

**Predicción de los niveles de NO_2 en el aire de Seúl
mediante el modelo Temporal Convolutional Network
(TCN)**

Carlos López Pérez^a (carlosdl@uninorte.edu.co),
Miguel Herrera Rocha^a (mherrerass@uninorte.edu.co)

^a Msc. en Estadística Aplicada, Departamento de Matemáticas y Estadística,
Universidad del Norte, Barranquilla, Colombia

Predicción de los niveles de NO_2 en el aire de Seúl mediante el modelo Temporal Convolutional Network (TCN)

Carlos López Pérez^a, Miguel Herrera Rocha^a

^aMsc. en Estadística Aplicada, Departamento de Matemáticas y Estadística, Universidad del Norte, Barranquilla, Colombia.

Resumen

El monitoreo de la calidad del aire es fundamental para comprender los efectos de la contaminación atmosférica en la salud pública y el medio ambiente. En este estudio, se aborda la predicción de los niveles de dióxido de nitrógeno (NO_2) en el aire de Seúl, Corea del Sur, utilizando datos de 1988 a 2021. Diversos enfoques, como modelos ARIMA, redes neuronales recurrentes (RNN), LSTM y transformers, han sido empleados para predecir contaminantes atmosféricos, incluyendo NO_2 . Sin embargo, hasta la fecha no se han encontrado estudios han explorado el uso de redes convolucionales temporales (Temporal Convolutional Networks, TCN) para esta tarea. En este trabajo, se implementa un modelo TCN para predecir los niveles de NO_2 , evaluando su capacidad para capturar dependencias temporales y realizar predicciones precisas. Los resultados muestran que el modelo TCN es capaz de identificar patrones complejos y no lineales en los datos históricos, superando a modelos tradicionales en términos de precisión. Este estudio destaca la efectividad de las TCN en la predicción de contaminantes atmosféricos y sugiere su potencial para mejorar la gestión de la calidad del aire a nivel global.

Palabras Clave: NO_2 , Temporal Convolutional Network, calidad del aire, serie temporal, transformers, modelos de predicción.

Email addresses: carlosdl@uninorte.edu.co (Carlos López Pérez),
mherrerass@uninorte.edu.co (Miguel Herrera Rocha)

1. Datos

Los datos utilizados en este estudio provienen del conjunto de datos histórico sobre la calidad del aire en Seúl, Corea del Sur desde 1988 hasta 2021. Este conjunto de datos contiene más de 5 millones de observaciones en diferentes puntos de medición por cada hora relacionadas con varios contaminantes atmosféricos tales como Ozono (O_3), PM2.5, PM10 y SO_2 y los niveles de dióxido de nitrógeno (NO_2). La variable de interés en este análisis es la concentración de NO_2 , que se mide en partículas por millón [ppm]. Este conjunto de datos está estructurado en formato de serie temporal lo que facilita su uso para modelos de predicción basados en este tipo de datos. Los autores proporcionan acceso completo al conjunto de datos, permitiendo su replicación para futuras investigaciones y puede ser consultado en la bibliografía (Hyun, 2021).

El dióxido de nitrógeno NO_2 es un contaminante atmosférico común que puede tener efectos negativos en la salud humana y el medio ambiente ya que la exposición a altos niveles ambientales puede aumentar el riesgo de infecciones del tracto respiratorio en la población debido a la interacción del contaminante con el sistema inmune (Stieb et al., 2021). El análisis de las concentraciones de NO_2 en el aire es crucial para entender los patrones de contaminación y tomar medidas adecuadas para su control.

1.1. Análisis exploratorio de los datos - EDA

El objetivo del presente análisis exploratorio de datos es comprender la estructura de los datos históricos de concentraciones de NO_2 en el aire de Seúl, identificar patrones temporales, detectar datos atípicos y determinar la existencia de tendencia o estacionalidad de la variable objetivo. Este proceso es fundamental para preparar los datos antes de aplicar modelos predictivos, ya que permite entender la distribución, tendencias y estacionalidad de la serie, así como manejar valores faltantes o atípicos.

El conjunto de datos inicial contenía un total de 5,984,782 registros, con mediciones de concentraciones. Una de las variables clave en la base de datos es

'loc', que hace referencia al punto de medición donde se tomaron las muestras de NO_2 . Esta variable indica el punto de cada medición, por tanto, se decidió trabajar con una única ubicación. Para ello, se identificaron los valores únicos de loc y se seleccionó la ubicación con el mayor número de registros, que resultó ser la ubicación 124. Esta locación contaba con 288,736 mediciones, lo que la convierte en la más representativa y adecuada para realizar un análisis detallado de las concentraciones de NO_2 a lo largo del tiempo.

Durante el análisis, se identificó una observación atípica en la concentración de NO_2 , cuyo valor era mucho más alto que el resto de los datos, registrando 0.687 ppm. Este valor (contrastaba notablemente con el percentil 99 que era de 0.08 ppm.) Dada esta discrepancia, se decidió tratar este valor como un outlier y reemplazarlo por un valor NaN para evitar distorsiones en el análisis. Posteriormente, este valor faltante fue imputado junto con otros valores faltantes en la serie. En total, se encontraron 6,898 valores faltantes en la variable de estudio, lo que representó aproximadamente el 2.39% del total de observaciones. Estos valores fueron imputados utilizando el método de interpolación cuadrática, lo que permitió mantener la integridad de la serie temporal sin perder información relevante. A continuación, se presenta un resumen estadístico de las concentraciones de NO_2 después de la imputación:

Métrica	Valor
Cantidad de datos	288736.00
Media	0.0318
Desviación estándar	0.0171
Valor Mínimo	0.0000
Cuartil 1	0.0190
Mediana	0.0290
Cuartil 3	0.0420
Valor Máximo	0.2040
Rango Intercuartílico (IRQ)	0.0230

Table 1: Resumen estadístico de las concentraciones de NO_2 .

La distribución de las concentraciones de NO_2 , visualizada en la gráfica de densidad, muestra un sesgo a la izquierda, con la mayoría de los valores concentrados en el rango inferior (entre 0.00 y 0.05 ppm), pero con una cola que se extiende hacia valores más altos. Esto indica que, aunque la mayoría de las mediciones son bajas, existen muy ocasionales picos de concentración que elevan la media.

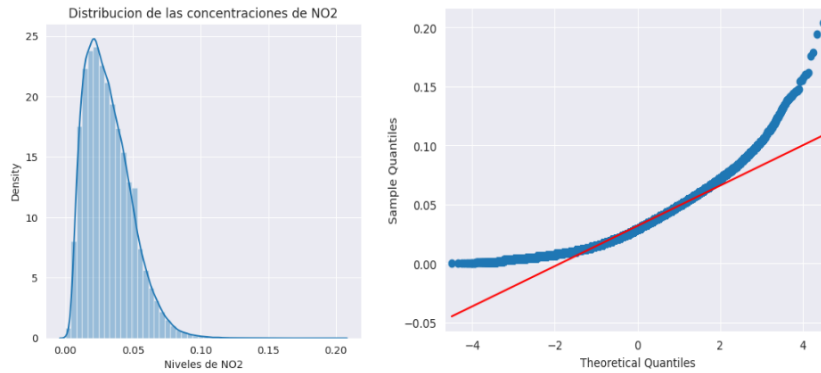


Figure 1: Distribución de la concentración de NO_2 en ppm

Por otro lado, el gráfico de cuantiles revela una desviación significativa de la normalidad, especialmente en los extremos, donde los valores observados no siguen la línea teórica de una distribución normal. Esto se confirma con la prueba de Shapiro-Wilk, que arrojó un p-valor extremadamente bajo ($p \leq 0.0001$) y un estadístico de 0.949, lo que lleva a rechazar la hipótesis nula de normalidad. En conclusión, las concentraciones de NO_2 no siguen una distribución normal.

A continuación se presenta en las gráficas 2 y 3 una visualización de la serie temporal con el objetivo de observar la tendencia general y posibles anomalías. En la siguiente gráfica se observan tres series superpuestas que representan diferentes niveles de agregación temporal de las concentraciones de NO_2 .

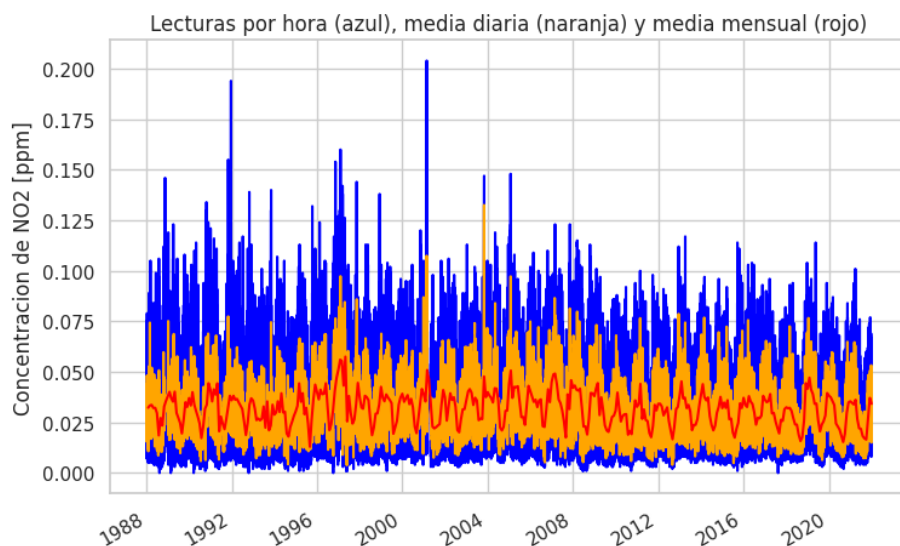


Figure 2: Visualización concentraciones de NO_2 en el tiempo

La serie graficada en azul muestra las mediciones originales de NO_2 tomadas por hora. Se puede apreciar una alta variabilidad, con fluctuaciones significativas a lo largo del tiempo. Para suavizar la variabilidad horaria y resaltar patrones a más largo plazo, se calculó la media diaria (naranja) y mensual (rojo) de las concentraciones de NO_2 , en esta gráfica podemos sospechar que no se encuentra una tendencia marcada pero que con las medias mensuales (rojo) se sospecha de un patrón estacional, esto será verificado mas adelante.

