

Ajuste y predicción mediante metodología Box-Jenkins para los niveles de contaminante NO_2 en la estación de Cuatro Caminos

Jorge Bengoa Pinedo
Universidad Politécnica de Madrid

Sergio Izquierdo Morona
Universidad Complutense de Madrid

Bogurad Barański Barańska
Universidad Politécnica de Madrid

Máster TECI - Series Temporales

2 de febrero de 2026

Resumen

En este trabajo se analiza la evolución temporal de la concentración de dióxido de nitrógeno (NO_2) en la estación de monitoreo de Cuatro Caminos (Madrid), utilizando datos mensuales desde 2010 hasta 2025. El objetivo principal es modelar el comportamiento del contaminante y generar predicciones precisas que sirvan de apoyo a la gestión ambiental. Metodológicamente, se ha realizado un preprocesamiento exhaustivo que incluye transformación de Box-Cox y diferenciación estacional. Se han contrastado dos enfoques: modelos deterministas de alisado exponencial y modelos estocásticos bajo la metodología Box-Jenkins (SARIMA). Los resultados indican que el modelo de Holt-Winters aditivo ofrece la mejor capacidad de generalización, obteniendo un RMSE de (2,88, $\mu g/m^3$) en el conjunto de validación, superando a los modelos SARIMA ajustados, los cuales presentaron dificultades para cumplir estrictamente la hipótesis de normalidad en los residuos. Se concluye que, pese a la complejidad estocástica de la serie, los métodos de suavizado resultan robustos frente a las irregularidades históricas de los datos, capturando eficazmente la estacionalidad anual asociada al tráfico urbano.

Palabras clave: Serie temporal, Contaminación, Metodología Box-Jenkins, Cuatro Caminos.

1. Introducción

En este trabajo se estudia la serie temporal del nivel de contaminante del dióxido de nitrógeno (NO_2) en la estación de medición de cuatro caminos en $\mu g/m^3$. Para ello, se accede a la base de datos de contaminantes del Ayuntamiento de Madrid, Ayuntamiento de Madrid (2026), en la cual se recogen los datos horarios día a día desde 2001 para todas las estaciones de contaminante de la ciudad. En esta base de datos la estación 38 es la que se corresponde a la de Cuatro Caminos y el contaminante número 10 es el que se corresponde con el NO_2 .

La elección NO_2 como objeto de estudio en esta serie temporal no es arbitraria, sino que se escoge por su impacto crítico en la salud pública urbana. La exposición a este contaminante está directamente relacionada con enfermedades respiratorias crónicas, inflamación de las vías aéreas y una reducción de la función pulmonar, afectando desproporcionadamente a grupos vulnerables como niños y ancianos tal como se recoge en World Health Organization (2021). Debido a estos peligros, la Organización Mundial de la Salud ha endurecido recientemente sus directrices, recomendando que las concentraciones medias anuales no superen los $10 \mu g/m^3$ y estableciendo un límite de $25 \mu g/m^3$ para la exposición diaria.

En el contexto europeo y español, la situación es especialmente preocupante. Según el informe más reciente del Centro Temático Europeo sobre Salud Humana y Medio Ambiente, la contaminación del aire continúa siendo el mayor riesgo ambiental para la salud en Europa, con una carga de enfermedad significativa atribuible específicamente al NO_2 , Soares et al. (2025) como se puede ver en la figura 1. Puesto que el tráfico rodado constituye la principal fuente de emisión de NO_2 en entornos urbanos, Madrid se erige como un caso de estudio idóneo para analizar la evolución temporal del contaminante y evaluar la eficacia de las políticas de mitigación implementadas en la capital.

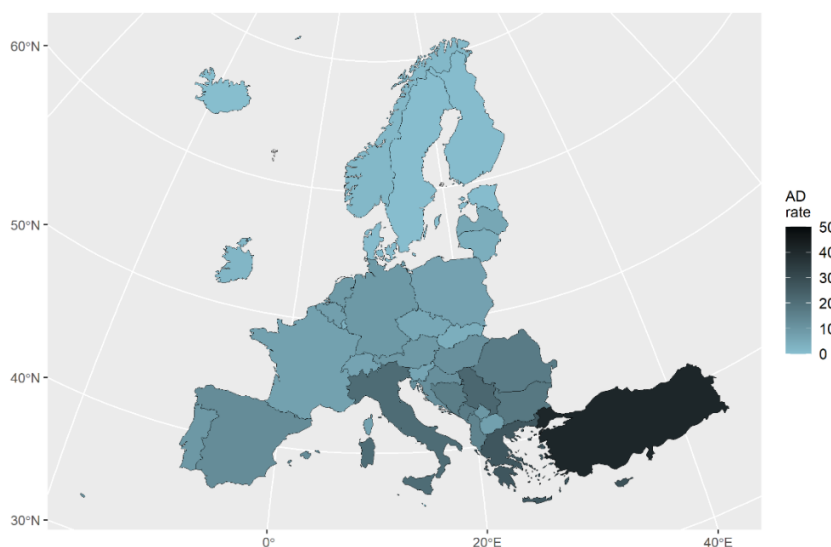


Figura 1: Impacto en la salud pública: Tasa de mortalidad por todas las causas atribuibles a la exposición al NO_2 (Muertes por cada 100.000 habitantes) en el año 2023.

Para analizar esta serie temporal y su comportamiento en la ciudad, la literatura reciente ha desplegado un abanico de técnicas que van desde la metodología Box-Jenkins vista en clase hasta métodos basados en redes neuronales.

En el ámbito de los modelos estocásticos, en Avila et al. (2023) abordaron la predicción

mediante modelos ARIMA, destacando la necesidad de modelar explícitamente las múltiples estacionalidades (diaria, semanal y anual) para obtener pronósticos robustos. En otros trabajos como Orta Blanco (2024) se utilizaron Procesos Gaussianos para estudiar el impacto de la zona de bajas emisiones Madrid Central. Este enfoque probabilístico se utiliza para cuantificar la incertidumbre del modelo y evaluar la efectividad de intervenciones públicas en la calidad del aire.

Por otro lado, según la literatura por la complejidad no lineal de los datos ha impulsado el uso de redes neuronales. En Iskandaryan et al. (2022) propusieron un modelo de aprendizaje profundo que integra redes convolucionales y LSTM bidireccionales, capturando simultáneamente dependencias espaciales y temporales. En una línea híbrida reciente, Fernández-Avilés et al. (2024) desarrollaron un modelo de Redes Neuronales Autorregresivas Aumentadas por Factores. Su metodología mejora la capacidad predictiva al sintetizar la información de múltiples variables en factores latentes antes de alimentar la red neuronal.

