion	1
NUM1,1	-0.76594	0.20074	-3.82	0.0001	1	log_precipitacion	1
NUM1,2	-1.39705	0.36059	-3.87	0.0001	2	log_precipitacion	1
DEN1,1	-0.33046	0.17874	-1.85	0.0645	2	log_precipitacion	1
DEN2,1	-0.79843	0.06949	-11.49	<.0001	365	log_precipitacion	1
DEN2,2	-0.58886	0.07443	-7.91	<.0001	730	log_precipitacion	1

Figura 46: Parámetros ajustado para el modelo con Log Precipitación

Todos los coeficientes del modelo son significativos. Además, al analizar los residuos mediante el test de Ljung-Box, se obtiene un p-valor de 0.0815, asumiendo el comportamiento de residuos como ruido blanco. La matriz de correlaciones entre los parámetros del modelo [**Anexo I**] contiene correlaciones asumibles admitiendo el modelo.

Los criterios de ajuste que se obtienen en esta estimación son; $AIC = 33773,36$ y $SSE_p = 69915,73$.

Observando los residuales y el error de predicción obtenidos con las observaciones de test, podemos concluir que el modelo ajustado no presenta signos de sobreajuste a los datos de entrenamiento, por lo que es válido para la predicción de valores futuros.

Al comparar este resultado con el modelo dinámico que incluye la temperatura, observamos una importante reducción del error en la predicción. Sin embargo, el criterio de información AIC ha aumentado. Esto se debe a que el modelo generaliza mejor a nuevos datos, pero se ajusta peor a los datos de entrenamiento.

Las predicciones para el año 2024 de este nuevo modelo se dan en el siguiente gráfico:

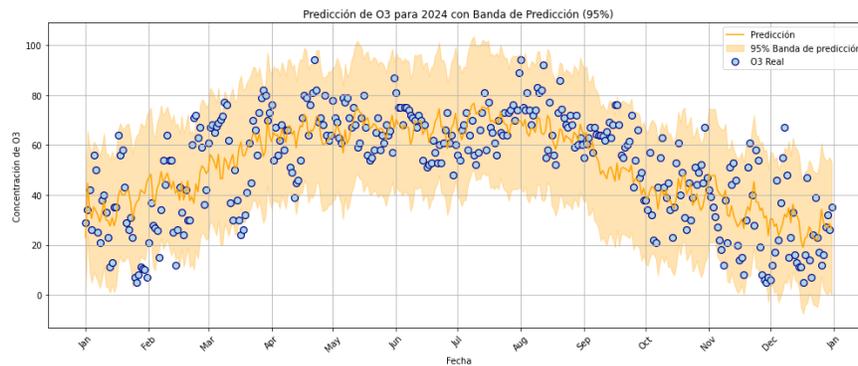


Figura 47: Predicciones 2024 Modelo dinámico con Precipitación

Este modelo dinámico muestra que la variable $\log(\text{precipitación})$ tiene un efecto negativo sobre los niveles de ozono. Este efecto se extiende también al retardo anterior, viendo la dependencia con valores anteriores. Esto lo podemos ver con la presencia de coeficientes negativos en el numerador del polinomio, indicando que estas precipitaciones pueden provocar una reducción en el nivel del ozono. Además, el modelo incluye componente estacional anual, lo que refleja la ciclicidad de este fenómeno a lo largo del tiempo.

6.3 Modelo dinámico con NO_2 como predictor

En esta sección vamos a tratar de mejorar la predicción del ozono añadiendo un nuevo predictor a nuestro modelo dinámico que venimos ajustando; el NO_2 . Como venimos viendo a lo largo de nuestro estudio, el NO_2 es uno de los agentes contaminantes más influyentes en la creación del ozono troposférico, por lo que no podía quedar fuera del ajuste del modelo. En el apartado del análisis descriptivo [4.4] hemos visto la fuerte correlación existente entre estos dos agentes contaminantes.

El dióxido de nitrógeno (NO_2) juega un papel clave en la formación del ozono troposférico. En presencia de luz solar, el NO_2 se descompone liberando oxígeno atómico (O), que reacciona con el oxígeno molecular (O_2) para formar ozono (O_3). Este proceso, junto con la presencia de compuestos orgánicos volátiles (COVs), impulsa reacciones fotoquímicas que aumentan la concentración de ozono.

Función de Transferencia

Con este apartado se va a concluir el modelo dinámico ajustado hasta ahora añadiendo una de las variables que parecen más pueden aportar a la predicción del O_3 , el dióxido de Nitrógeno. Para ver cuánto aporta este nuevo predictor y si es conveniente su adición al modelo, vamos a ampliar la anterior función de transferencia. Para ello, nos emplearemos de la CCF para determinar los parámetros del polinomio del modelo dinámico.

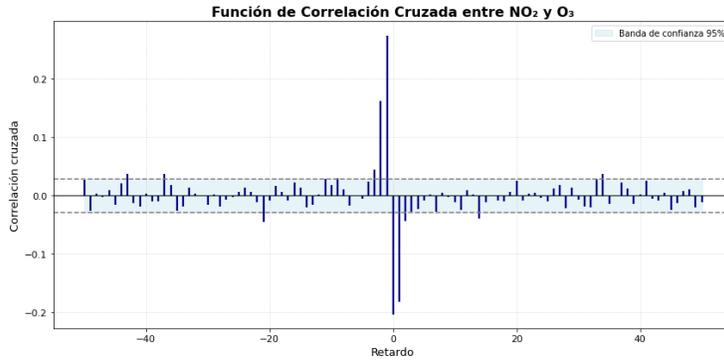


Figura 48: CCF de O_3 y NO_2

En el análisis realizado con la función de correlación cruzada (CCF) utilizando hasta 700 retardos, se han observado correlaciones estadísticamente significativas entre las concentraciones de dióxido de nitrógeno y ozono en los retardos correspondientes a periodos anuales. Por tanto, estos retardos serán considerados en el modelado. Asimismo, se identificaron retardos con correlaciones relevantes en retardos 7 y 14, lo que sugiere la presencia de una componente periódica de tipo semanal que también será tomada en cuenta. Con base en los resultados obtenidos a partir del análisis de la CCF, se ha procedido a ajustar la siguiente función de transferencia.