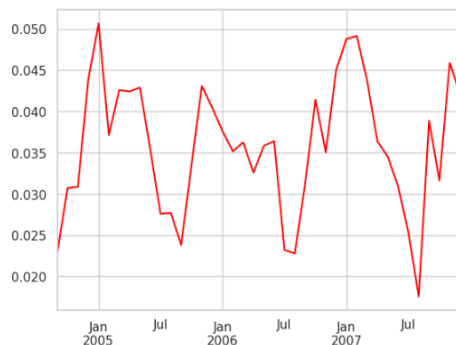


Figure 3: Ampliación de la serie de medias mensuales entre 2005-2008

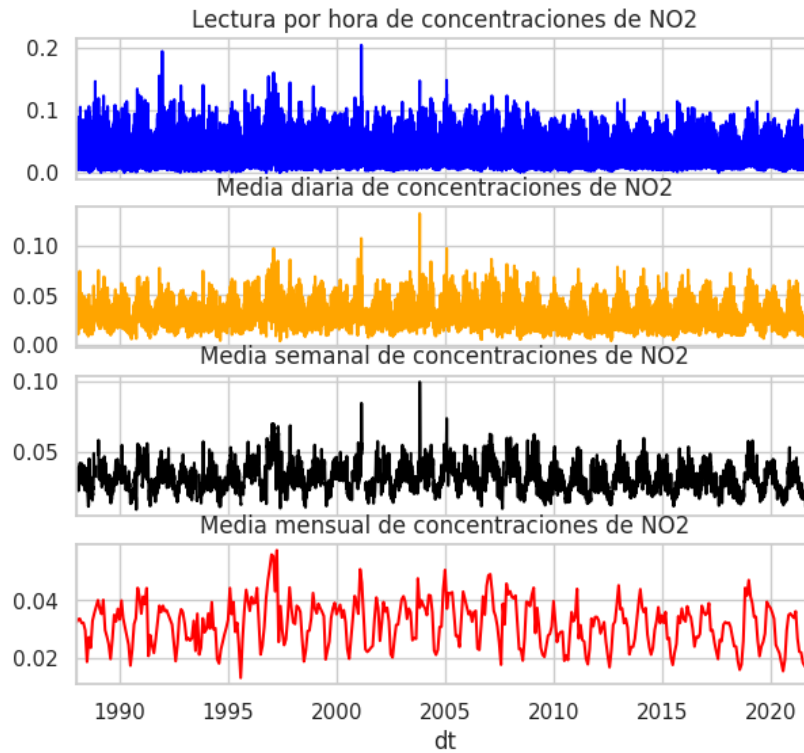


Figure 4: Concentraciones de NO_2 variando su agrupación temporal

Se modela la tendencia de la serie de tiempo con una regresión lineal, en esta, los residuos del modelo de tendencia aplicado a las concentraciones de NO_2 oscilan entre 0.15 y -0.05. Es importante verificar si los residuos muestran algún patrón o autocorrelación que indique la presencia de estructura adicional no capturada por el modelo. Se recomienda realizar pruebas de autocorrelación y ruido blanco en los residuos para validar la adecuación del modelo y, en caso de ser necesario, incorporar términos adicionales como componentes estacionales o de media móvil.

Si se amplía en una sección de la serie con las medias mensuales, es posible ver un patrón que se repite de forma anual en donde las concentraciones de NO_2 parecen llegar a su punto mas bajo en los meses de Julio y agosto (ver gráfica 4).

Residuos del modelo de tendencia para las concentraciones de NO₂

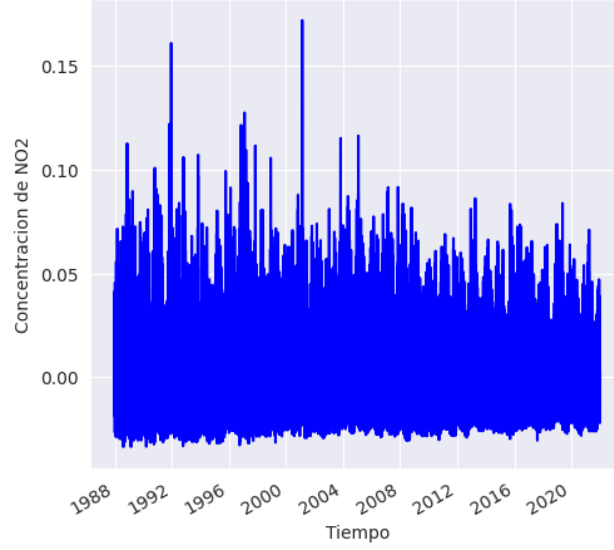


Figure 5: Residuos del modelo de tendencia para las concentraciones de NO₂

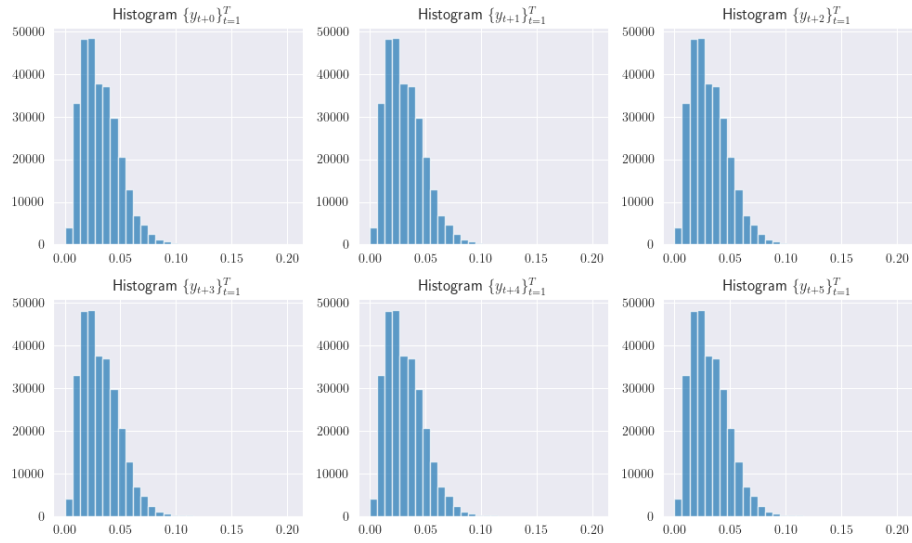


Figure 6: Distribución de NO₂ para Diferentes Horizontes Temporales

El análisis de los histogramas de la serie temporal y_{t+k} para $k = 0, 1, 2, 3, 4, 5$ sugiere que la distribución de la variable se mantiene estable a lo largo del

tiempo. La forma de los histogramas no muestra cambios significativos, lo que indica una dinámica estacionaria o con baja variabilidad en su estructura. No se observa un desplazamiento evidente en la media o la dispersión de los datos, lo que sugiere que la serie no tiene tendencias marcadas ni cambios abruptos. Esto es relevante para el modelado, ya que implica que métodos estadísticos que asumen estacionalidad (La cual se confirmara con la prueba de Dickey Fuller) pueden ser adecuados para su análisis y predicción.

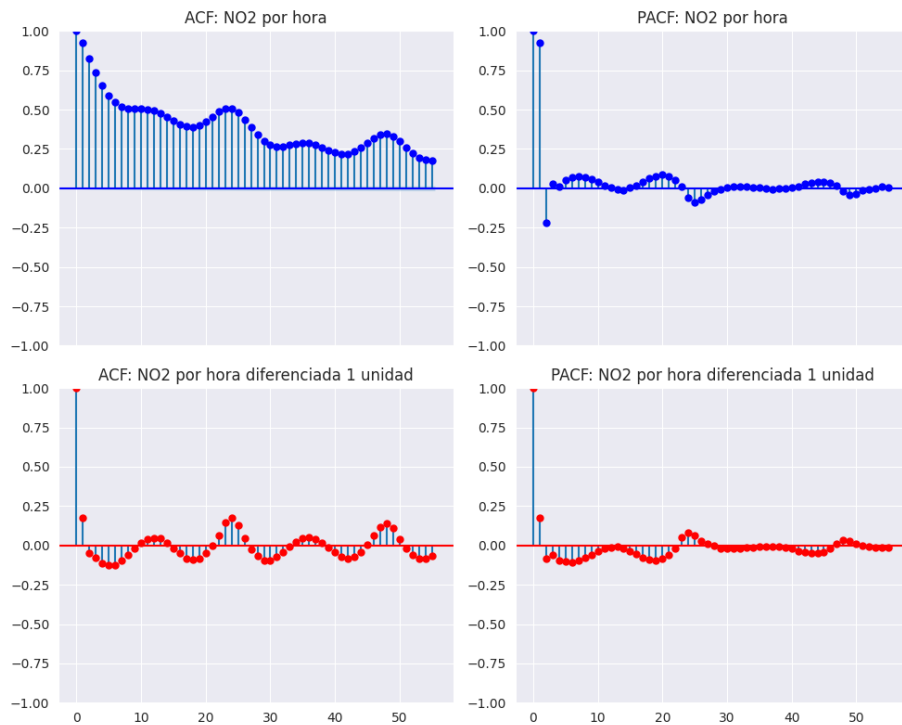


Figure 7: ACF y PACF de la serie de tiempo original y diferenciada 1 unidad

Se desea verificar la estacionalidad de las mediciones por hora de las concentraciones de NO_2 por lo que se realiza una prueba Dickey fuller para confirmar la estacionalidad. En los resultados de la prueba de Dickey-Fuller para la concentración de NO_2 , obtuvimos un valor p menor a 0.0001 sin diferenciar los datos, lo que indica que podemos rechazar la hipótesis nula de que la serie tiene una raíz unitaria (es decir, que la serie no es estacionaria) al nivel de significancia

del 5%. Esto nos confirma que la serie es estacionaria.