Finalmente, enfoques alternativos han optado por tratar la evolución diaria de la contaminación como un objeto matemático continuo. En Sancho Val et al. (2026) aplicaron el Análisis de Datos Funcionales mediante B-splines. Esta técnica permite analizar la curva diaria de NO_2 en su totalidad, facilitando la detección de patrones anómalos que los métodos discretos tradicionales podrían pasar por alto.

2. Metodología

En este trabajo se aborda el análisis de la serie temporal mediante dos enfoques metodológicos clásicos. En un primer lugar, se asume un enfoque determinista basado en técnicas de alisado exponencial (suavizados). Posteriormente, se adopta una perspectiva estocástica para el ajuste de modelos SARIMA (Seasonal Autoregressive Integrated Moving Average).

Aún así, antes de la calibración de cualquiera de estos modelos, se llevará a cabo un Análisis Exploratorio de Datos (EDA) exhaustivo para comprender la estructura subyacente de la serie. Cabe destacar que, aunque la serie sea horaria para facilitar el análisis y estabilizar la serie se trabaja con las medias mensuales de la serie, fijando el periodo estacional en $s = 12$.

2.1. Pipeline de trabajo y métricas de evaluación

Para evaluar la capacidad de generalización de los modelos propuestos, se ha procedido a particionar los datos en dos subconjuntos independientes: un conjunto de entrenamiento, que abarca el periodo $2010 \leq t < 2024$, y un conjunto de validación, correspondiente al intervalo $2024 \leq t \leq 2025$.

Es fundamental destacar el rigor necesario en el preprocesamiento para evitar la filtración de datos. Todos los parámetros de las transformaciones necesarias (como el parámetro λ de la transformación Box-Cox o los órdenes de diferenciación) se estiman exclusivamente sobre el conjunto de entrenamiento. Posteriormente, estas transformaciones ya calibradas se aplican directamente sobre el conjunto de validación.

Para cuantificar la bondad de ajuste y comparar el desempeño predictivo, se utilizará la Raíz del Error Cuadrático Medio (RMSE). La elección de esta métrica se justifica por dos razones principales:

- **Interpretabilidad dimensional:** Al aplicar la raíz cuadrada, el error se expresa en las

mismas unidades físicas que la variable objetivo (concentración de NO_2 en $\mu g/m^3$), facilitando su lectura directa.

- **Penalización de grandes desvíos:** Dada la naturaleza cuadrática del error antes de la raíz, esta métrica penaliza de manera severa los errores de gran magnitud. Esto es crítico en el contexto de la calidad del aire, donde no detectar un pico de contaminación es más grave que pequeños errores en niveles bajos.

La fórmula empleada es:

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

donde n denota el tamaño de la muestra de validación, y_i es el valor observado de la serie en el momento i , e \hat{y}_i representa la predicción generada por el modelo para ese mismo instante.

2.2. Análisis exploratorio de los datos

Metodológicamente, el estudio comienza con una descomposición aditiva-multiplicativa preliminar de la serie temporal. El objetivo de esta fase es inspeccionar visualmente la estructura subyacente de los datos, permitiendo identificar la presencia de tendencias deterministas y patrones estacionales marcados. Adicionalmente, la inspección gráfica permite una evaluación cualitativa de la heterocedasticidad, sugiriendo la necesidad de estabilizar la varianza.

Para formalizar esta decisión, se calcula el parámetro óptimo λ de la transformación de Box-Cox que maximiza la verosimilitud logarítmica. Se establece un criterio de decisión práctico para mantener la interpretabilidad del modelo: si el valor estimado se encuentra próximo a la identidad, concretamente si $\lambda \in [0,8, 1,2]$, se opta por no transformar la serie original. En caso contrario, es decir, si $\lambda \in (-\infty, 0,8) \cup (1,2, \infty)$, se aplica la transformación para garantizar la homocedasticidad requerida por los modelos lineales. Cabe mencionar que la decisión final se valida iterativamente según la calidad de los residuos del modelo.

Posteriormente, se aborda el requisito de estacionariedad mediante el análisis de diferenciación. Siguiendo la metodología estándar de Box-Jenkins (Box et al., 2015):

1. En primer lugar, se evalúa la necesidad de diferenciación estacional ∇_{12}^D dada la periodicidad mensual de la serie.
2. Una vez aplicada la diferenciación estacional, si la serie resultante aún presenta tendencia, se aplican diferenciaciones regulares ∇^d sucesivas hasta eliminarla.

Tras aplicar las diferenciaciones estacionales, se evalúa la estacionariedad de la serie mediante la prueba de Dickey-Fuller (Dickey & Fuller, 1979). Este contraste se formaliza a través del siguiente contraste de hipótesis:

$$\begin{cases} H_0 : & \text{(Existe raíz unitaria, la serie no es estacionaria)} \\ H_1 : & \text{(No existe raíz unitaria, la serie sí es estacionaria)} \end{cases} \quad (2)$$

Finalmente, cabe señalar que el tratamiento específico de valores atípicos y el análisis de intervención se reservan para una segunda etapa de refinamiento. Estos solo se incorporarán al modelo en caso de que el ajuste inicial no supere los diferentes tests de hipótesis.^{x⁰}

2.3. Ajuste de modelos deterministas: Métodos de Alisado Exponencial

Esta sección se elabora con ayuda del libro de Peña (2010). Para los suavizados se utiliza un suavizado exponencial que permite suavizar las fluctuaciones a corto plazo y resaltar la ciclo-tendencia. Para ello, se asignan ponderaciones a las observaciones que decaen exponencialmente de manera que los valores más recientes tienen mayor impacto en la estimación.

La elección del algoritmo de alisado específico depende de la estructura de la serie, la cual está intrínsecamente relacionada con las necesidades de diferenciación detectadas en la fase exploratoria:

- **Ausencia de tendencia y estacionalidad (Nivel constante):** Si la serie no requiere diferenciaciones, se aplica el Alisado Exponencial Simple. Este método actualiza el nivel de la serie (a_t) mediante un parámetro de suavizado α (donde típicamente $0,05 \leq \alpha \leq 0,3$):

$$a_t = \alpha x_t + (1 - \alpha)a_{t-1}, \quad a_1 = x_1 \quad (3)$$

Donde la predicción a un paso es

$$\hat{x}_{t+h} = a_t \quad (4)$$

- **Presencia de tendencia (Diferenciación regular):** Si la serie presenta una tendencia lineal (requiere $\nabla^d x_t$), se emplea el Método de Holt (Suavizado Doble). Este modelo introduce una segunda ecuación para estimar la pendiente o tendencia local (b_t) mediante un parámetro β :

$$\begin{aligned} a_t &= \alpha x_t + (1 - \alpha)(a_{t-1} + b_{t-1}), \quad 0 < \alpha < 1 \\ b_t &= \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}, \quad 0 < \beta < 1 \end{aligned} \quad (5)$$