Parámetro	Estimación de probabilidad máxima				Retardo	Variable	Desplazamiento
	Estimación	Error estándar	t valor	Approx Pr > t			
MA1,1	-0.03592	0.01398	-2.57	0.0102	7	O3	0
MA2,1	0.96423	0.05963	16.17	<.0001	365	O3	0
AR1,1	0.72917	0.01410	51.73	<.0001	1	O3	0
AR1,2	-0.07210	0.01732	-4.16	<.0001	2	O3	0
AR1,3	0.03009	0.01406	2.14	0.0324	3	O3	0
NUM1	-0.56784	0.01430	-39.70	<.0001	0	NO2	0
NUM1,1	0.42343	0.01627	26.02	<.0001	1	NO2	0
NUM1,2	0.17215	0.01417	12.15	<.0001	2	NO2	0
DEN1,1	-0.05296	0.02441	-2.17	0.0300	7	NO2	0
DEN2,1	0.04834	0.02329	2.08	0.0379	365	NO2	0

Figura 49: Parámetros ajustados para Función Transferencia con NO_2

$$(1 - B^{365})O_{3,t} = \frac{(-0,568 - 0,423 B - 0,172B^2)(1 - B)(1 - B^{365})}{(1 + 0,053 B^7)(1 + 0,048 B^{365})} \cdot NO_{2,t} + \frac{(1 + 0,036 B^7)(1 - 0,964 B^{365})}{1 - 0,729 B + 0,072 B^2 - 0,03 B^3} a_t$$

El test de Ljung-Box para los residuos obtenidos tiene un p-valor para retardos del periodo anual de 0.0933, por lo que asumiremos que su comportamiento es de ruido blanco. En la siguiente figura, podemos ver la ACF de Residuales del modelo.

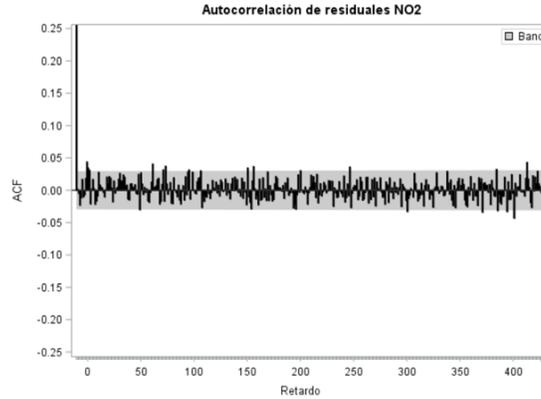


Figura 50: Análisis Residual Función Transferencia NO_2

La matriz de correlaciones no supone correlaciones fuertes entre parámetros [**Anexo I**]. Los criterios de medida del ajuste para el modelo dinámico se estiman en; $AIC = 32649,69$ y $SSE_p = 68950,13$.

El modelo dinámico con el NO_2 como predictor mejora mucho la capacidad de predicción del ozono, siendo la mayor que hemos obtenido en el estudio. Además, conseguimos un mejor ajuste a los datos de entrenamiento que en el modelo anterior.

El siguiente gráfico contiene la predicción y la banda de confianza para el modelo dinámico ajustado:

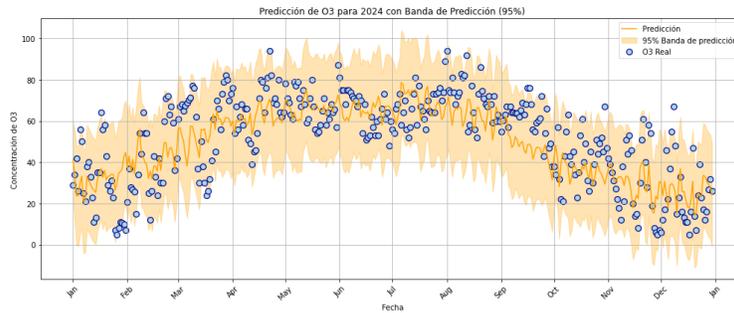


Figura 51: Predicciones 2024 Modelo dinámico con NO_2

Este modelo dinámico refleja la estrecha relación entre el dióxido de nitrógeno (NO_2) y el ozono (O_3). Como hemos comentado, en la atmósfera, el NO_2 actúa como precursor del ozono. Sin embargo, el modelo también muestra que niveles elevados de NO_2 están asociados con una disminución del ozono, lo cual concuerda con el fenómeno conocido como titulación del ozono, donde el NO_2 reacciona rápidamente con el O_3 y lo destruye.

6.4 Comparación de modelos

Cuadro 14: Comparación de modelos con medidas AIC y SSEp

	Univariante O_3	Mod. Dinámico Temp.	Mod. Dinámico Precip.	Mod. Dinámico NO_2
AIC	36.733,43	33.866,89	33.773,36	32.649,69
SSEp	73.168,02	72.114,09	69.915,73	68.950,13

Los resultados presentados en la Tabla 14 evidencian que, entre los modelos evaluados, el **modelo dinámico con NO_2 como variable explicativa** presenta el mejor ajuste global a los datos de concentración de O_3 , al tener el menor valor del SSEp (68.950,13), lo cual indica una mayor precisión en la predicción. También cuenta con el valor más bajo del AIC (32.649,69), lo que sugiere también que es el modelo que mejor equilibra la bondad de ajuste y la complejidad del modelo. Los demás modelos dinámicos cuentan con criterios AIC muy parecidos a este.

Un aspecto destacable en la comparación entre modelos dinámicos es que la temperatura es la única variable explicativa cuya función de transferencia contiene constante significativa. Aunque en el modelo univariante de la temperatura dicha constante no era significativa, se optó por eliminarla debido a que su valor no era significativo, pero estaba próximo al umbral de significación. Sin embargo, al incorporar la temperatura como variable explicativa en el modelo dinámico del ozono, la constante sí alcanzó significación estadística, lo que sugiere que su efecto conjunto con el ozono aporta información relevante al modelo.

En conjunto, estos resultados indican que la incorporación de variables explicativas climáticas o contaminantes mejora el ajuste frente al modelo univariante, y que la selección final del modelo puede depender del criterio al que damos mayor prioridad, ya sea precisión predictiva (SSEp) o bondad de ajuste (AIC). El modelo con NO_2 , al mostrar un equilibrio favorable en ambos criterios, puede considerarse el **modelo más robusto para la predicción de O_3** . Este comportamiento es coherente con la naturaleza química del ozono, cuya formación está estrechamente ligada a la presencia de precursores como el dióxido de nitrógeno.

A continuación, la representación de los modelos ajustados para la explicación del ozono, junto a sus bandas de confianza del 95 %:

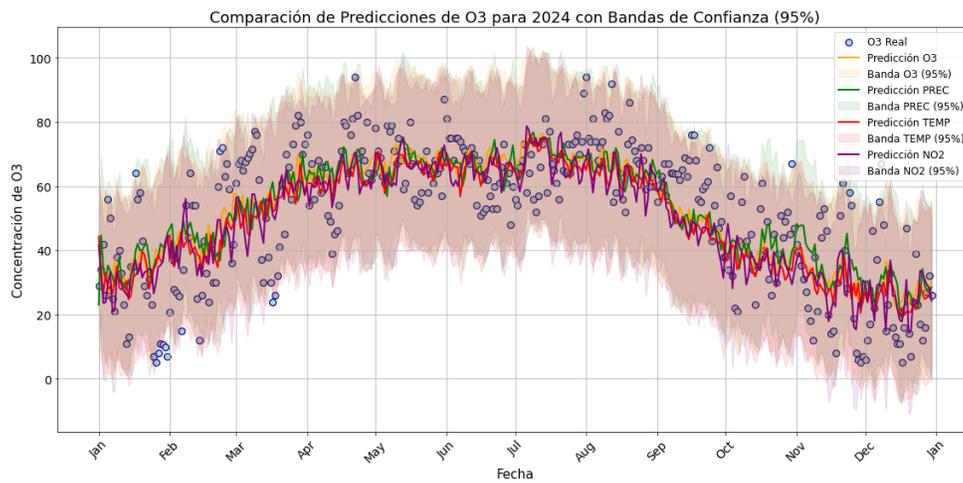


Figura 52: Comparación Modelos O_3

7 Conclusiones, limitaciones y líneas futuras

En este trabajo se ha llevado a cabo un estudio exhaustivo sobre la modelización de la concentración de ozono (O_3) en la estación meteorológica Puente Poniente - M^a Luisa Sánchez, mediante series temporales, utilizando la herramienta SAS como entorno principal de desarrollo. Se han ajustado modelos univariantes y de *función de transferencia* basados en el enfoque *Box-Jenkins*, incorporando distintas variables explicativas como la temperatura, la precipitación y el NO_2 .

Los resultados muestran que los modelos de función de transferencia presentan una mejor precisión en la predicción en comparación con el modelo univariante, tal y como se refleja en el *AIC* y en el *SSEp* obtenidos. En particular, el modelo que incluye NO_2 como variable predictiva tiene el menor valor para el criterio de predicción y el criterio *AIC*.

Por otro lado, desde el punto de vista metodológico, los modelos *Box-Jenkins* han ofrecido una buena base para capturar la dinámica temporal del O_3 , mientras que los modelos de *función de transferencia* han permitido modelar la influencia de otras variables sobre el contaminante.

En cuanto a la complejidad computacional, si bien los modelos de función de transferencia implican un mayor esfuerzo en términos de identificación y estimación, la herramienta SAS ha demostrado ser eficaz y robusta para su implementación y análisis. Sobre todo, si tenemos en cuenta que el periodo más largo que manejamos es 365 y periodos tan largos dan problemas en las funciones ya implementadas en R o Python.

Además, las bandas de confianza permiten no solo visualizar la evolución esperada de O_3 , sino también transmitir de forma clara la incertidumbre asociada a las predicciones que hemos realizado.

Es importante señalar que los análisis realizados en este trabajo se han llevado a cabo desde un punto de vista estadístico, con el objetivo de mostrar técnicas de modelización probabilística aplicadas a datos atmosféricos. Sin embargo, hay que ser conscientes de la complejidad de este tipo de datos, y de la limitación que supone no disponer de un conocimiento profundo en áreas como la climatología. La integración de estos conocimientos en el estudio podría contribuir a mejorar tanto la calidad de los modelos como la interpretación de los resultados obtenidos.

Se plantean **líneas de trabajo futuro** que podrían ampliarse:

- **Incorporación de nuevas variables explicativas:** Analizar la influencia de otros contaminantes atmosféricos (como PM_{10} , SO_2 o CO), así como de factores meteorológicos adicionales (humedad relativa, radiación solar) podría ayudar a explicar los modelos y mejorar su capacidad predictiva.
- **Modelización multivariante de series temporales:** Ajustar modelos que consideren múltiples variables explicativas, tanto contaminantes como meteorológicas. Podría tener interés ajustar modelos de función de transferencia con más de una variable explicativa, creando modelos más realistas. Por otro lado, también podríamos emplear modelos multivariantes como los VAR (Vectores Autorregresivos), que permiten capturar las relaciones dinámicas entre las series temporales [24]. Este enfoque sería más ambicioso, mejorando la capacidad explicativa y predictiva de los modelos ajustados.
- **Exploración de métodos de aprendizaje automático:** Comparar el rendimiento de los modelos

clásicos utilizados en este trabajo con técnicas más recientes como redes neuronales recurrentes (RNN, LSTM), que podrían capturar relaciones no lineales más complejas.

- **Validación cruzada:** Implementar esquemas de validación más exhaustivos, como *cross-validation* con ventanas deslizantes, para evaluar la estabilidad de los modelos a lo largo del tiempo.

Estas líneas permitirían no solo perfeccionar la modelización del ozono troposférico, sino también avanzar hacia sistemas de alerta temprana más precisos y útiles en la gestión de la calidad del aire.

Bibliografía

- [1] *Modelos ARIMA: Serie Temporal*. Disponible en <https://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>
- [2] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). Disponible en <https://otexts.com/fpp3/>
- [3] Departamento de Estadística e Investigación Operativa. *Apuntes de la asignatura Análisis de Series Temporales*. Universidad de Valladolid (UVa), Facultad de Ciencias. Material no publicado.
- [4] Brockwell, P. J., & Davis, R. A. (2016). *Series Estacionarias. Introduction to Time Series and Forecasting* (3ª ed.). Springer.
- [5] Bandara, K., Hyndman, R. J., & Bergmeir, C. (2021). *MSTL: A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns*. Disponible en <https://arxiv.org/abs/2107.13462>
- [6] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*.
- [7] La Vanguardia. (25 de agosto de 2024). *Valladolid activa la Situación 1 preventiva por un episodio de contaminación por ozono*. Disponible en <https://www.lavanguardia.com/vida/20240825/9889716/valladolid-activa-situacion-1-preventiva-episodio-contaminacion-ozono-agenciaslv202408.html>
- [8] Martín Sobrino, M. Á. (2012). *Estudio multifractal de la influencia de factores meteorológicos y químicos en la concentración de ozono troposférico* Disponible en <https://helvia.uco.es/xmlui/handle/10396/7977>
- [9] National Geographic España. (2023). *Noticia Ozono Troposférico*. Disponible en https://www.nationalgeographic.com.es/medio-ambiente/espana-se-producen-entre-1300-y-1500-muertes-prematuras-por-contaminacion-ozono_20281
- [10] Climate and Clean Air Coalition (CCAC). *Ozono troposférico*. <https://www.ccacoalition.org/es/short-lived-climate-pollutants/tropospheric-ozone>. Consultado el 7 de junio de 2025.
- [11] Maji, S., Dikshit, A. K., & Deshpande, A. W. (2019). Relationship between rainfall and ground-level ozone concentration in an urban area of western India. *Current Pollution Reports*, 5 <https://link.springer.com/article/10.1007/s40726-019-00118-3>
- [12] Rodríguez, J. (2023). *Relación entre la meteorología y la contaminación por ozono en la región de Barcelona*. Universitat Politècnica de Catalunya. <https://upcommons.upc.edu/handle/2117/417793>