El análisis de la función de autocorrelación (ACF) y la autocorrelación parcial (PACF) sugiere que la serie de NO_2 por hora presenta una fuerte dependencia temporal y una estructura estacional clara. La ACF de la serie original muestra un decaimiento exponencial y/o sinusoides amortiguadas, mientras que la PACF presenta un pico significativo en el primer y segundo rezago, lo que según lo visto gráficamente, un modelo auto-regresivo podría ser útil para modelar esta serie de tiempo pero debe ser confirmado con pruebas estadísticas.

La prueba de Dickey-Fuller (DF) confirma que la serie es estacionaria, lo que indica que no presenta tendencia o raíz unitaria. Sin embargo, la función de autocorrelación (ACF) muestra autocorrelación significativa (En la prueba de Ljung-Box) en varios lags, lo que sugiere la presencia de una estructura de dependencia temporal en la serie. Por otro lado, la función de autocorrelación parcial (PACF) no muestra autocorrelación parcial significativa, lo que indica la ausencia de un componente autorregresivo (AR) claro. Estos resultados no son contradictorios, sino que reflejan que la serie, aunque estacionaria, tiene una dependencia temporal que podría ser capturada por un componente de media móvil (MA).

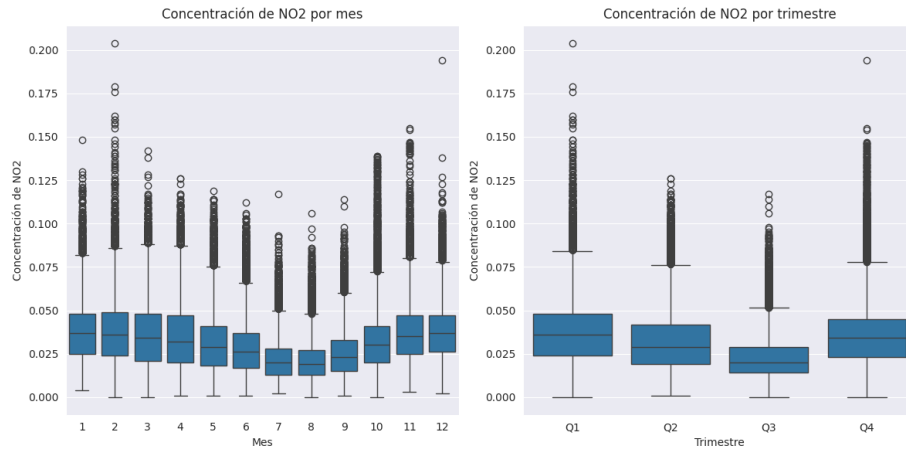


Figure 8: Concentración de NO_2 por trimestre

En adición a lo visto con anterioridad, se presentan las concentraciones de NO_2 por trimestre en donde vemos que dependiendo del trimestre de medición se obtuvieron diferentes medias de concentración de NO_2 , cabe resaltar que al tomar las medias por mes y por trimestre la serie temporal se comporta de forma diferente a si se toman los datos por hora (el cual es la serie original).

2. Métodos

El dióxido de nitrógeno (NO_2) es un contaminante atmosférico producido principalmente por la combustión de combustibles fósiles, incluyendo emisiones de vehículos y procesos industriales. Su presencia en el aire tiene efectos adversos tanto en la salud humana como en el medio ambiente (Stieb et al., 2021). Exposiciones prolongadas a niveles elevados de NO_2 pueden agravar enfermedades respiratorias, reducir la función pulmonar y aumentar la susceptibilidad a infecciones respiratorias (Kashtan et al., 2024). Además, este gas contribuye a la formación de ozono troposférico y partículas finas, amplificando los efectos de la contaminación del aire.

El análisis y predicción de la concentración de NO_2 en entornos urbanos es esencial para implementar políticas efectivas de control de la contaminación y mitigar sus impactos (de Vries et al., 2024). Para ello, se han desarrollado diversos modelos de series temporales que permiten analizar la evolución de las concentraciones de este contaminante y realizar predicciones a corto y largo plazo.

2.1. *Revisión Teórica*

En este estudio, se comparan múltiples enfoques de modelado, desde métodos estadísticos tradicionales hasta técnicas avanzadas de aprendizaje profundo, con el objetivo de identificar la metodología más efectiva para predecir los niveles de NO_2 en Seúl, Corea del Sur por hora. A continuación, se describen brevemente los modelos utilizados en el análisis para comparar con el modelo principal.

2.1.1. *Suavizamiento exponencial*

El Suavizamiento exponencial (SE) es una técnica de series temporales que asigna mayor peso a las observaciones más recientes, reduciendo la influencia de datos más antiguos. En este estudio se utilizaran los modelos de suavización exponencial de primer y segundo orden.

- **SE(1):** Este modelo se utiliza para series de tiempo sin tendencia ni estacionalidad. Este modelo Ajusta los datos dando más peso a las observaciones recientes y esta dado por:

$$\tilde{y}_t = \lambda y_t + (1 - \lambda)\tilde{y}_{t-1}$$

La suavización exponencial de primer orden, puede verse como combinación lineal de la observación actual y la observación suavizada en un tiempo previo. Definiendo a λ como factor de corrección siendo $|\lambda| < 1$

- **SE(2):** Extensión del SE(1) que incluye una componente de tendencia lineal. Dada la suavización exponencial de segundo orden definida por:

$$\tilde{y}_T^{(2)} = \lambda \tilde{y}_T^{(1)} + (1 - \lambda)\tilde{y}_{T-1}^{(2)}$$

Donde $\tilde{y}_T^{(1)}$ y $\tilde{y}_T^{(2)}$ denotan las suavizaciones exponenciales de primer y segundo orden respectivamente y el predictor de \tilde{y}_T esta dado por

$$\tilde{y}_T = 2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}$$

2.1.2. *Suavizamiento exponencial - Holt-Winters*

Los modelos de suavización de Holt-Winters son variantes del Suavizado Exponencial de Holt-Winters, que se diferencian principalmente en cómo manejan la componente estacional de la serie de tiempo. En donde SE-HW(1) es aditivo, SE-HW(2) es multiplicativo, y SE-HW(3) es una variante mixta.

- **SE-HW(1) - Holt-Winters Aditivo**

Este modelo asume que la componente estacional es aditiva, es decir, la magnitud de la estacionalidad es constante a lo largo del tiempo. Es adecuado para series de tiempo donde la estacionalidad no depende del nivel de la serie.

$$\hat{y}_{t+h} = \ell_t + hb_t + s_{t+h-m}$$

Donde,

$$\text{Nivel: } \ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$\text{Tendencia: } b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$$

$$\text{Estacionalidad: } s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$$

- s_t : Componente estacional.
- m : Longitud del ciclo estacional (por ejemplo, 12 para datos mensuales con estacionalidad anual).
- γ : Parámetro de suavizado estacional.

- **SE-HW(2) - Holt-Winters multiplicativo:** Este modelo asume que la componente estacional es multiplicativa, es decir, la magnitud de la estacionalidad varía proporcionalmente con el nivel de la serie. Es adecuado para series de tiempo donde la estacionalidad aumenta o disminuye con el nivel de la serie.

$$\text{Pronóstico: } \hat{y}_{t+h} = (\ell_t + hb_t) \cdot s_{t+h-m}$$

$$\text{Nivel: } \ell_t = \alpha \left(\frac{y_t}{s_{t-m}} \right) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$\text{Tendencia: } b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$$

$$\text{Estacionalidad: } s_t = \gamma \left(\frac{y_t}{\ell_{t-1} + b_{t-1}} \right) + (1 - \gamma)s_{t-m}$$

- **SE-HW(3) - Holt-Winters mixto:** Este modelo combina componentes aditivos y multiplicativos, permitiendo que algunas partes de la serie tengan estacionalidad aditiva y otras multiplicativa. Es útil cuando la serie de tiempo muestra un comportamiento mixto, donde la estacionalidad no es completamente aditiva ni multiplicativa.