Las predicciones se generan proyectando la tendencia:

$$\hat{x}_{t+h} = a_t + hb_t \quad (6)$$

- **Presencia de estacionalidad y tendencia (Diferenciación estacional y regular):** Si la serie requiere diferenciación estacional (∇_s^D), se aplica el Método de Holt-Winters (Suavizado Triple). Este enfoque generaliza el anterior añadiendo una tercera ecuación y un parámetro γ para actualizar los índices estacionales, permitiendo capturar patrones cíclicos como el comportamiento mensual del NO_2 .

$$\begin{aligned} a_t &= \alpha x_t / R_{t-s} + (1 - \alpha)(a_{t-1} + b_{t-1}), \quad 0 < \alpha < 1 \\ b_t &= \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}, \quad 0 < \beta < 1 \\ R_t &= \gamma(x_t / a_t) + (1 - \gamma)R_{t-s}, \quad 0 < \gamma < 1 \end{aligned} \quad (7)$$

Las predicciones se generan proyectando la tendencia:

$$\hat{x}_{t+h} = a_t + hb_t + R_{t-s+h} \quad (8)$$

Aunque se ajuste el modelo en función de las diferenciaciones realizadas, como el coste computacional no es muy elevado se entrena cada uno de los suavizados en el conjunto de entrenamiento y se elige el que tiene menor RMSE en el conjunto de validación. El que tenga un error menor en esta métrica es el modelo elegido para realizar las predicciones a futuro.

2.4. Ajuste de modelos estocásticos basados en modelos SARIMA

Siguiendo la metodología descrita en Box et al. (2015), el modelado estocástico se estructura en un proceso iterativo de tres etapas: identificación, estimación y validación. Una vez determinadas las transformaciones necesarias para garantizar la estacionariedad (diferenciación regular d y estacional D , analizadas en el EDA), el objetivo se centra en identificar los órdenes de la parte autorregresiva y de medias móviles.

La estructura general de un modelo estocástico estacional SARIMA(p, d, q)(P, D, Q) $_s$ viene definida por la ecuación:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D X_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (9)$$

donde:

- s : es el periodo estacional (fijado en $s = 12$ para datos mensuales).
- $\phi_p(B)$ y $\theta_q(B)$: son los polinomios de la parte regular (AR y MA).
- $\Phi_P(B^s)$ y $\Theta_Q(B^s)$: son los polinomios de la parte estacional.
- $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$: es el término de error (ruido blanco).

Para la identificación de los órdenes (p, q, P, Q) se examinan los correlogramas estimados, aplicando el principio de parsimonia (buscar el modelo más simple posible, con órdenes ≤ 3):

- **Parte Regular - Parámetro p (AR):** Se determina inspeccionando la Función de Autocorrelación Parcial (FAP). El orden p se estima contando el número de barras consecutivas significativas iniciales. Es decir, si se observan k barras que superan las bandas de confianza seguidas de un corte brusco, se sugiere un modelo AR($p = k$). Por el contrario, si las barras presentan un decaimiento suave sin cortarse, indica la presencia de un componente de Medias Móviles (MA).
- **Parte Regular - Parámetro q (MA):** Se determina inspeccionando la Función de Autocorrelación Simple (FAS). El orden q corresponde al número de barras consecutivas significativas (omitiendo la barra del retardo 0). Si se encuentran k barras significativas seguidas de un corte, se sugiere un modelo MA($q = k$). Si, en cambio, se observa un decaimiento lento en las barras, es indicativo de un proceso Autorregresivo (AR).
- **Parte Estacional - Parámetro P (SAR):** Se analizan las barras situadas en los múltiplos del periodo estacional ($s, 2s, 3s \dots$) en la FAP. El orden P será igual al número de barras estacionales consecutivas que sean significativas. Por ejemplo, si solo la barra en $s = 12$ es significativa y la de $2s = 24$ no, se sugiere un $P = 1$.
- **Parte Estacional - Parámetro Q (SMA):** Se analizan las barras situadas en los múltiplos del periodo estacional ($s, 2s, 3s \dots$) en la FAS. El orden Q se estima contando cuántas de estas barras estacionales son significativas antes de que la correlación se anule. Típicamente, una única barra significativa en el lag s indica un $Q = 1$.

A partir de la identificación preliminar basada en el análisis cualitativo de los correlogramas, se procede a una etapa de refinamiento iterativo. Dado que el espacio de búsqueda es acotado con $3^4 = 81$ posibles modelos, se ejecuta una exploración exhaustiva sobre las combinaciones de los órdenes p, q, P y Q , manteniendo constantes los grados de diferenciación (d, D) determinados en la fase exploratoria.

El criterio de selección en esta etapa es estricto en cuanto a la robustez estadística. Únicamente se consideran válidos aquellos modelos en los que todos sus coeficientes estimados son estadísticamente significativos. Para ello, se verifica que el p-valor asociado a cada coeficiente sea inferior al nivel de significación establecido ($\alpha = 0,05$), rechazando cualquier especificación que incluya parámetros redundantes o no explicativos.

Una vez seleccionados aquellos modelos cuyos coeficientes resultan estadísticamente significativos, se procede a la validación de los supuestos de los residuos. Para ello, se realizan los siguientes contrastes de hipótesis sobre los residuos ($\hat{\varepsilon}_t$) de los modelos ajustados donde se buscan p-valores superiores a α :

- **Comprobación de independencia (Test de Ljung-Box)** (LJUNG & BOX, 1978). Se busca verificar la ausencia de autocorrelación serial en los residuos hasta un retardo k :

$$\begin{cases} H_0 : \rho_1 = \dots = \rho_k = 0 & \text{(Independencia: Existe Ruido Blanco)} \\ H_1 : \exists \rho_j \neq 0 & \text{(Existe autocorrelación, el modelo es incompleto)} \end{cases} \quad (10)$$

- **Comprobación de homocedasticidad (Test de Breusch-Pagan)** (Breusch & Pagan, 1979). Se evalúa si la varianza de los residuos es constante en el tiempo:

$$\begin{cases} H_0 : \text{Var}(\varepsilon_t) = \sigma^2 & \text{(Homocedasticidad: Varianza constante)} \\ H_1 : \text{Var}(\varepsilon_t) \neq \sigma^2 & \text{(Heterocedasticidad: Varianza volátil)} \end{cases} \quad (11)$$

- **Comprobación de normalidad (Test de Kolmogorov-Smirnov corrección Lilliefors)** (Massey, 1951). Se contrasta si la distribución empírica de los residuos se ajusta a una distribución Normal:

$$\begin{cases} H_0 : \hat{\varepsilon}_t \sim N(\mu, \sigma^2) & \text{(Los residuos siguen una distribución Normal)} \\ H_1 : \hat{\varepsilon}_t \not\sim N(\mu, \sigma^2) & \text{(Los residuos NO siguen una distribución Normal)} \end{cases} \quad (12)$$

Por último, se calcula la métrica del RMSE para todos los modelos con coeficientes significativos aunque no pasen todos los tests de hipótesis. De esta manera, se pueden comparar los modelos y verificar si no pasar alguno de los test es crítico para el funcionamiento del modelo.