- [13] Agencia de Protección Ambiental de Estados Unidos (EPA), *Óxidos de Nitrógeno (NOx): ¿Por qué y cómo se controlan?*
Disponible en: <https://www3.epa.gov/ttn/catc1/cica/files/fnoxdocs.pdf>
- [14] Ayuntamiento de Valladolid. *Red de Control de la Contaminación Atmosférica de Valladolid (RCCAVA)*. Disponible en <https://www.valladolid.es/es/rccava>
- [15] Agencia Estatal de Meteorología (AEMET). *Datos abiertos*. Disponible en https://www.aemet.es/es/datos_abiertos
- [16] *Artículo sobre Filtro de Kalman*. Disponible en <https://diposit.ub.edu/dspace/bitstream/2445/150978/2/150978.pdf>
- [17] Parra, R. (2017). *Efecto Fin de Semana en la calidad del aire de la ciudad de Cuenca, Ecuador*. Disponible en <http://doi.org/10.18272/aci.v9i15.291>
- [18] (Sicard et al., 2020) Amplified ozone pollution in cities during the COVID-19 lockdown. *Science of The Total Environment*. <https://doi.org/10.1016/j.scitotenv.2020.139542>
- [19] Dominique Bäumer y Bernhard Vogel, *Un patrón inesperado de periodicidades semanales distintas en variables climatológicas en Alemania*, *Geophysical Research Letters*, vol. 34, núm. 3, 2007. <https://doi.org/10.1029/2006GL028559>
- [20] *Ciclos semanales en precipitación y otras variables meteorológicas en una región contaminada de Europa*, *Atmospheric Chemistry and Physics*, vol. 11, 2011. <https://doi.org/10.5194/acp-11-4095-2011>
- [21] *Relación temperatura-ozono y sus mecanismos durante olas de calor extremas*. <https://doi.org/10.5194/acp-23-1873-2023>
- [22] *Dependencia del ozono superficial respecto a la temperatura en el hemisferio norte durante el verano*. <https://doi.org/10.5194/acp-22-1601-2022>
- [23] Aránzazu de Juan Fernández. *Función de Transferencia*. Disponible en: https://www.cartagena99.com/recursos/alumnos/apuntes/Funcion_de_Transferencia_I.pdf
- [24] Peña, D. (2010). *Análisis de series temporales*.
- [25] SAS Institute Inc. *PROC ARIMA Statement*. Disponible en: https://documentation.sas.com/doc/en/etsug/15.2/etsug_arima_syntax02.htm
- [26] Alonso Losa, Lucía. *Aplicación del análisis de series temporales a variables atmosféricas*. Disponible en: <https://uvadoc.uva.es/handle/10324/50477>.
- [27] Marcos Parra, Pablo. *Comparativa de modelos en Python para la predicción de una serie temporal diaria de temperatura*. Disponible en: <https://uvadoc.uva.es/handle/10324/74269>.

Anexo I

Serie Ozono

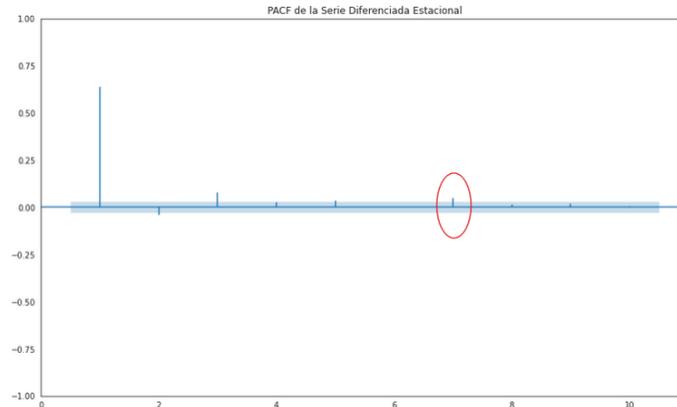


Figura 53: PACF Serie O3 Zoom Retardo 7

Correlaciones de las estimaciones de parámetro						
Parámetro	MU	MA1,1	AR1,1	AR1,2	AR1,3	AR1,4
MU	1.000	-0.015	0.001	0.000	-0.001	-0.002
MA1,1	-0.015	1.000	-0.018	-0.003	0.008	0.001
AR1,1	0.001	-0.018	1.000	-0.545	0.061	-0.052
AR1,2	0.000	-0.003	-0.545	1.000	-0.541	-0.014
AR1,3	-0.001	0.008	0.061	-0.541	1.000	-0.114
AR1,4	-0.002	0.001	-0.052	-0.014	-0.114	1.000

Figura 54: Matriz Correlación Ozono Modelo 1

Correlaciones de las estimaciones de parámetro						
Parámetro	MU	MA1,1	MA2,1	AR1,1	AR1,2	AR1,3
MU	1.000	0.000	-0.016	0.001	0.000	-0.001
MA1,1	0.000	1.000	-0.005	0.007	0.024	0.063
MA2,1	-0.016	-0.005	1.000	-0.018	-0.004	0.008
AR1,1	0.001	0.007	-0.018	1.000	-0.545	0.060
AR1,2	0.000	0.024	-0.004	-0.545	1.000	-0.543
AR1,3	-0.001	0.063	0.008	0.060	-0.543	1.000

Figura 55: Matriz Correlación Ozono Modelo 2

Correlaciones de las estimaciones de parámetro							
Parámetro	MU	MA1,1	MA1,2	AR1,1	AR1,2	AR1,3	AR1,4
MU	1.000	0.000	-0.020	0.001	0.000	-0.001	-0.002
MA1,1	0.000	1.000	-0.262	-0.005	-0.000	-0.011	0.041
MA1,2	-0.020	-0.262	1.000	-0.008	0.002	0.012	-0.008
AR1,1	0.001	-0.005	-0.008	1.000	-0.544	0.060	-0.050
AR1,2	0.000	-0.000	0.002	-0.544	1.000	-0.540	-0.014
AR1,3	-0.001	-0.011	0.012	0.060	-0.540	1.000	-0.115
AR1,4	-0.002	0.041	-0.008	-0.050	-0.014	-0.115	1.000

Figura 56: Matriz Correlación Ozono Modelo 3

Correlaciones de las estimaciones de parámetro											
Parámetro	MU	MA1,1	MA1,2	MA1,3	MA1,4	MA1,5	AR1,1	AR1,2	AR1,3	AR1,4	AR1,5
MU	1.000	0.001	0.001	-0.002	-0.010	-0.002	-0.003	-0.000	-0.004	-0.006	-0.008
MA1,1	0.001	1.000	0.093	0.012	-0.143	-0.093	0.031	-0.016	0.002	0.049	0.037
MA1,2	0.001	0.093	1.000	0.134	-0.131	-0.065	0.004	0.008	-0.009	0.058	0.029
MA1,3	-0.002	0.012	0.134	1.000	-0.172	-0.100	0.021	-0.027	0.004	0.029	0.013
MA1,4	-0.010	-0.143	-0.131	-0.172	1.000	-0.148	0.048	-0.007	0.027	0.000	0.448
MA1,5	-0.002	-0.093	-0.065	-0.100	-0.148	1.000	-0.077	0.020	-0.016	-0.052	-0.676
AR1,1	-0.003	0.031	0.004	0.021	0.048	-0.077	1.000	-0.545	0.065	-0.042	0.121
AR1,2	-0.000	-0.016	0.008	-0.027	-0.007	0.020	-0.545	1.000	-0.542	-0.017	-0.016
AR1,3	-0.004	0.002	-0.009	0.004	0.027	-0.016	0.065	-0.542	1.000	-0.116	0.049
AR1,4	-0.006	0.049	0.058	0.029	0.000	-0.052	-0.042	-0.017	-0.116	1.000	0.032
AR1,5	-0.008	0.037	0.029	0.013	0.448	-0.676	0.121	-0.016	0.049	0.032	1.000