2.1.3. **ARIMA(p,d,q):**

El modelo ARIMA combina términos autorregresivos (AR), de integración (I) y de medias móviles (MA). Se denota como ARIMA(p,d,q) y se define como:

$$\phi(B)(1 - B)^d y_t = \delta + \phi(B)\epsilon_t$$

En donde el pronóstico para un periodo τ en el futuro se puede calcular como:

$$y_{T+\tau} = \mu + \sum_{i=0}^{\tau-1} \psi_i \epsilon_{T+\tau-i} + \sum_{i=\tau}^{\infty} \psi_i \epsilon_{T+\tau-i}$$

2.1.4. **GARCH:**

GARCH, por sus siglas en inglés (*Generalized Autoregressive Conditional Heteroskedasticity*) es un modelo utilizado para capturar la volatilidad en series temporales, comúnmente aplicado en datos financieros, pero también útil en problemas ambientales donde la variabilidad de la contaminación presenta heterocedasticidad condicional (Hasanov et al., 2024).

El modelo $GARCH(p, q)$ se define mediante dos ecuaciones: una para la media y otra para la varianza condicional.

Ecuación de la media: $y_t = \mu + \epsilon_t$

$$\text{Ecuación de la varianza: } \sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

Donde:

- y_t : Valor de la serie en el tiempo t .

- μ : Media incondicional de la serie.
- ϵ_t : Término de error en el tiempo t , con $\epsilon_t \sim N(0, \sigma_t^2)$.
- σ_t^2 : Varianza condicional en el tiempo t .
- ω : Término constante (intercepto) en la ecuación de la varianza.
- α_i : Coeficientes asociados a los términos ϵ_{t-i}^2 (efectos de los errores al cuadrado rezagados).
- β_j : Coeficientes asociados a los términos σ_{t-j}^2 (efectos de las varianzas condicionales rezagadas).
- p : Orden de los términos ϵ_{t-i}^2 (componente ARCH).
- q : Orden de los términos σ_{t-j}^2 (componente GARCH).

Para garantizar que la varianza condicional sea positiva y estacionaria, se imponen las siguientes restricciones:

$$\omega > 0, \quad \alpha_i \geq 0, \quad \beta_j \geq 0, \quad \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$$

2.1.5. Modelos de Deep-Learning

En adición a los modelos previamente vistos, también se utilizará para comparar con el modelo escogido, 3 modelos de deep-learning los cuales son MLP, RNN y LSTM, estos se definirán a continuación

- **MLP (Multi-Layer Perceptron)**: Este el nombre de una red neuronal de propagación hacia adelante moderna que consta de neuronas totalmente conectadas con funciones de activación no lineales, organizadas en capas, que se destaca por ser capaz de distinguir datos que no son linealmente separables (Bisong, 2019). La salida de una neurona en una capa de perceptrón esta dada por:

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right)$$

Donde:

- f : Función de activación (Por ejemplo, ReLU, sigmoide).
- w_i : Pesos de la red.
- b : Término de sesgo (bias).
- x_i : Entradas de la red.

- **RNN (Redes Neuronales Recurrentes):** Es un tipo de red neuronal creada para procesar datos secuenciales como las series de tiempo, donde el orden de los elementos es importante. A diferencia de las redes neuronales de propagación hacia adelante, las RNN utilizan conexiones recurrentes, donde la salida de una neurona en un paso de tiempo se devuelve como entrada a la red en el siguiente paso de tiempo (Sherstinsky, 2020).

$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

Donde:

- h_t : Estado oculto en el tiempo t .
 - W_h : Matriz de pesos para el estado oculto anterior.
 - W_x : Matriz de pesos para la entrada actual.
 - b : Término de sesgo.
 - f : Función de activación (Por ejemplo: tanh, ReLU).
- **LSTM (Long-Short Term Memory):** Las LSTM son una variante de las RNN diseñadas para capturar dependencias de largo plazo mediante el uso de celdas de memoria y puertas y tratando de evitar el desvanecimiento del gradiente (Siami-Namini et al., 2019). Estas se han utilizado para el análisis de series de tiempo, por tal motivo se utilizara como modelo para comparar con el original.

$$\text{Formulación: } \left\{ \begin{array}{l} \text{Puerta de olvido: } f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \\ \text{Puerta de entrada: } i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \text{Celda candidata: } \tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \\ \text{Actualización de celda: } C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \\ \text{Puerta de salida: } o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \\ \text{Estado oculto: } h_t = o_t \cdot \tanh(C_t) \end{array} \right.$$

Donde:

- σ : Función sigmoide.
- W_f, W_i, W_C, W_o : Matrices de pesos.
- b_f, b_i, b_C, b_o : Términos de sesgo.
- C_t : Estado de la celda en el tiempo t .
- h_t : Estado oculto en el tiempo t .

2.2. Temporal Convolutional Network

Las redes neuronales convolucionales temporales (TCN) han surgido como una alternativa a las redes neuronales recurrentes (RNN) para el modelado de datos secuenciales, como las series de tiempo. A diferencia de las RNN, que procesan los datos de manera secuencial, las TCN utilizan convoluciones causales y dilatadas, lo que permite el procesamiento en paralelo y una mejor captura de dependencias a largo plazo sin los problemas de desvanecimiento del gradiente (Lea et al., 2016).

Una TCN se compone de capas convolucionales unidimensionales que aplican filtros sobre secuencias temporales. La salida de una capa convolucional está dada por:

$$y_t = \sum_{i=0}^{k-1} w_i \cdot x_{t-d-i} + b$$

donde w_i representa los pesos del filtro convolucional, $x_{t-d \cdot i}$ es la entrada en el tiempo t desplazada por un factor de dilatación d , k es el tamaño del kernel y b es un término de sesgo.

Las TCN presentan dos características fundamentales que las diferencian de las redes recurrentes: *Convoluciones causales* las cuales garantizan que la salida en un instante t solo dependa de valores anteriores en la secuencia, evitando el acceso a información futura y *Convoluciones dilatadas* las cuales permiten expandir el campo receptivo de la red sin aumentar significativamente el número de parámetros, facilitando la captura de patrones de largo alcance en la serie temporal.

La formulación completa de una TCN con múltiples capas está dada por:

$$h_t^{(l)} = \sigma \left(\sum_{i=0}^{k-1} W_i^{(l)} h_{t-d \cdot i}^{(l-1)} + b^{(l)} \right)$$

donde $h_t^{(l)}$ es la activación de la capa l en el tiempo t , $W_i^{(l)}$ son los pesos del filtro convolucional en la capa l , d es el factor de dilatación, y σ es la función de activación (por ejemplo, ReLU).

Además, las TCN incorporan **conexiones residuales**, lo que permite mejorar la estabilidad del entrenamiento y facilita la propagación del gradiente en redes profundas. La estructura final de la red se define como:

$$h_t^{(l)} = \text{BatchNorm} \left(h_t^{(l)} + \text{Conv1D} \left(h_t^{(l-1)} \right) \right)$$

donde BatchNorm representa la normalización por lotes y Conv1D es una capa convolucional unidimensional. Gracias a estas propiedades, las TCN han demostrado un rendimiento competitivo en tareas de predicción de series temporales, superando en muchos casos a los modelos recurrentes tradicionales (Lara-Benítez et al., 2020).

3. Resultados

3.1. $ES(1)$

El modelo de suavizamiento exponencial simple ($ES(1)$) se utilizó para predecir los niveles de dióxido de nitrógeno (NO_2). La Figura 9 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución sesgada a la izquierda, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal. Además, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos.

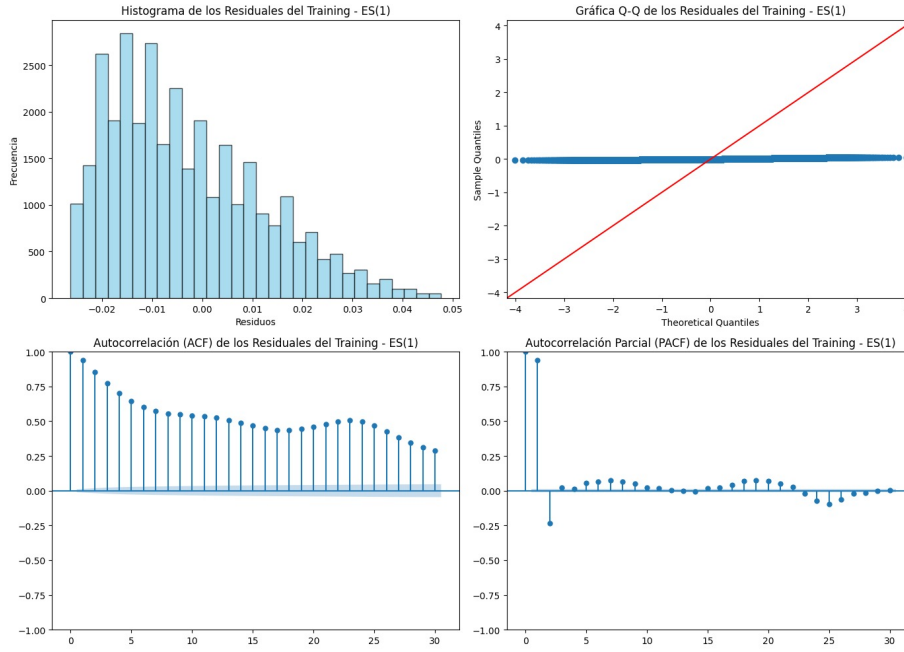


Figure 9: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - $ES(1)$.

La Figura 10 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los residuales en el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica las predicciones tienen un error

porcentual del 70.12% respecto a los valores reales. Esto sugiere problemas en la precisión de la predicción; cabe resaltar que el RMSE de este modelo es de 0.016 lo que indica que error es relativamente alto en comparación con la escala de los datos (La media de las mediciones de NO_2 es de 0.0318) .