3. Resultados

En esta sección se presentan los resultados obtenidos del análisis de la serie temporal:

3.1. Análisis exploratorio de los datos

El primer paso es dividir la serie temporal en entrenamiento y validación tal como se puede ver en la figura 2.

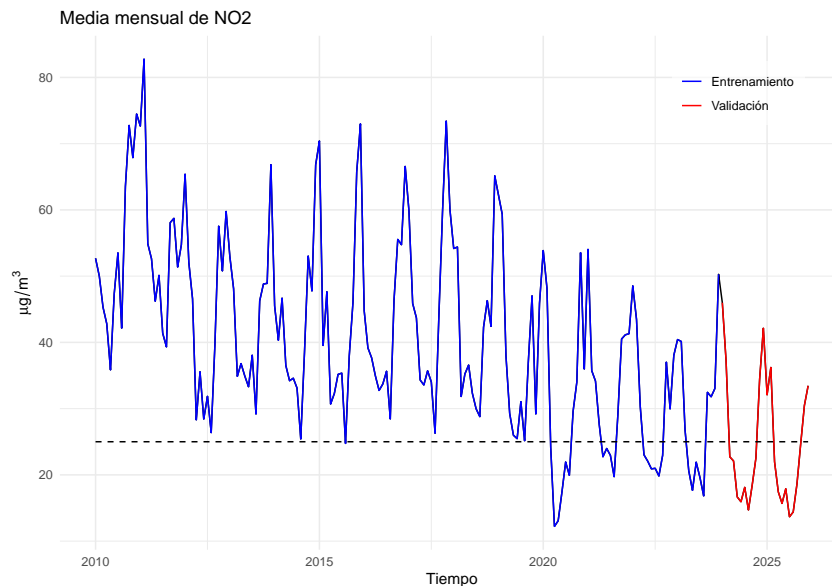


Figura 2: Representación de la serie temporal del NO_2 dividida en conjuntos de entrenamiento y test junto con la recta umbral del límite diario seguro de contaminante ($25\mu g/m^3$) según la Organización Mundial de la Salud.

Antes de proceder al modelado, se efectúa una descomposición de la serie mediante los métodos aditivo y multiplicativo. Aun así, de la función `decompose` solo se pueden obtener conclusiones sobre la tendencia de la serie debido a que la componente estacional no se calcula de manera fiable. En la Figura 3, correspondiente al esquema aditivo, se observa como la serie presenta una tendencia clara que justifica la necesidad de diferenciación en etapas posteriores.

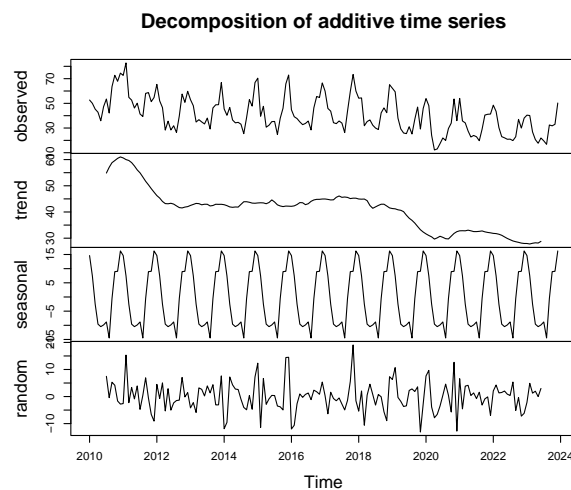


Figura 3: Decomposición de la serie mediante la función `decompose` en R.

La inspección visual de la serie original de la figura 2 revela indicios de heterocedasticidad, observándose específicamente una contracción de la varianza en el tramo final de la muestra. Al calcular el parámetro óptimo de Box-Cox, se obtiene un valor de $\lambda \approx 0,797$. Si bien este valor se sitúa en la frontera del intervalo de no-transformación, se decide aplicar la transformación priorizando la evidencia visual. Como se aprecia en la Figura 4, la serie transformada tiene un comportamiento mucho más estable y homogéneo, lo cual facilitará el ajuste de los modelos posteriores.

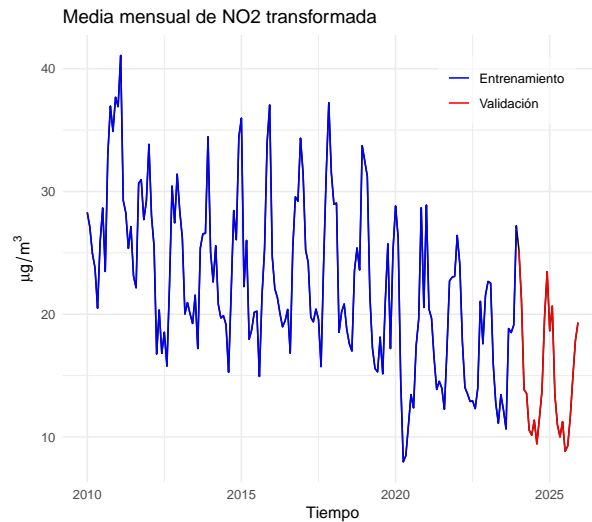


Figura 4: Serie original transformada mediante Box-Cox con $\lambda = 0,7968376$.

Se analiza la presencia de valores atípicos en el conjunto de entrenamiento para evaluar su posible impacto en el ajuste del modelo. Como muestra la Figura 5, se detectan dos anomalías estadísticas: una leve antes de 2018 y otra más pronunciada en 2021. Siguiendo el protocolo mencionado en la sección de metodología, se decide mantener estos valores en la muestra, reservando su posible tratamiento únicamente para casos en los que la calidad predictiva del modelo sea baja.

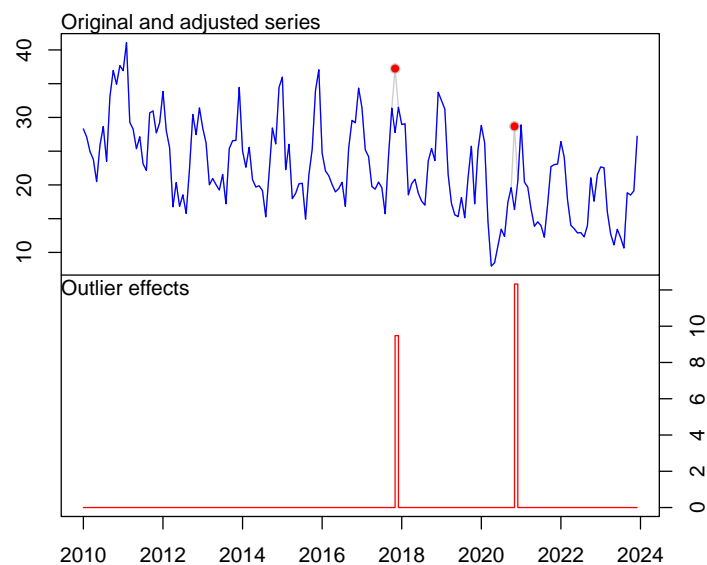


Figura 5: Representación de los posibles valores atípicos en el conjunto de entrenamiento.