Figura 57: Matriz Correlación Ozono Modelo 4

Estimación de probabilidad máxima					
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo
MU	0.29019	0.10050	2.89	0.0039	0
MA1,1	0.96312	0.04919	19.58	<.0001	365
AR1,1	0.65255	0.01331	49.04	<.0001	1
AR1,2	-0.10542	0.01585	-6.65	<.0001	2
AR1,3	0.06581	0.01339	4.92	<.0001	3
AR1,4	0.05175	0.01065	4.86	<.0001	7
Constante Estimación		0.097303			
Varianza Estimación		113.3511			
Error Std Estimación		10.64665			
AIC		36742.81			
SBC		36781.61			
Número de residuales		4748			

Figura 58: Ajuste Ozono modelo 1

Estimación de probabilidad máxima					
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo
MU	0.29109	0.09227	3.15	0.0016	0
MA1,1	-0.04723	0.01341	-3.52	0.0004	7
MA2,1	0.96604	0.05332	18.12	<.0001	365
AR1,1	0.65445	0.01330	49.19	<.0001	1
AR1,2	-0.10621	0.01585	-6.70	<.0001	2
AR1,3	0.06987	0.01334	5.24	<.0001	3
Constante Estimación		0.111161			
Varianza Estimación		113.2842			
Error Std Estimación		10.6435			
AIC		36751.8			
SBC		36790.59			
Número de residuales		4748			

Figura 59: Ajuste Ozono modelo 2

Estimación de probabilidad máxima					
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo
MU	0.28909	0.10554	2.74	0.0062	0
MA1,1	-0.03314	0.01147	-2.89	0.0039	13
MA1,2	0.95809	0.03912	24.49	<.0001	365
AR1,1	0.65121	0.01333	48.86	<.0001	1
AR1,2	-0.10435	0.01587	-6.58	<.0001	2
AR1,3	0.06623	0.01341	4.94	<.0001	3
AR1,4	0.04988	0.01068	4.67	<.0001	7
Constante Estimación		0.097432			
Varianza Estimación		113.5359			
Error Std Estimación		10.65532			
AIC		36737.64			
SBC		36782.9			
Número de residuales		4748			

Figura 60: Ajuste Ozono modelo 3

Estimación de probabilidad máxima					
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo
MU	0.26444	0.11057	2.39	0.0168	0
MA1,1	-0.02378	0.01231	-1.93	0.0533	14
MA1,2	-0.02670	0.01249	-2.14	0.0326	21
MA1,3	-0.02774	0.01280	-2.17	0.0303	28
MA1,4	0.85304	0.03928	21.72	<.0001	365
MA1,5	0.09187	0.02375	3.87	0.0001	730
AR1,1	0.65129	0.01345	48.44	<.0001	1
AR1,2	-0.10488	0.01592	-6.59	<.0001	2
AR1,3	0.06230	0.01345	4.63	<.0001	3
AR1,4	0.04794	0.01069	4.49	<.0001	7
AR1,5	-0.04955	0.01743	-2.84	0.0045	365
Constante Estimación		0.103898			
Varianza Estimación		114.7164			
Error Std Estimación		10.71057			
AIC		36733.43			
SBC		36804.55			
Número de residuales		4748			