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos.

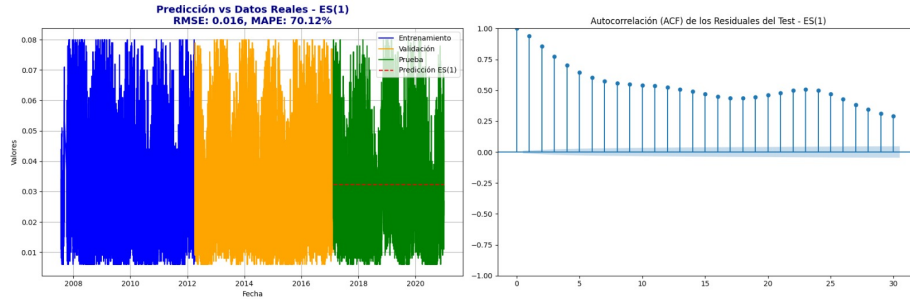


Figure 10: Predicción vs Datos Reales y ACF de los Residuales del Test - ES(1).

Teniendo en cuenta estos resultados, podemos concluir que este modelo es incapaz de manejar bien la variabilidad de la serie por lo que se debe buscar uno que ayude a capturar mejor la estructura de los datos.

3.2. $ES(2)$

Para el modelo de Suavizamiento exponencial simple ($ES(2)$) La Figura 11 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución sesgada a la izquierda, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal. Además, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos.

La Figura 12 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los residuales en

el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica las predicciones tienen un error porcentual del 70.12% respecto a los valores reales. Esto sugiere problemas en la precisión de la predicción; cabe resaltar que el RMSE de este modelo es de 0.016 lo que indica que error es relativamente alto en comparación con la escala de los datos.

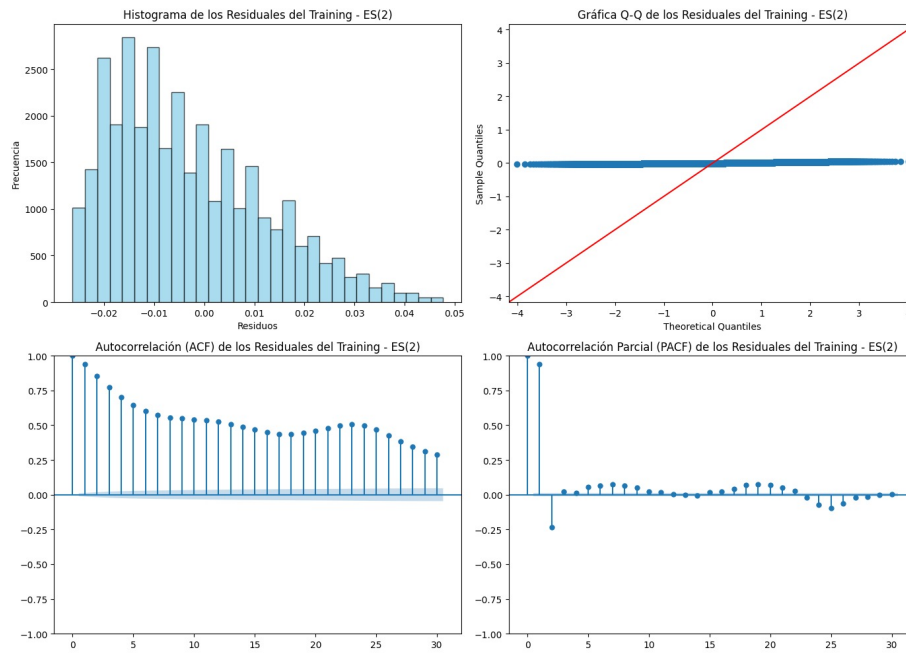


Figure 11: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - ES(2).

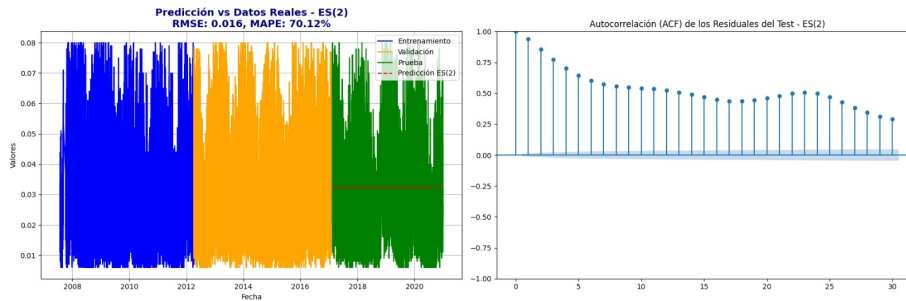


Figure 12: Predicción vs Datos Reales y ACF de los Residuales del Test - ES(2).

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos. Teniendo en cuenta estos resultados, podemos concluir que este modelo arroja resultados iguales que el ES(1) y por ende, también es incapaz de manejar bien la variabilidad de la serie por lo que se debe buscar uno que ayude a capturar mejor la estructura de los datos.

3.3. ES-HW(1)

Para el modelo de Suavizamiento exponencial Holter-Winters ES-HW(1) La Figura 13 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución sesgada a la izquierda, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal. Además, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos.

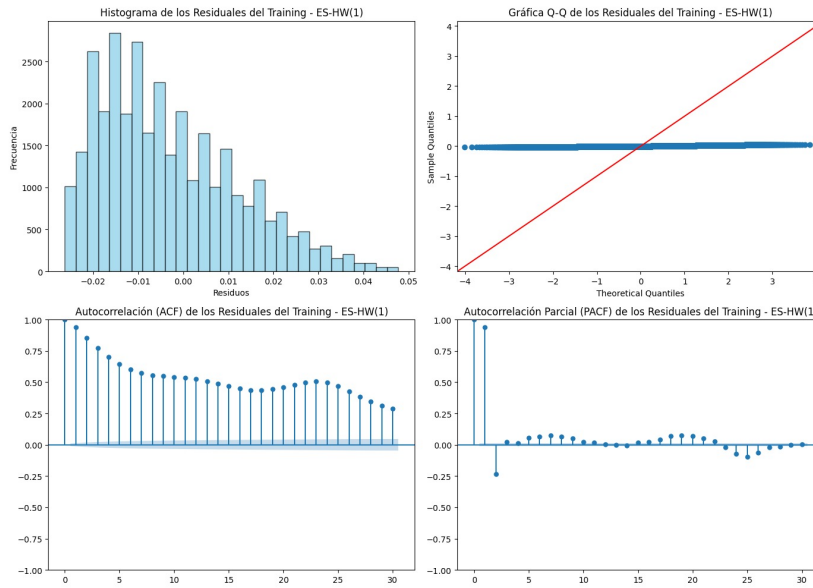


Figure 13: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - ES-HW(1).

La Figura 14 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los residuales en el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica las predicciones tienen un error porcentual del 70.12% respecto a los valores reales. Esto sugiere problemas en la precisión de la predicción; cabe resaltar que el RMSE de este modelo es de 0.016 lo que indica que error es relativamente alto en comparación con la escala de los datos.

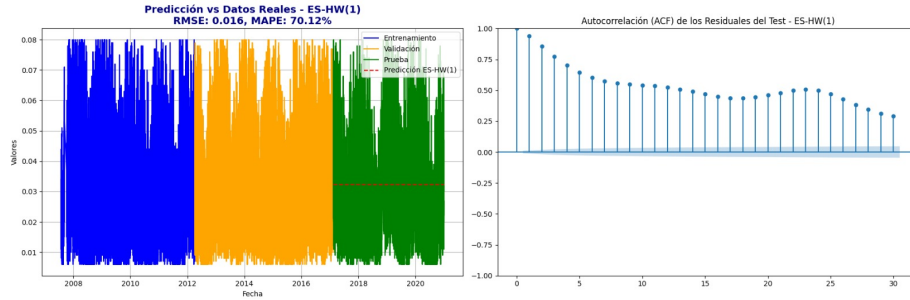


Figure 14: Predicción vs Datos Reales y ACF de los Residuales del Test - ES-HW(1).

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos. Teniendo en cuenta estos resultados, podemos concluir que este modelo arroja resultados iguales que los ES(1), ES(2) presentados anteriormente y por ende, también es incapaz de manejar bien la variabilidad de la serie por lo que se debe buscar uno que ayude a capturar mejor la estructura de los datos.

3.4. ES-HW(2)

Para el modelo de Suavizamiento exponencial Holter-Winters ES-HW(2) La Figura 15 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución sesgada a la izquierda, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal. Además, la función de autocorrelación (ACF) y la función de autocorrelación

parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos.