Por último, el análisis exploratorio de los datos indica que se debe realizar una única diferenciación estacional y ninguna diferenciación regular. Para comprobar que, efectivamente, la serie ha sido desestacionalizada y es adecuada para el modelado, se realiza el Test de Dickey-Fuller donde se obtiene el siguiente resultado:

$$\text{p-valor} = 0,02345 < \alpha = 0,05 \quad (13)$$

Dado que el p-valor es menor que el nivel de significación, rechazamos la hipótesis nula, lo que indica que la serie transformada es **estacionaria**.

3.2. Ajuste de modelos deterministas basados en suavizados

De la teoría el único modelo de suavizado que debería ajustarse es el suavizado Holt-Winters debido a que existe componente estacional. No obstante, como computacionalmente no es costoso, se ajustan todos los suavizados para ver cuales es razonable estudiar en la validación. Los resultados de las predicciones se pueden ver en la figura 6.

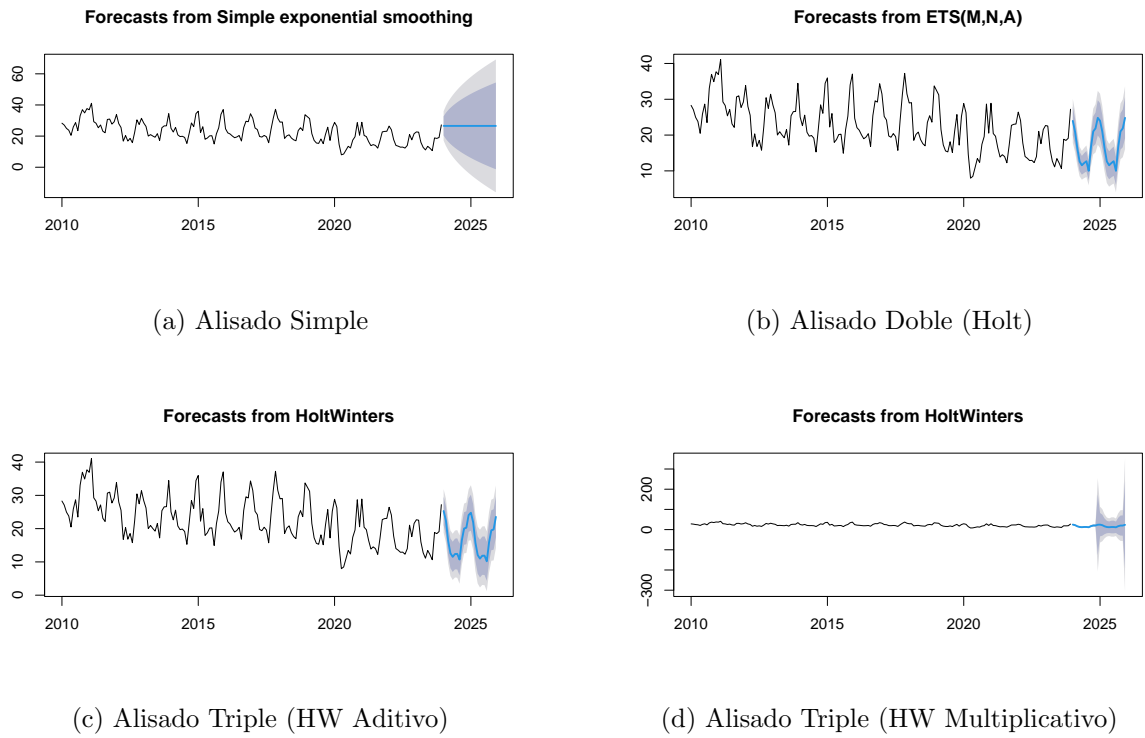


Figura 6: Comparación de los ajustes realizados: (a) muestra el ajuste mediante Alisado Simple donde se puede ver que no existe predicción, (b) muestra el alisado de Holt donde se ve que las predicciones son razonables (c) y (d) representan los ajustes estacionales de Holt-Winters en sus variantes aditiva y multiplicativa respectivamente donde claramente se puede ver como el enfoque multiplicativo no es bueno prediciendo debido a la gran varianza que tiene asociada.

Por tanto, se ha calculado el RMSE para los modelos de alisado de Holt y Holt-Winters aditivo, cuyos resultados se detallan en la Tabla 1. A la vista de las métricas, se concluye que el modelo óptimo es el Holt-Winters aditivo. Cabe destacar que este modelo presenta un mejor desempeño (menor error) en el conjunto de validación que en el de entrenamiento. La causa de este fenómeno se observa en la Figura 4: el inicio de la serie temporal muestra un comportamiento irregular que difiere de la tendencia general, lo cual penaliza el error promedio durante el entrenamiento. En contraste, el periodo de validación es más estable, permitiendo al modelo realizar predicciones más precisas.

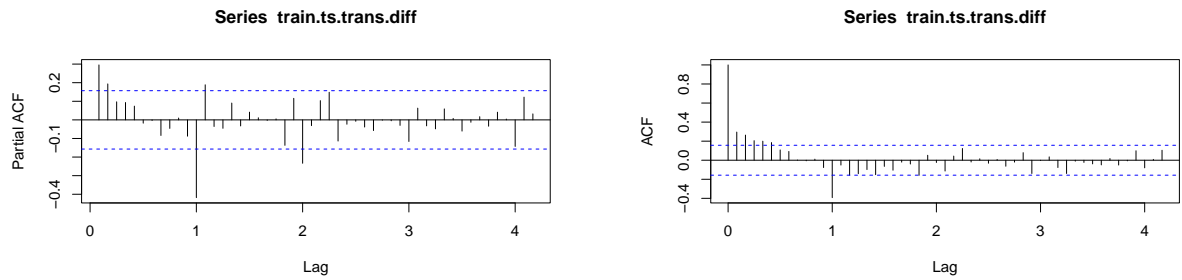
Tabla 1: Modelos de alisado escogidos.

| Modelo | Alpha | Beta | Gamma | RMSE Entrenamiento | RMSE Validación |
|----------------------|--------|--------|--------|--------------------|-----------------|
| Holt-Winters aditivo | 0.1790 | 0.0085 | 0.3250 | 3.3308 | 2.8808 |
| Holt | 0.2753 | 0.0001 | - | 3.0061 | 3.3431 |

3.3. Ajuste de modelos estocásticos basados en modelos SARIMA

En primer lugar, se ajusta el modelo manualmente a través la FAP y la FAS de la figura 7. Del análisis visual se obtiene el siguiente modelo SARIMA:

$$(2, 0, 0)(2, 1, 1)_{[12]} \quad (14)$$



(a) Función de Autocorrelación Parcial (FAP)

(b) Función de Autocorrelación Simple (FAS)

Figura 7: Correlogramas estimados de la serie estacionaria. La gráfica (a) muestra la FAP, utilizada para identificar el orden autorregresivo (p y P), mientras que la (b) muestra la FAS, utilizada para identificar el orden de medias móviles (q y Q).