Figura 61: Ajuste Ozono modelo 4

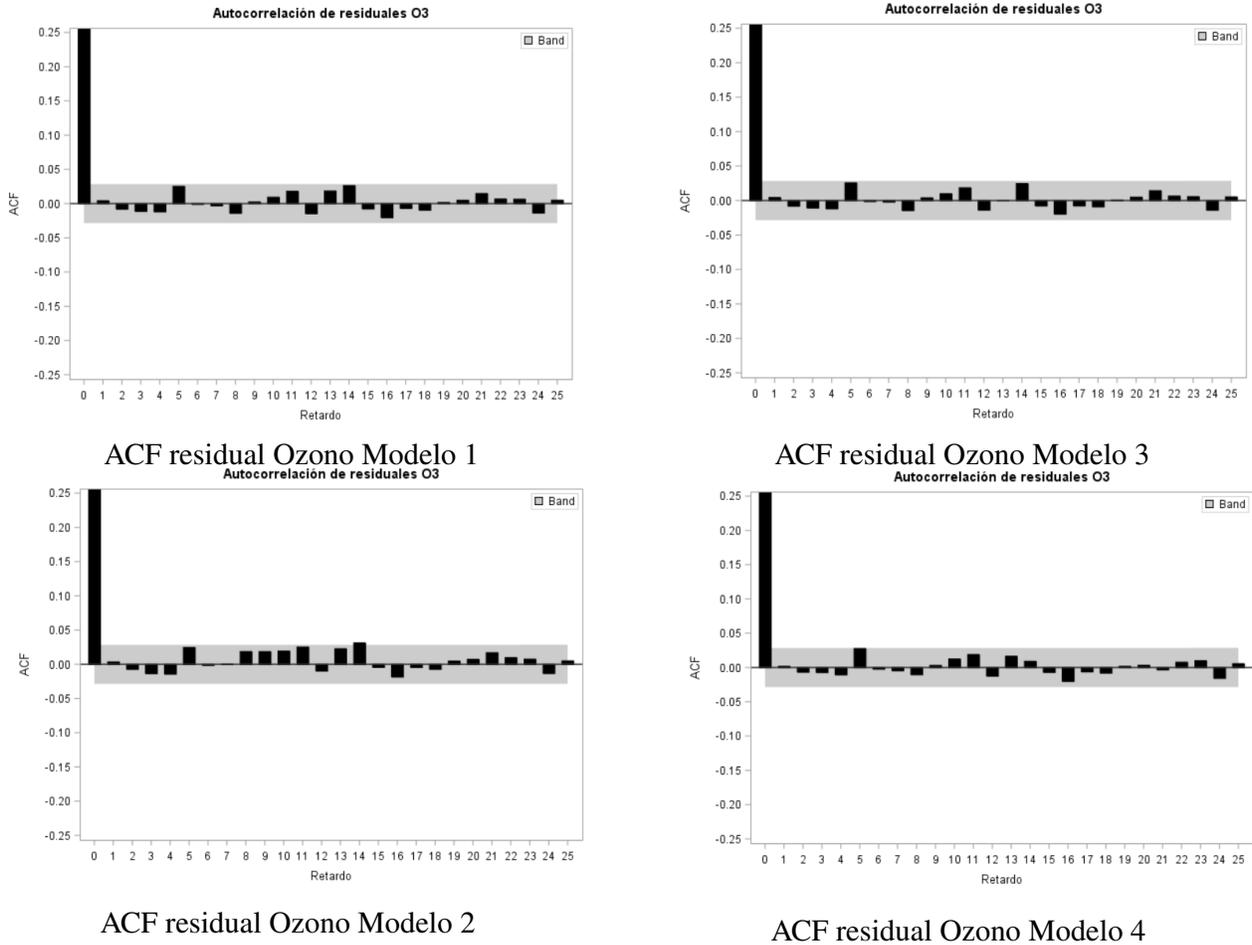


Figura 62: ACF Residual Modelo ozono

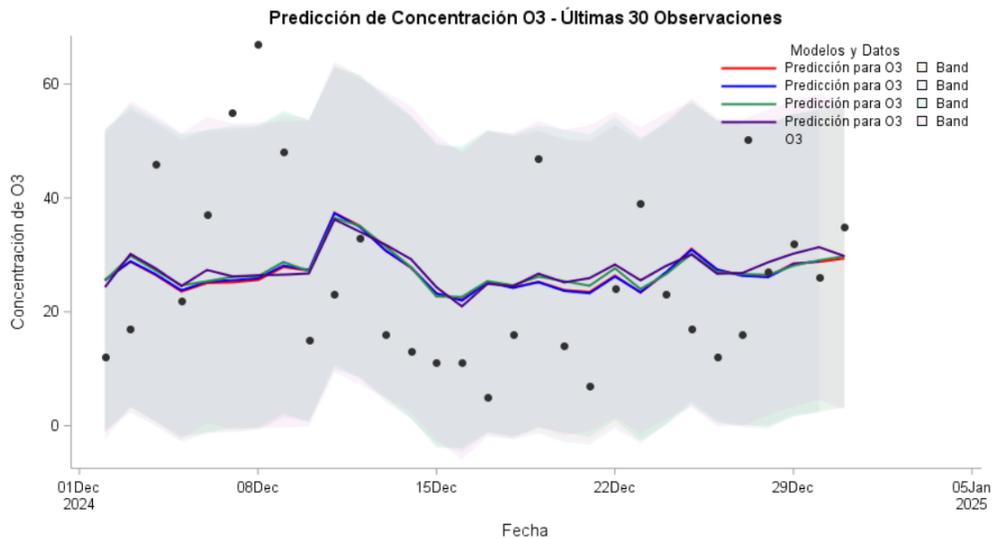


Figura 63: Predicción 2024 Serie O_3 últimas 30 obvs.

Serie Temperatura

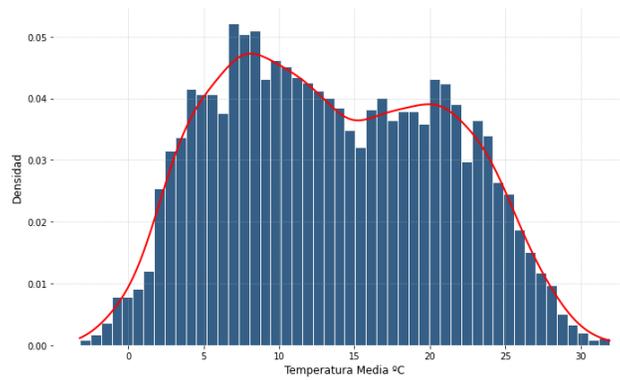
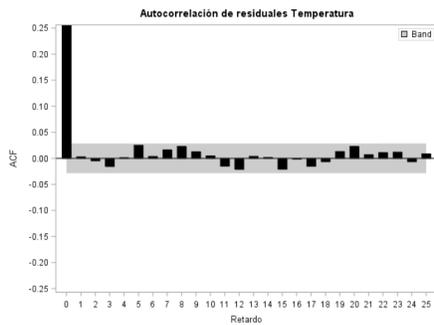
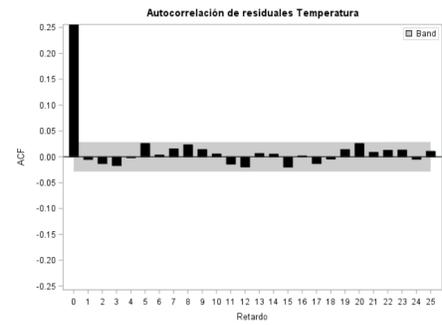


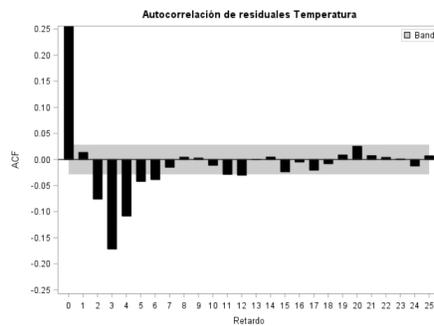
Figura 64: Distribución Temperatura



ACF residual Temperatura Modelo 1



ACF residual Temperatura Modelo 2



ACF residual Temperatura Modelo 3

Figura 65: ACF Residuales Modelo ARIMA Temperatura Media

Estimación de probabilidad máxima					
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo
MA1,1	0.74697	0.01654	45.16	<.0001	1
MA1,2	0.24921	0.01636	15.23	<.0001	2
MA2,1	0.02657	0.01240	2.14	0.0321	3
MA2,2	0.97343	0.05613	17.34	<.0001	365
AR1,1	0.66987	0.01295	51.72	<.0001	1

Varianza Estimación	3.95764
Error Std Estimación	1.989382
AIC	20858.93
SBC	20891.26
Número de residuales	4747

Estimación de probabilidad máxima					
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo
MA1,1	0.75411	0.47155	1.60	0.1098	1
MA1,2	0.24588	0.11622	2.12	0.0344	2
MA2,1	0.02566	0.01274	2.01	0.0440	3
MA2,2	0.88083	0.02767	31.84	<.0001	365
MA2,3	0.04312	0.01695	2.54	0.0109	730
AR1,1	0.68257	0.01301	52.45	<.0001	1

Varianza Estimación	4.145547
Error Std Estimación	2.036062
AIC	20881.38
SBC	20920.17
Número de residuales	4747

Figura 66: Ajuste Temperatura Modelo 2

Estimación de probabilidad máxima					
Parámetro	Estimación	Error estándar	t valor	Approx Pr > t	Retardo
MA1,1	-1.16067	0.0013013	-891.92	<.0001	1
MA1,2	-0.16067	0.01827	-8.79	<.0001	2
MA2,1	0.77336	0.02563	30.18	<.0001	365
MA2,2	0.12885	0.01766	7.30	<.0001	730
AR1,1	-1.12989	0.01566	-72.17	<.0001	1
AR1,2	-0.25365	0.0060795	-41.72	<.0001	2
AR1,3	-0.12361	0.01403	-8.81	<.0001	3