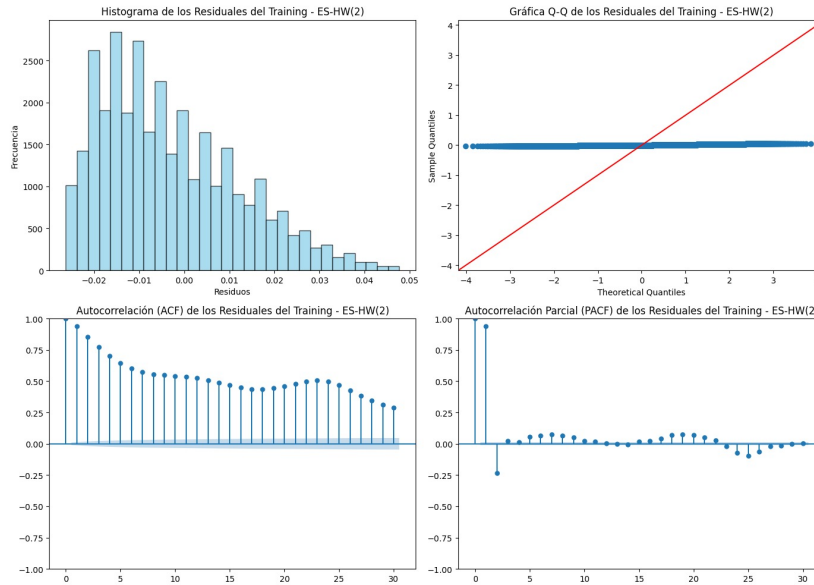


Figure 15: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - ES-HW(2).

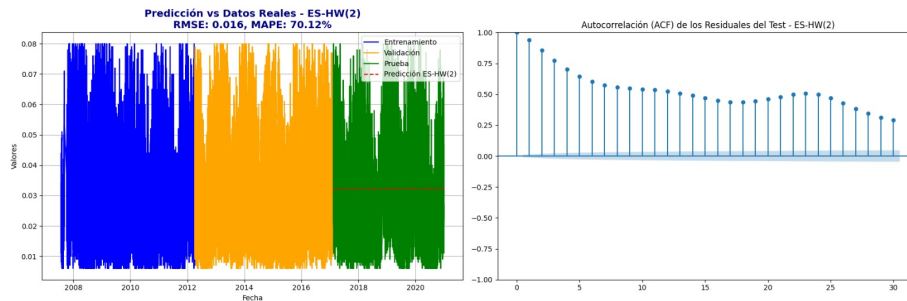


Figure 16: Predicción vs Datos Reales y ACF de los Residuales del Test - ES-HW(2).

La Figura 16 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los residuos en el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica las predicciones tienen un error porcentual del 70.12% respecto a los valores reales. Esto sugiere problemas en

la precisión de la predicción; cabe resaltar que el RMSE de este modelo es de 0.016 lo que indica que error es relativamente alto en comparación con la escala de los datos.

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos. Teniendo en cuenta estos resultados, podemos concluir que este modelo arroja resultados iguales que los ES(1), ES(2) y ES-HW(1) presentados anteriormente y por ende, también es incapaz de manejar bien la variabilidad de la serie por lo que se debe buscar uno que ayude a capturar mejor la estructura de los datos.

3.5. ES-HW(3)

Para el modelo de Suavizamiento exponencial Holter-Winters ES-HW(3) La Figura 17 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución sesgada a la izquierda, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal.

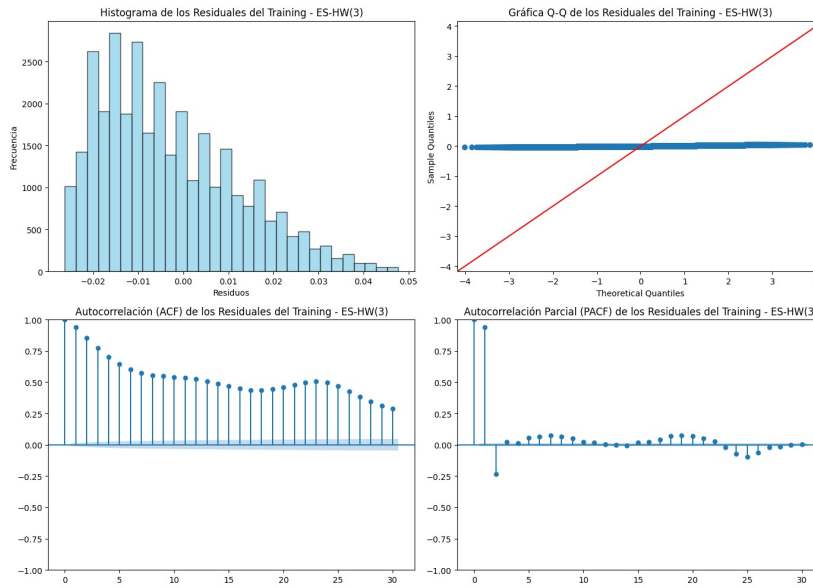


Figure 17: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - ES-HW(3).

En adición a esto, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos. La Figura 18 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los residuales en el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica las predicciones tienen un error porcentual del 70.12% respecto a los valores reales. Esto sugiere problemas en la precisión de la predicción; cabe resaltar que el RMSE de este modelo es de 0.016 lo que indica que error es relativamente alto en comparación con la escala de los datos.

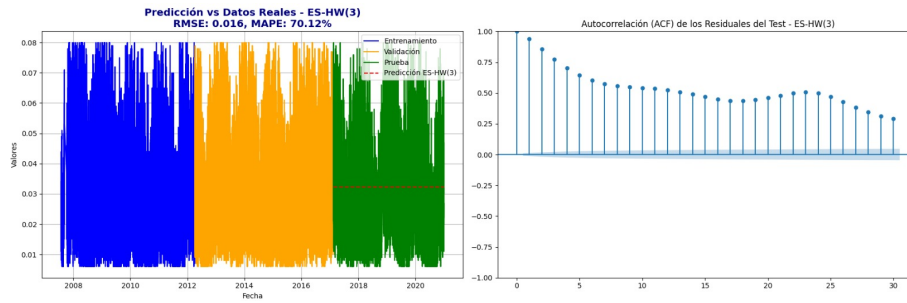


Figure 18: Predicción vs Datos Reales y ACF de los Residuales del Test - ES-HW(3).

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos. Teniendo en cuenta estos resultados, podemos concluir que este modelo arroja resultados iguales que los ES(1), ES(2), ES-HW(1) y ES-HW(2) presentados anteriormente y por ende, también es incapaz de manejar bien la variabilidad de la serie por lo que se debe buscar uno que ayude a capturar mejor la estructura de los datos.

3.6. ARIMA

Para el modelo ARIMA, la Figura 19 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución sesgada a la izquierda, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal.

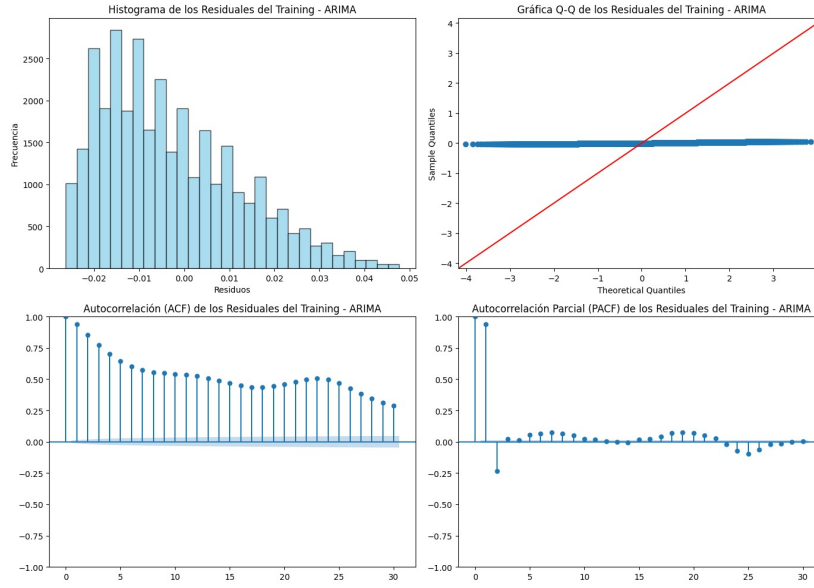


Figure 19: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - ARIMA.

En adición a esto, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos. La Figura 20 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los residuales en el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica que las predicciones tienen un error porcentual del 70.12% respecto a los valores reales. Esto sugiere problemas en la precisión de la predicción; cabe resaltar que el RMSE de este modelo es de 0.016 lo que indica que el error es relativamente alto en comparación con la escala de los datos.

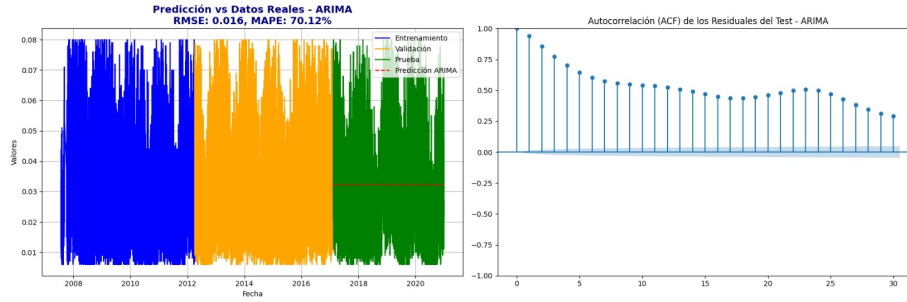


Figure 20: Predicción vs Datos Reales y ACF de los Residuales del Test - ARIMA.

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos. Teniendo en cuenta estos resultados, podemos concluir que este modelo arroja resultados iguales que los ES(1), ES(2), ES-HW(1), ES-HW(2) Y ES-HW(3) presentados anteriormente y por ende, también es incapaz de manejar bien la variabilidad de la serie por lo que se debe buscar uno que ayude a capturar mejor la estructura de los datos.

3.7. GARCH

Para el modelo GARCH, la Figura 21 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución sesgada a la izquierda, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal.