Tal como se mencionó en la metodología una vez ajustado un modelo SARIMA inicial se realiza un barrido en todos los posibles modelos que tengan las mismas diferenciaciones para escoger aquel que tenga menor error con el conjunto de validación y que a su vez cumpla los tests de: independencia, homocedasticidad y normalidad. En la tabla 2 se pueden ver los resultados de ejecución y como ningún modelo cumple normalidad si se selecciona $\alpha = 0,05$. No obstante, las celdas en amarillo se encuentran en la zona de significancia marginal (p-valor entre 0.01 y 0.05). Si la normalidad de un modelo se encuentra en la misma y se cumplen además las hipótesis de independencia y homocedasticidad, será considerado válido. Es relevante destacar que el modelo inicial estimado no aparece debido a que tiene alguno de sus coeficientes no significativo.

Tabla 2: Resultados de los tests de hipótesis realizados en los modelos SARIMA. En rojo se marcan aquellos p-valores con significancia menor a 0,01, en verde aquellos con significancia mayor a 0,05 y en amarillo el resto.

| Modelo | Independencia | Homocedasticidad | Normalidad |
|-----------------------------|---------------|------------------|------------|
| $(0, 0, 0)(0, 1, 1)_{[12]}$ | 0.0000 | 0.1672 | 0.0006 |
| $(0, 0, 0)(1, 1, 0)_{[12]}$ | 0.0000 | 0.2488 | 0.0058 |
| $(0, 0, 0)(2, 1, 3)_{[12]}$ | 0.0000 | 0.1369 | 0.0017 |
| $(0, 0, 1)(0, 1, 0)_{[12]}$ | 0.0000 | 0.4620 | 0.0301 |
| $(0, 0, 1)(0, 1, 1)_{[12]}$ | 0.0001 | 0.5475 | 0.0006 |
| $(0, 0, 1)(1, 1, 0)_{[12]}$ | 0.0000 | 0.7799 | 0.0052 |
| $(0, 0, 1)(2, 1, 0)_{[12]}$ | 0.0001 | 0.7007 | 0.0005 |

| Modelo | Independencia | Homocedasticidad | Normalidad |
|------------------------------------|---------------|------------------|------------|
| (0, 0, 2)(0, 1, 0) _[12] | 0.0000 | 0.6523 | 0.0017 |
| (0, 0, 2)(0, 1, 1) _[12] | 0.0152 | 0.7426 | 0.0010 |
| (0, 0, 2)(1, 1, 0) _[12] | 0.0030 | 0.9804 | 0.0144 |
| (0, 0, 2)(2, 1, 0) _[12] | 0.0143 | 0.9332 | 0.0075 |
| (0, 0, 3)(0, 1, 1) _[12] | 0.0794 | 0.8809 | 0.0036 |
| (0, 0, 3)(2, 1, 0) _[12] | 0.0746 | 0.9121 | 0.0013 |
| (1, 0, 0)(0, 1, 0) _[12] | 0.0000 | 0.7017 | 0.0058 |
| (1, 0, 0)(0, 1, 1) _[12] | 0.0122 | 0.8926 | 0.0028 |
| (1, 0, 0)(1, 1, 0) _[12] | 0.0004 | 0.6947 | 0.0017 |
| (1, 0, 0)(2, 1, 0) _[12] | 0.0504 | 0.7125 | 0.0026 |
| (1, 0, 1)(0, 1, 0) _[12] | 0.0001 | 0.9166 | 0.0007 |
| (1, 0, 1)(0, 1, 1) _[12] | 0.3251 | 0.8681 | 0.0026 |
| (1, 0, 1)(1, 1, 0) _[12] | 0.1118 | 0.6236 | 0.0139 |
| (1, 0, 1)(2, 1, 0) _[12] | 0.6094 | 0.6972 | 0.0075 |
| (2, 0, 0)(0, 1, 0) _[12] | 0.0000 | 0.9473 | 0.0002 |
| (2, 0, 0)(0, 1, 1) _[12] | 0.0485 | 0.7601 | 0.0082 |
| (2, 0, 0)(1, 1, 0) _[12] | 0.0562 | 0.5643 | 0.0101 |
| (2, 0, 0)(2, 1, 0) _[12] | 0.1683 | 0.6077 | 0.0033 |
| (2, 0, 1)(2, 1, 3) _[12] | 0.0382 | 0.8533 | 0.0041 |
| (3, 0, 0)(0, 1, 1) _[12] | 0.0264 | 0.6382 | 0.0126 |
| (3, 0, 0)(2, 1, 0) _[12] | 0.1897 | 0.5170 | 0.0297 |
| (3, 0, 0)(2, 1, 3) _[12] | 0.0271 | 0.5754 | 0.0314 |
| (3, 0, 2)(2, 1, 0) _[12] | 0.4104 | 0.6916 | 0.0172 |

A continuación, en la tabla 3 se recogen los valores del error cuadrático medio de los distintos modelos candidatos. Entre estos, solo 4 candidatos cumplen todas las hipótesis (normalidad relajada) y se seleccionan en magenta aquellos con menor valor del error.

Tabla 3: Resultados de los tests de hipótesis realizados en los modelos SARIMA junto a su error. En rojo se marcan aquellos p-valores con significancia menor a 0,01, en verde aquellos con significancia mayor a 0,05 y en amarillo el resto.