Varianza Estimación	4.924406
Error Std Estimación	2.2191
AIC	21595.42
SBC	21640.68
Número de residuales	4747

Figura 67: Ajuste Temperatura Modelo 3

Modelos Dinámicos

Correlaciones de las estimaciones de parámetro										
Variable Parámetro	O3 MA1,1	O3 MA2,1	O3 AR1,1	O3 AR1,2	O3 AR1,3	temperatura_media NUM1	temperatura_media NUM1,1	temperatura_media DEN1,1	temperatura_media DEN1,2	
O3 MA1,1	1.000	-0.017	0.001	0.028	0.077	-0.005	0.002	-0.010	0.009	
O3 MA2,1	-0.017	1.000	-0.006	-0.006	0.012	0.009	-0.005	0.007	-0.013	
O3 AR1,1	0.001	-0.006	1.000	-0.546	0.040	0.092	-0.048	-0.014	-0.032	
O3 AR1,2	0.028	-0.006	-0.546	1.000	-0.541	-0.089	-0.003	0.016	0.025	
O3 AR1,3	0.077	0.012	0.040	-0.541	1.000	0.015	0.059	-0.009	0.006	
temperatura_media NUM1	-0.005	0.009	0.092	-0.089	0.015	1.000	-0.482	-0.125	-0.355	
temperatura_media NUM1,1	0.002	-0.005	-0.048	-0.003	0.059	-0.482	1.000	0.107	0.215	
temperatura_media DEN1,1	-0.010	0.007	-0.014	0.016	-0.009	-0.125	0.107	1.000	-0.670	
temperatura_media DEN1,2	0.009	-0.013	-0.032	0.025	0.006	-0.355	0.215	-0.670	1.000	

Figura 68: Correlaciones Función de Transferencia Temperatura

Correlaciones de las estimaciones de parámetro												
Variable Parámetro	O3 MA1,1	O3 MA2,1	O3 AR1,1	O3 AR1,2	O3 AR1,3	log_precipitacion NUM1	log_precipitacion NUM1,1	log_precipitacion NUM1,2	log_precipitacion DEN1,1	log_precipitacion DEN2,1	log_precipitacion DEN2,2	
O3 MA1,1	1.000	-0.015	0.003	0.027	0.076	0.023	-0.003	-0.016	-0.013	0.031	0.017	
O3 MA2,1	-0.015	1.000	-0.009	-0.006	0.004	-0.064	0.034	0.009	-0.005	-0.029	-0.043	
O3 AR1,1	0.003	-0.009	1.000	-0.543	0.050	-0.005	0.003	-0.013	-0.015	0.015	0.019	
O3 AR1,2	0.027	-0.006	-0.543	1.000	-0.541	-0.007	0.007	0.010	0.010	-0.010	-0.033	
O3 AR1,3	0.076	0.004	0.050	-0.541	1.000	0.011	-0.003	-0.005	-0.008	0.012	0.032	
log_precipitacion NUM1	0.023	-0.064	-0.005	-0.007	0.011	1.000	-0.211	-0.182	-0.028	0.062	0.168	
log_precipitacion NUM1,1	-0.003	0.034	0.003	0.007	-0.003	-0.211	1.000	-0.034	-0.185	0.009	-0.006	
log_precipitacion NUM1,2	-0.016	0.009	-0.013	0.010	-0.005	-0.182	-0.034	1.000	0.830	-0.091	-0.135	
log_precipitacion DEN1,1	-0.013	-0.005	-0.015	0.010	-0.008	-0.028	-0.185	0.830	1.000	-0.095	-0.105	
log_precipitacion DEN2,1	0.031	-0.029	0.015	-0.010	0.012	0.062	0.009	-0.091	-0.095	1.000	0.492	
log_precipitacion DEN2,2	0.017	-0.043	0.019	-0.033	0.032	0.168	-0.006	-0.135	-0.105	0.492	1.000	

Figura 69: Correlaciones Función de Transferencia Precipitación

Correlaciones de las estimaciones de parámetro										
Variable Parámetro	O3 MA1,1	O3 MA2,1	O3 AR1,1	O3 AR1,2	O3 AR1,3	NO2 NUM1	NO2 NUM1,1	NO2 NUM1,2	NO2 DEN1,1	NO2 DEN2,1
O3 MA1,1	1.000	-0.027	-0.003	0.024	0.084	0.006	0.002	0.021	-0.010	0.015
O3 MA2,1	-0.027	1.000	-0.001	0.002	-0.000	-0.008	-0.008	-0.014	-0.010	0.015
O3 AR1,1	-0.003	-0.001	1.000	-0.582	0.038	-0.192	0.113	0.073	0.010	-0.001
O3 AR1,2	0.024	0.002	-0.582	1.000	-0.587	0.060	0.083	0.049	0.013	-0.005
O3 AR1,3	0.084	-0.000	0.038	-0.587	1.000	0.045	-0.131	-0.032	-0.066	-0.000
NO2 NUM1	0.006	-0.008	-0.192	0.060	0.045	1.000	-0.590	-0.397	0.041	-0.025
NO2 NUM1,1	0.002	-0.008	0.113	0.083	-0.131	-0.590	1.000	0.594	-0.024	-0.054
NO2 NUM1,2	0.021	-0.014	0.073	0.049	-0.032	-0.397	0.594	1.000	-0.015	-0.062
NO2 DEN1,1	-0.010	-0.010	0.010	0.013	-0.066	0.041	-0.024	-0.015	1.000	-0.054
NO2 DEN2,1	0.015	0.015	-0.001	-0.005	-0.000	-0.025	-0.054	-0.062	-0.054	1.000

Figura 70: Correlaciones Función de Transferencia NO_2

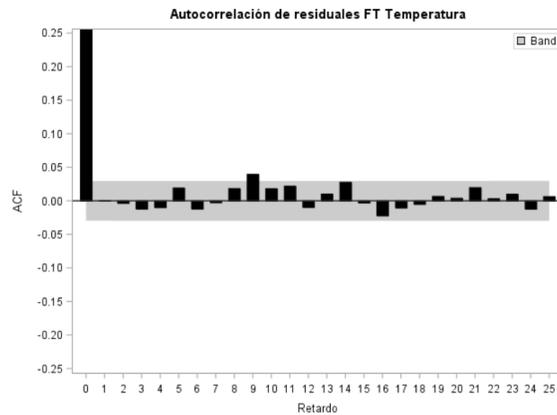


Figura 71: Residuales Modelo Función de Transferencia

Anexo II

Código Python Descomposición MSTL

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 from statsmodels.tsa.seasonal import MSTL
6 from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
7 from scipy.signal import periodogram
8
9 # Cargar el archivo .xls
10 df = pd.read_excel('datosMet_03_imputado.xls')
11 df = df.drop('FECHA', axis=1)
12
13 # Crear la columna de fecha con un rango de fechas
14 fecha_inicio = '2010-01-01'
15 fecha_fin = '2024-12-31'
16 fechas = pd.date_range(start=fecha_inicio, end=fecha_fin, freq='D')
17
18 # Asegurarnos de que el n mero de filas coincida con el n mero de fechas
19 if len(fechas) == len(df):
20     df['Fecha'] = fechas
21 else:
22     print("Error: La longitud de las fechas no coincide con la longitud de los
23         datos.")
24
25 # Convertir la columna de fecha a datetime
26 df['Fecha'] = pd.to_datetime(df['Fecha'])
27
28 # Establecer la columna de fecha como ndice
29 df.set_index('Fecha', inplace=True)
30 df['Fecha'] = fechas
31
32 # Crear la descomposicion MSTL
33 mstl = MSTL(df['03'], periods=[7, 365])
34 result = mstl.fit()
35
36 plt.figure(figsize=(400, 300))
37 result.plot()
38 plt.show()
```

Código SAS Modelo ARIMA Ozono