En adición a esto, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos. La Figura 22 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los residuales en el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica las predicciones tienen un error porcentual del 69.71% respecto a los valores reales. Esto sugiere prob-

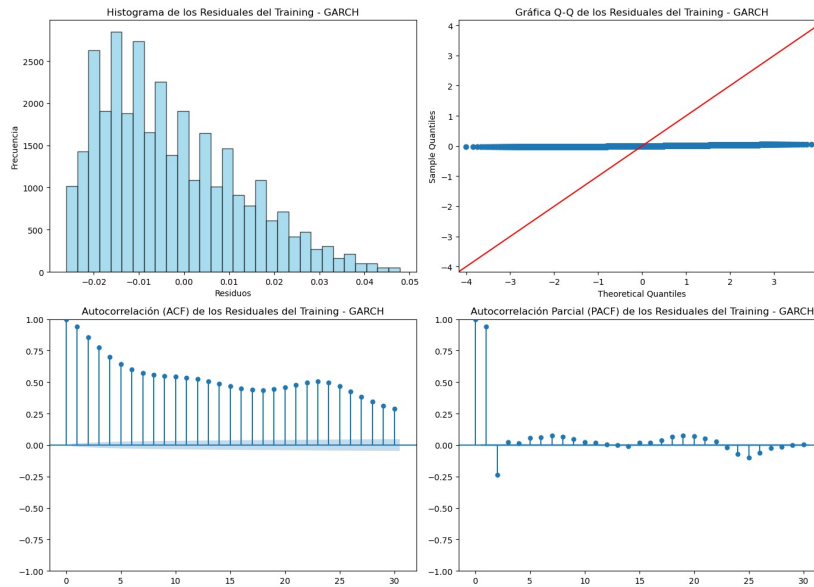


Figure 21: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - GARCH.

lema en la precisión de la predicción aunque este resultado ha sido mejor en comparación con los presentados con anterioridad; cabe resaltar que el RMSE de este modelo es de 0.016 lo que indica que error sigue siendo relativamente alto en comparación con la escala de los datos.

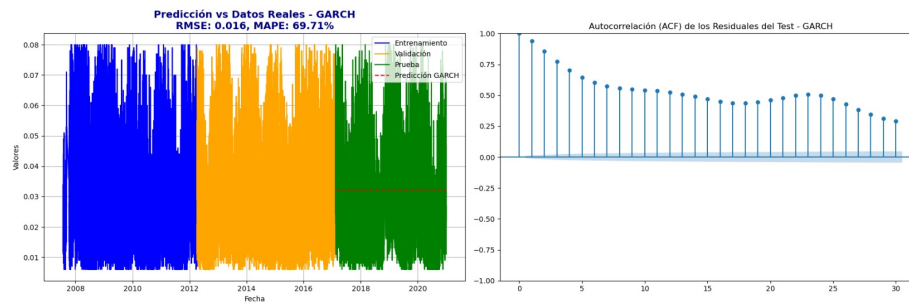


Figure 22: Predicción vs Datos Reales y ACF de los Residuales del Test - GARCH.

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos.

Teniendo en cuenta estos resultados, podemos concluir que este modelo arroja resultados parecidos al resto de modelos, sin embargo presentó un MAPE menor a dichos modelos. DE igual forma al estar los residuos autocorrelacionados, esto significa que este modelo puede no ser adecuado para las predicciones del NO_2 por hora.

3.8. MLP

Para el modelo de multilayer perceptron MLP, la Figura 23 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución simétrica y centrada en 0, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal.

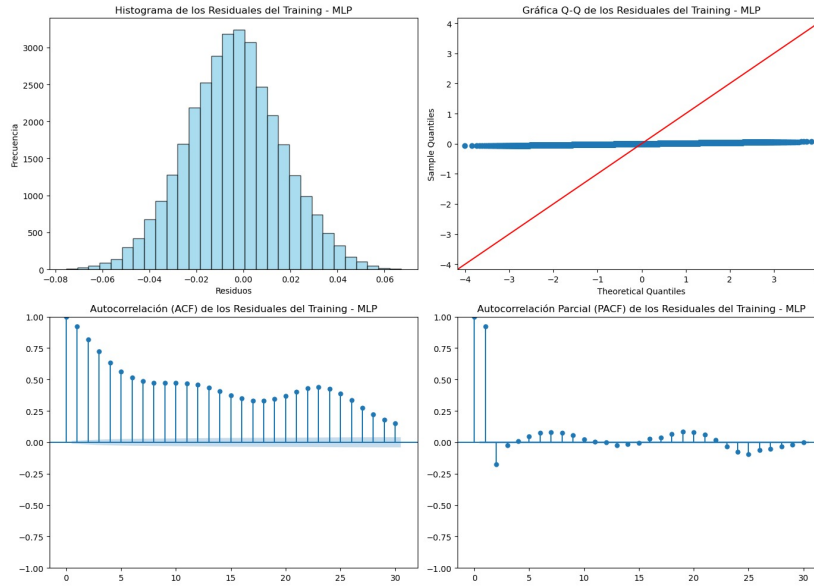


Figure 23: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - MLP.

En adición a esto, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos. La Figura 24 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los

residuales en el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica las predicciones tienen un error porcentual del 81.47% respecto a los valores reales. Esto sugiere problemas en la precisión de la predicción; cabe resaltar que el RMSE de este modelo es de 0.021 lo que indica que error es relativamente alto en comparación con la escala de los datos.

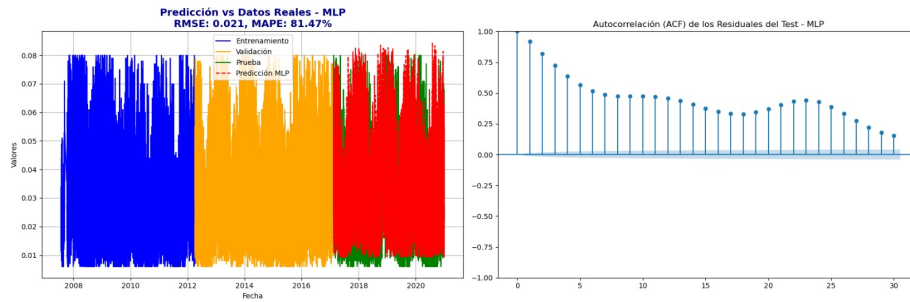


Figure 24: Predicción vs Datos Reales y ACF de los Residuales del Test - MLP.

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos. Teniendo en cuenta estos resultados, podemos concluir que este modelo obtuvo peor rendimiento que los presentados anteriormente y por ende, también es incapaz de manejar bien la variabilidad de la serie por lo que se debe buscar uno que ayude a capturar mejor la estructura de los datos.

3.9. RNN

Para el modelo RNN, la Figura 25 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución simétrica y centrada en 0, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal.

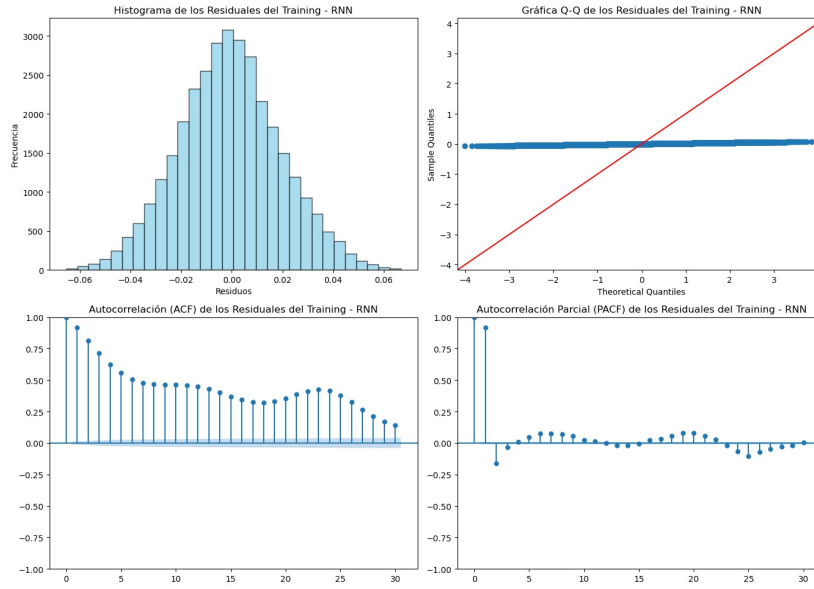


Figure 25: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - RNN.

En adición a esto, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos. La Figura 26 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los residuos en el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica las predicciones tienen un error porcentual alto del 74.92% respecto a los valores reales. Esto sugiere problemas en la precisión de la predicción; cabe resaltar que el RMSE de este modelo es de 0.020 lo que indica que error es relativamente alto en comparación con la escala de los datos.

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos. Teniendo en cuenta estos resultados, podemos concluir que este modelo no tuvo un mejor rendimiento que los presentados anteriormente y es incapaz

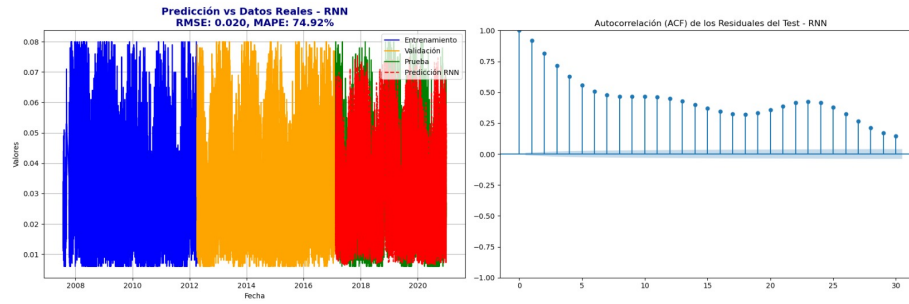


Figure 26: Predicción vs Datos Reales y ACF de los Residuales del Test - RNN.

de manejar bien la variabilidad de la serie por lo que se debe buscar uno que ayude a capturar mejor la estructura de los datos.