| Modelo | Independencia | Homocedasticidad | Normalidad | RMSE (train) | RMSE (val) |
|------------------------------------|---------------|------------------|------------|--------------|------------|
| (0, 0, 0)(0, 1, 1) _[12] | NO | SÍ | NO | 4.0531 | 3.5227 |
| (0, 0, 0)(1, 1, 0) _[12] | NO | SÍ | NO | 4.1000 | 3.2865 |
| (0, 0, 0)(2, 1, 3) _[12] | NO | SÍ | NO | 3.9201 | 3.5917 |
| (0, 0, 1)(0, 1, 0) _[12] | NO | SÍ | NO | 4.2458 | 3.5412 |
| (0, 0, 1)(0, 1, 1) _[12] | NO | SÍ | NO | 3.7374 | 3.8851 |
| (0, 0, 1)(1, 1, 0) _[12] | NO | SÍ | NO | 3.8509 | 3.3284 |
| (0, 0, 1)(2, 1, 0) _[12] | NO | SÍ | NO | 3.7432 | 3.5380 |
| (0, 0, 2)(0, 1, 0) _[12] | NO | SÍ | NO | 4.1365 | 3.6381 |
| (0, 0, 2)(0, 1, 1) _[12] | NO | SÍ | NO | 3.5665 | 4.1563 |
| (0, 0, 2)(1, 1, 0) _[12] | NO | SÍ | NO | 3.6802 | 3.4406 |
| (0, 0, 2)(2, 1, 0) _[12] | NO | SÍ | NO | 3.6035 | 3.5702 |
| (0, 0, 3)(0, 1, 1) _[12] | SÍ | SÍ | NO | 3.5081 | 4.2694 |
| (0, 0, 3)(2, 1, 0) _[12] | SÍ | SÍ | NO | 3.5580 | 3.5936 |
| (1, 0, 0)(0, 1, 0) _[12] | NO | SÍ | NO | 4.1648 | 3.6213 |
| (1, 0, 0)(0, 1, 1) _[12] | NO | SÍ | NO | 3.4760 | 4.7911 |

| Modelo | Independencia | Homocedasticidad | Normalidad | RMSE (train) | RMSE (val) |
|------------------------------------|---------------|------------------|------------|--------------|------------|
| (1, 0, 0)(1, 1, 0) _[12] | NO | SÍ | NO | 3.6973 | 3.4407 |
| (1, 0, 0)(2, 1, 0) _[12] | SÍ | SÍ | NO | 3.5537 | 3.7018 |
| (1, 0, 1)(0, 1, 0) _[12] | NO | SÍ | NO | 3.9917 | 4.1054 |
| (1, 0, 1)(0, 1, 1) _[12] | SÍ | SÍ | NO | 3.0437 | 4.0382 |
| (1, 0, 1)(1, 1, 0) _[12] | SÍ | SÍ | NO | 3.4807 | 3.6146 |
| (1, 0, 1)(2, 1, 0) _[12] | SÍ | SÍ | NO | 3.3184 | 3.7311 |
| (2, 0, 0)(0, 1, 0) _[12] | NO | SÍ | NO | 4.0448 | 3.8182 |
| (2, 0, 0)(0, 1, 1) _[12] | NO | SÍ | NO | 3.3101 | 5.1379 |
| (2, 0, 0)(1, 1, 0) _[12] | SÍ | SÍ | NO | 3.5523 | 3.5725 |
| (2, 0, 0)(2, 1, 0) _[12] | SÍ | SÍ | NO | 3.4424 | 3.7190 |
| (2, 0, 1)(2, 1, 3) _[12] | NO | SÍ | NO | 3.1133 | 5.5767 |
| (3, 0, 0)(0, 1, 1) _[12] | NO | SÍ | NO | 3.2204 | 5.3714 |
| (3, 0, 0)(2, 1, 0) _[12] | SÍ | SÍ | NO | 3.3924 | 3.7661 |
| (3, 0, 0)(2, 1, 3) _[12] | NO | SÍ | NO | 2.9320 | 6.4224 |
| (3, 0, 2)(2, 1, 0) _[12] | SÍ | SÍ | NO | 3.3215 | 3.7474 |

En la tabla 4 se detallan los modelos SARIMA escogidos junto a sus ecuaciones obtenidas directamente a partir de los coeficientes de R en `fitArima`¹.

| Modelo | Ecuación | RMSE (train) | RMSE (val) |
|--------------------|---|--------------|------------|
| (2, 0, 0)(1, 1, 0) | $(1 - 0,3201B^1 - 0,2784B^2)(1 + 0,4758B^{1s})\nabla_s^1 X_t = \epsilon_t$ | 3.5523 | 3.5725 |
| (1, 0, 1)(1, 1, 0) | $(1 - 0,9086B^1)(1 + 0,4894B^{1s})\nabla_s^1 X_t = (1 + 0,6426B^1)\epsilon_t$ | 3.4807 | 3.6146 |
| (3, 0, 2)(2, 1, 0) | $(1 - 0,1276B^1 - 0,6914B^2 - 0,0362B^3) \cdot (1 + 0,6231B^{1s} + 0,2932B^{2s})\nabla_s^1 X_t = (1 - 0,1395B^1 + 0,4984B^2)\epsilon_t$ | 3.3215 | 3.7474 |

Tabla 4: Modelos ARIMA escogidos para el estudio comparativo.

¹Nota sobre la parametrización en R: Para convertir la salida de la función `fitArima` a la ecuación teórica $\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D X_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t$, debe tenerse en cuenta que R utiliza una parametrización específica (Hyndman & Athanasopoulos, 2018, Cap. 8.7):

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \dots - \Phi_P B^{sP})y_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ})\epsilon_t \quad (15)$$

- **Términos AR (ϕ):** Los coeficientes `ar` en R corresponden directamente a los parámetros ϕ_i del polinomio autorregresivo $\phi_p(B)$.
- **Términos MA (θ):** Los coeficientes `ma` en R corresponden a θ_i en el polinomio de media móvil $\theta_q(B)$.
- **Términos SAR (Φ):** Los coeficientes etiquetados como `sar` en la salida de R corresponden a los parámetros Φ_i del componente autorregresivo estacional $\Phi_P(B^s)$.
- **Términos SMA (Θ):** Los coeficientes etiquetados como `sma` corresponden a los parámetros Θ_i del componente de media móvil estacional $\Theta_Q(B^s)$.

4. Discusión

Una vez obtenidos los resultados anteriores se entrenan ahora 4 modelos predictivos, uno basado en el suavizado de Holt-Winters y otros 3 basados en modelos ARIMA. Pues como se puede ver si se comparan las tablas 1 y 4 el modelo determinista mediante suavizados tiene una mejor generalización y ajusta mejor nuevos datos.

No obstante, tal y como se vio en la figura 4 los modelos cometen un gran error al principio del entrenamiento y posiblemente si se tomasen los datos a partir de 2012 los modelos ARIMA mejorarían significativamente.

Por último, como el objetivo es predecir se reentrenan todos los modelos con todos los datos de la serie, es decir, se juntan los conjuntos de entrenamiento y validación en uno solo porque para una predicción a futuro los más relevante son los datos más recientes. En la figura 8 se recogen las predicciones de estos 4 modelos.