```
1
2 * Modelo 2;
3 proc arima data=Datostotales2024 plots(unpack)=all;
4     identify var=03(365);
5     estimate p=(1,2,3) q=(7)(365) method=ml;
6     forecast back=365 lead=365 out=pred2 ;
```

```
7 run;
8
9
10 /* Calculo del SSEp */
11 proc arima data=Datostotales2023;
12     identify var=O3(365);
13     estimate p=(1,2,3,7) q=(365) method=ml;
14     forecast lead=365 OUT=s1;
15     estimate p=(1,2,3) q=(7)(365) method=ml;
16     forecast lead=365 OUT=s2;
17 run; quit;
18
19 data w; set Datostotales2024; keep O3; if _N_ > 5113; run;
20 data w1; set s1; keep FORECAST; if _N_ > 5113; rename FORECAST=FORE1; run;
21 data w2; set s2; keep FORECAST; if _N_ > 5113; rename FORECAST=FORE2; run;
22
23 data x; merge w w1 w2;
24     res1 = O3 - FORE1;
25     res2 = O3 - FORE2;
26     ssep1 + res1 * res1;
27     ssep2 + res2 * res2;
28 run;
29
30
31 **** Gr fíco Predicciones 2024 ****;
32 data pred1_2024;
33     set pred1(rename=(forecast=pred1 L95=icinf1 U95=icsup1));
34     if _N_ > (5479 - 365);
35 run;
36
37 data pred2_2024;
38     set pred2(rename=(forecast=pred2 L95=icinf2 U95=icsup2));
39     if _N_ > (5479 - 365);
40 run;
41
42
43 /* Fechas para el año 2024 */
44 data fechas_2024;
45     format Fecha date9.;
46     do i = 1 to 365;
47         Fecha = '01JAN2024'd + i - 1;
48         output;
49     end;
50 run;
51
52 /* Seleccionar los últimos 365 datos de O3 */
53 data O3_2024;
54     set Datostotales2024(keep=Fecha O3);
55     if _N_ > (5479 - 365); /* Seleccionamos las últimas 365 observaciones */
56 run;
57
58
59 data pred_total_2024;
60     merge fechas_2024
```

```
61     pred1_2024(keep=pred1 icinf1 icsup1)
62     pred2_2024(keep=pred2 icinf2 icsup2)
63     O3_2024(keep=Fecha O3);
64 run;
65
66
67
68 ods html style=journal;
69
70 proc sgplot data=pred_total_2024;
71     title "Comparacion de predicciones de modelos ARIMA para 2024";
72
73     band x=Fecha lower=icinf1 upper=icsup1 /
74         fillattrs=(color=lightred transparency=0.85)
75         legendlabel="IC Modelo 1";
76     series x=Fecha y=pred1 /
77         lineattrs=(color=red thickness=1.5)
78         legendlabel="Modelo 1";
79
80     band x=Fecha lower=icinf2 upper=icsup2 /
81         fillattrs=(color=lightblue transparency=0.85)
82         legendlabel="IC Modelo 2";
83     series x=Fecha y=pred2 /
84         lineattrs=(color=blue thickness=1.5)
85         legendlabel="Modelo 2";
86
87     scatter x=Fecha y=O3 /
88         markerattrs=(symbol=circlefilled color=grayAA size=6)
89         legendlabel="Datos reales";
90
91     xaxis label="Fecha" labelattrs=(size=12);
92     yaxis label="Concentracion de O3" labelattrs=(size=12);
93 run;
94
95
96 ods html close;
```

Código SAS Modelo Temperatura Media

```
1
2 /* Diferenciaciones */
3 proc arima data=Datostotales2024 plots(unpack)=all;
4     identify var=temperatura_media;
5     identify var=temperatura_media(365);
6     identify var=temperatura_media(1,365);
7     run;
8
9 /* Modelo 1 */
10 proc arima data=Datostotales2023 plots(unpack)=all;
11     identify var=temperatura_media(1,365);
12     estimate p=(1) q=(1,2)(3,365) method=ml noint;
13     forecast back=365 lead=365 out=pred1;
```

```
14 run;quit;
15
16
17 /* Calculo SSEp y RSME% */
18
19 proc arima data=Datostotales2023;
20     identify var=temperatura_media(365);
21     estimate p=(1) q=(1,2)(3,365) method=ml noint;           *modelo 1;
22     forecast lead=365 OUT=s1;
23     estimate p=1 q=(1,2)(3,365,730) method=ml noint;       *modelo 2;
24     forecast lead=365 OUT=s2;
25 run;
26
27 data w;set Datostotales2024;keep temperatura_media;if _N_>5113;
28 run;
29 data w1;set s1;keep FORECAST; IF _N_>5113;RENAME FORECAST=FORE1;
30 RUN;
31 data w2;set s2;keep FORECAST; IF _N_>5113;RENAME FORECAST=FORE2;
32 RUN;
33
34 data x;merge w w1 w2;
35     res1=temperatura_media-FORE1;
36     res2=temperatura_media-FORE2;
37     ssep1+res1*res1;
38     ssep2+res2*res2;
39 RUN;
40
41 data x;
42     merge w w1;
43     res1 = temperatura_media - FORE1;
44     ap1 = .;
45     if temperatura_media ne 0 then do;
46         ap1 = abs(res1 / temperatura_media) * 100;
47     end;
48     res1_sq = res1**2;
49 run;
50
51 proc means data=x mean;
52     var res1_sq temperatura_media;
53     output out=stats mean=mean_res1_sq mean_real;
54 run;
55
56 data rmsep;
57     set stats;
58     rmse1 = sqrt(mean_res1_sq);
59     rmsep1 = (rmse1 / mean_real) * 100;
60 run;
```

Código SAS Función de Transferencia

```
1 proc arima data=Datostotales2023 plots(unpack)=all;
2     identify var=03(365) crosscorr=(temperatura_media(1,365)) nlag=50;
```

```
3 estimate p=(1,2,3) q=(7)(365)
4 input=(1 $ (1)/(3,365,730) temperatura_media )
5 method=ml
6 noint;
7 forecast out=salida1;
8 run;quit;
```