3.10. LSTM

Para el modelo LSTM, la Figura 27 presenta el análisis de los residuales en el conjunto de entrenamiento. El histograma sugiere una distribución simétrica y centrada en 0, mientras que el gráfico Q-Q muestra que los residuos no siguen una distribución normal.

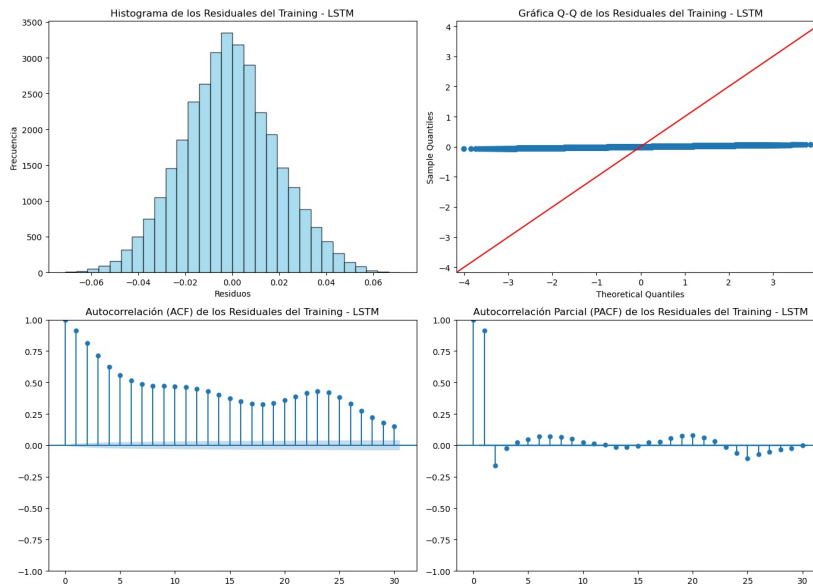


Figure 27: Histograma, QQ-Plot, ACF y PACF de los Residuales del Training - LSTM.

En adición a esto, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) indican que los residuos presentan dependencia temporal, lo que sugiere que el modelo no captura completamente la estructura de los datos.

La Figura 28 muestra la comparación entre los datos reales y las predicciones del modelo, junto con la función de autocorrelación (ACF) de los residuales en el conjunto de prueba. Se observa que, aunque el modelo captura la tendencia general de la serie, el MAPE obtenido indica las predicciones tienen un error porcentual alto del 75.65% respecto a los valores reales. Esto sugiere problemas en la precisión de la predicción; cabe resaltar que el RMSE de este modelo es de 0.020 lo que indica que error es relativamente alto en comparación con la escala de los datos.

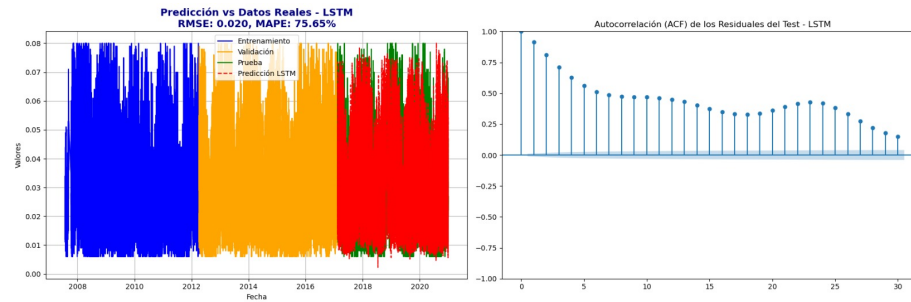


Figure 28: Predicción vs Datos Reales y ACF de los Residuales del Test - LSTM.

También se identifica que los residuos en el conjunto de test del modelo están autocorrelacionados significativamente, esto significa que el modelo no ha capturado correctamente toda la estructura de dependencia temporal en los datos. Teniendo en cuenta estos resultados, podemos concluir que este modelo no tuvo un mejor rendimiento que los presentados anteriormente y es incapaz de manejar bien la variabilidad de la serie por lo que se debe buscar uno que ayude a capturar mejor la estructura de los datos.

3.11. Comparación de los modelos

En esta seccion se presenta la comparacion de los rendimientos de los modelos de benchmark para la prediccion de los niveles por hora del dióxido de nitrógeno (NO_2) en la ciudad de Seul, segun los resultados presentados en la tabla, vemos que los modelos no tienen buenos rendimientos a la hora de realizar las predicciones y que varios de estos presentaron resultados muy similares.

Modelo	MAPE	MAE	RMSE	MSE	LB (Valor p)
$ES^{(1)}$	70,12%	0.014445	0.016	0.000256	< 0.0001
$ES^{(2)}$	70,12%	0.014445	0.016	0.000256	< 0.0001
$ES - HW^{(1)}$	70,12%	0.014445	0.016	0.000256	< 0.0001
$ES - HW^{(2)}$	70,12%	0.014445	0.016	0.000256	< 0.0001
$ES - HW^{(3)}$	70,12%	0.014445	0.016	0.000256	< 0.0001
ARIMA	70,12%	0.014445	0.016	0.000256	< 0.0001
GARCH	69,71%	0.013139	0.015527	0.000256	< 0.0001
MLP	73,65%	0.015872	0.020061	0.000441	< 0.0001
RNN	74,00%	0.015916	0.020091	0.000400	< 0.0001
LSTM	75,65%	0.015999	0.020178	0.000400	< 0.0001

Table 2: Comparación de los resultados de los modelos de benchmark

El mejor de estos fue el modelo GARCH el cual obtuvo un RMSE de 0.0155, un MAPE de 69.71% y un MAE de 0.0013 mostrando una ligera ventaja sobre los otros métodos tradicionales. Este resultado sugiere que la modelización de la volatilidad en la serie temporal podría haber permitido una mejor captura de las variaciones en los datos. Sin embargo varios modelos tradicionales como $ES(1)$, $ES(2)$, $ES-HW(1)$, $ES-HW(2)$, $ES-HW(3)$ y ARIMA presentan un rendimiento muy similar, con un MAPE del 70.12% y métricas de error relativamente bajas y homogéneas.

Por otro lado, los modelos basados en Deep Learning (MLP, RNN y LSTM) obtuvieron los peores resultados entre los modelos evaluados. En particular, el modelo LSTM, que suele ser adecuado para datos secuenciales, obtuvo un

MAPE del 75.65%, siendo el de peor desempeño en la tabla. Esto sugiere que los modelos de redes neuronales no lograron capturar patrones significativos en los datos.

4. Conclusiones

Aquí van las conclusiones finales

References

- Bisong, E. (2019). The multilayer perceptron (mlp). In *Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners* (pp. 401–405). Springer.
- Hasanov, A. S., Brooks, R., Abrorov, S., & Burkhanov, A. U. (2024). Structural breaks and garch models of exchange rate volatility: Re-examination and extension. *Journal of Applied Econometrics*, *39*, 1403–1407.
- Hyun, W. (2021). Seoul air quality historical data. URL: <https://www.kaggle.com/datasets/williamhyun/seoulairqualityhistoricdata/data>.
- Kashtan, Y., Nicholson, M., Finnegan, C. J., Ouyang, Z., Garg, A., Lebel, E. D., Rowland, S. T., Michanowicz, D. R., Herrera, J., Nadeau, K. C. et al. (2024). Nitrogen dioxide exposure, health outcomes, and associated demographic disparities due to gas and propane combustion by us stoves. *Science Advances*, *10*, eadm8680.
- Lara-Benítez, P., Carranza-García, M., Luna-Romera, J. M., & Riquelme, J. C. (2020). Temporal convolutional networks applied to energy-related time series forecasting. *Applied Sciences*, *10*. URL: <https://www.mdpi.com/2076-3417/10/7/2322>. doi:10.3390/app10072322.
- Lea, C., Vidal, R., Reiter, A., & Hager, G. D. (2016). Temporal convolutional networks: A unified approach to action segmentation. In G. Hua, &

- H. Jégou (Eds.), *Computer Vision – ECCV 2016 Workshops* (pp. 47–54). Cham: Springer International Publishing.
- Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.
- Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The performance of lstm and bilstm in forecasting time series. In *2019 IEEE International conference on big data (Big Data)* (pp. 3285–3292). IEEE.
- Stieb, D. M., Berjawi, R., Emode, M., Zheng, C., Salama, D., Hocking, R., Lyrette, N., Matz, C., Lavigne, E., & Shin, H. H. (2021). Systematic review and meta-analysis of cohort studies of long term outdoor nitrogen dioxide exposure and mortality. *PloS one*, *16*, e0246451.
- de Vries, W., Posch, M., Simpson, D., de Leeuw, F. A., van Grinsven, H. J., Schulte-Uebbing, L. F., Sutton, M. A., & Ros, G. H. (2024). Trends and geographic variation in adverse impacts of nitrogen use in europe on human health, climate, and ecosystems: A review. *Earth-Science Reviews*, (p. 104789).