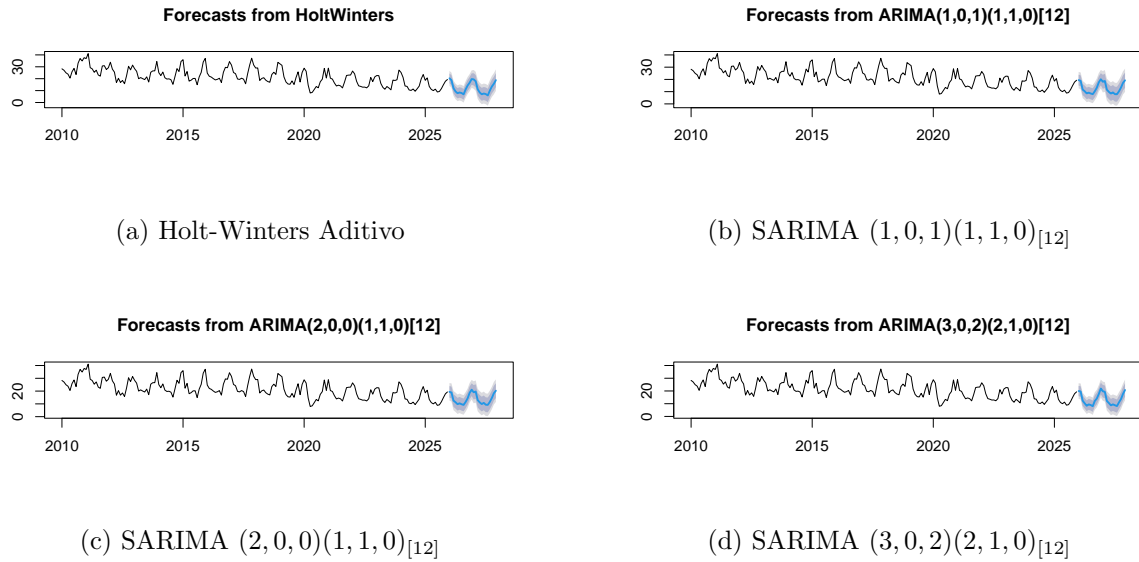


Figura 8: Predicciones de los 4 modelos ajustados:.

5. Conclusiones

El análisis de la serie temporal de las concentraciones de NO_2 en la estación de Cuatro Caminos ha permitido caracterizar la dinámica de la contaminación urbana y evaluar la eficacia de distintas herramientas de predicción. A partir de los resultados obtenidos y la discusión planteada, se extraen las siguientes conclusiones principales:

- **Dinámica de la serie y preprocesamiento:** El análisis exploratorio confirmó un periodo estacional mensual ($s = 12$) y una tendencia decreciente a largo plazo, consistente con la implementación de políticas de reducción de emisiones en la última década. Asimismo, se identificó la naturaleza heterocedástica de los datos, requiriendo una transformación de Box-Cox ($\lambda \approx 0,8$) y una diferenciación estacional ($D = 1$) para garantizar la estacionariedad, confirmada posteriormente mediante el test de Dickey-Fuller.
- **Superioridad de los métodos de alisado:** Los modelos deterministas demostraron una mayor capacidad de generalización que los estocásticos en este escenario específico. El modelo de **Holt-Winters aditivo** resultó ser el más robusto, alcanzando el menor error en el conjunto de validación ($RMSE = 2,88 \mu g/m^3$).
- **Limitaciones de los modelos SARIMA:** Si bien se identificaron modelos estadísticamente significativos, estos presentaron dificultades para cumplir estrictamente con la hipótesis de normalidad en los residuos (Test de Kolmogorov-Smirnov). Esto sugiere que la serie contiene dinámicas no lineales o valores atípicos puntuales que la estructura lineal de los modelos ARIMA no logra capturar en su totalidad.
- **Impacto de la volatilidad histórica:** Se observó un fenómeno particular donde el error de entrenamiento fue sistemáticamente superior al de validación en el caso concreto de Holt-Winters. Esto evidencia que el comportamiento del contaminante antes de 2012 presentaba una variabilidad y unos niveles medios mucho más erráticos que en la actualidad.

Como líneas de trabajo futuro, se propone acotar el rango histórico de entrenamiento (por ejemplo, utilizando datos únicamente a partir de 2012) para reducir el ruido estructural y mejorar el ajuste de los modelos estocásticos. Asimismo, sería pertinente explorar modelos multivariantes (ARIMAX) que incorporen variables exógenas meteorológicas, o técnicas de aprendizaje profundo (Redes LSTM) capaces de modelar las no linealidades detectadas en los residuos.

Referencias

- Avila, M. L., Alonso, A. M., Peña, D., et al. (2023, abril). Modelling Multiple Seasonalities with ARIMA: Forecasting Madrid NO2 Hourly Pollution Levels [Version 1]. <https://doi.org/10.21203/rs.3.rs-2860239/v1>
- Ayuntamiento de Madrid. (2026). Calidad del aire. Datos horarios desde 2001 [Accedido: 30-01-2026]. <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnnextoid=f3c0f7d512273410VgnVCM2000000c205a0aRCRD>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5.^a ed.). John Wiley & Sons.
- Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5), 1287-1294. Consultado el 30 de enero de 2026, desde <http://www.jstor.org/stable/1911963>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), 427-431. Consultado el 30 de enero de 2026, desde <http://www.jstor.org/stable/2286348>
- Fernández-Avilés, G., Mattera, R., & Scepi, G. (2024). Factor-Augmented Autoregressive Neural Network to forecast NOx in the city of Madrid. *Socio-Economic Planning Sciences*, 95, 101958. <https://doi.org/https://doi.org/10.1016/j.seps.2024.101958>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd). OTexts. <https://otexts.com/fpp2/>
- Iskandaryan, D., Ramos, F., & Trilles, S. (2022). Bidirectional convolutional LSTM for the prediction of nitrogen dioxide in the city of Madrid. *PLOS ONE*, 17(6), 1-20. <https://doi.org/10.1371/journal.pone.0269295>
- LJUNG, G. M., & BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303. <https://doi.org/10.1093/biomet/65.2.297>
- Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253), 68-78. Consultado el 30 de enero de 2026, desde <http://www.jstor.org/stable/2280095>
- Orta Blanco, I. S. (2024, septiembre). *Study of Air Quality due to Madrid Central with Gaussian Processes* [Master's Thesis]. Universidad Carlos III de Madrid [Supervisors: Miguel Cárdenas Montes and Juan Miguel Marín Diazaraque]. <https://hdl.handle.net/20.500.14855/4761>
- Peña, D. (2010). Análisis de series temporales. Alianza Editorial.
- Sancho Val, J., Hernando, C. C., & de Baños, L. M. (2026). Functional data analysis of air quality time series in Madrid Using FPCA and splines. *Atmospheric Environment*, 367, 121741. <https://doi.org/https://doi.org/10.1016/j.atmosenv.2025.121741>
- Soares, J., Plass, D., Kienzler, S., González Ortiz, A., Gsell, A., & Horálek, J. (2025, noviembre). Assessing the environmental burden of disease related to air pollution in Europe in 2023. ETC HE Report 2025/8. <https://doi.org/10.5281/zenodo.17658760>
- World Health Organization. (2021). *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide* (inf. téc.). World Health Organization. Geneva. <https://apps.who.int/iris/handle/10665/345329>