



# CALIDAD DEL AIRE CIUDAD DE MADRID

---

TRABAJO FIN DE MÁSTER

## ¿SABÍAS QUÉ...?

La contaminación atmosférica constituye una de las amenazas silenciosas más persistentes para la salud humana. Cada respiración expone a nuestros pulmones a partículas y gases nocivos que, incluso en bajas concentraciones, pueden alterar el sistema respiratorio y cardiovascular. Ni la edad ni la buena condición física ofrecen plena protección: todos somos vulnerables al aire que respiramos.

**Cristóbal León-Salas Bujalance**

Big Data y Data Science aplicados a la Economía y a la Administración y Dirección de Empresas - UNED

NOVIEMBRE 2025

## 0.1. LEYENDAS Y ABREVIATURAS

Símbolo/ Acrónimo	Significado	Observaciones / Unidad
<b>PM<sub>10</sub></b>	Partículas <sup>1</sup> en suspensión con diámetro ≤ 10 µm	µg/m <sup>3</sup>
<b>PM<sub>2.5</sub></b>	Partículas en suspensión con diámetro ≤ 2,5 µm	µg/m <sup>3</sup>
<b>NO<sub>2</sub></b>	Dióxido de nitrógeno	µg/m <sup>3</sup>
<b>NO</b>	Monóxido de nitrógeno	µg/m <sup>3</sup>
<b>NOx</b>	Óxidos de nitrógeno (NO + NO <sub>2</sub> )	µg/m <sup>3</sup>
<b>O<sub>3</sub></b>	Ozono troposférico	µg/m <sup>3</sup>
<b>SO<sub>2</sub></b>	Dióxido de azufre	µg/m <sup>3</sup>
<b>CO</b>	Monóxido de carbono	En el <i>RAW DATA</i> se presenta en mg/m <sup>3</sup> . Para el análisis se pasará a µg/m <sup>3</sup>
<b>ML</b>	<i>Machine Learning</i>	Aprendizaje automático
<b>DL</b>	<i>Deep Learning</i>	Aprendizaje profundo
<b>ARIMA/SARIMA</b>	Modelos autorregresivos integrados	Modelos clásicos de series temporales
<b>XGBoost</b>	Extreme Gradient Boosting	Modelo basado en boosting por gradiente. Maneja relaciones no lineales y efectos de interacción
<b>CatBoost</b>	Algoritmo Categorical Boosting	Especializado en el tratamiento eficiente de variables categóricas sin necesidad de transformaciones extensivas
<b>LightGBM</b>	Light Gradient Boosting Machine	Optimizado para grandes volúmenes de datos y alta velocidad de entrenamiento
<b>LSTM</b>	Long Short-Term Memory Network	Arquitectura recurrente diseñada para capturar dependencias temporales de largo alcance.
<b>CNN</b>	Convolutional Neural Network	Red convolucional capaz de detectar patrones locales y variaciones estructurales
<b>Modelo Hurdle</b>	Modelo estadístico de dos partes	Combina un modelo binario (ocurrencia) y otro continuo (intensidad) para series con exceso de ceros
<b>Modelo Logit</b>	Modelo de regresión logística binaria	Estima la probabilidad de ocurrencia de un evento (p. ej., incendio)

<sup>1</sup> No es una sustancia única, sino una mezcla compleja de partículas sólidas y líquidas suspendidas en el aire

Símbolo/ Acrónimo	Significado	Observaciones / Unidad
<b>Modelo Gamma</b>	Modelo lineal generalizado con distribución Gamma	Predice variables continuas positivas y sesgadas, como la superficie incendiada.
<b>Prophet</b>	Modelo aditivo generalizado de series temporales	Captura tendencias y estacionalidades mediante funciones de Fourier y splines
<b>Series de Fourier</b>	Representación determinista de componentes periódicos	Descompone una serie temporal en funciones seno y coseno para modelar estacionalidades semanales y anuales
<b>RMSE</b>	<i>Root Mean Squared Error</i>	Error cuadrático medio
<b>MAE</b>	<i>Mean Absolute Error</i>	Error absoluto medio
<b>W/m<sup>2</sup></b>	Vatios por metro cuadrado	Medida de radiación solar instantánea
<b>Wh/m<sup>2</sup></b>	Vatios-hora por metro cuadrado	Energía solar acumulada
<b>OMS</b>	Organización Mundial de la Salud	Recomienda valores límite de las sustancias nocivas que han de cumplirse antes de 2030
<b>ODS</b>	Objetivos de Desarrollo Sostenible	Agenda global de la ONU para el desarrollo sostenible (Agenda 2030)
<b>EEA</b>	Agencia Europea de Medio Ambiente	Encargada de ofrecer información ambiental independiente y de apoyar la elaboración de políticas sostenibles en Europa.
<b>CCSSO</b>	Centro Canadiense de Seguridad y Salud Ocupacional	Promueve un ambiente de trabajo seguro y saludable suministrando información y asesoría sobre seguridad y salud ocupacional.
<b>COV</b>	Compuestos Orgánicos Volátiles	Son gases formados por moléculas orgánicas (basadas en carbono) que se evaporan fácilmente a temperatura ambiente

## 0.2. ÍNDICE

### ÍNDICE CONTENIDO

1. INTRODUCCIÓN .....	1
1.1. CONTEXTO Y MOTIVACIÓN.....	1
1.2. MARCO INSTITUCIONAL .....	1
2. OBJETO Y ALCANCE.....	2
2.1. OBJETIVO GENERAL.....	2
2.2. OBJETIVOS ESPECÍFICOS.....	2
2.3. ALCANCE Y LIMITACIONES .....	3
3. MARCO NORMATIVO Y REGULADOR.....	3
3.1. NORMATIVA INTERNACIONAL Y EUROPEA.....	3
i. Normativa europea .....	3
ii. Guías de Calidad del Aire de la OMS .....	4
3.2. NORMATIVA NACIONAL .....	5
3.3. NORMATIVA AUTONÓMICA Y MUNICIPAL.....	5
3.4. IMPLICACIONES DEL ESTUDIO .....	6
4. LENGUAJE DE PROGRAMACIÓN Y LIBRERIAS .....	6
5. BASE DE DATOS Y PREPROCESADO .....	7
5.1. FUENTE DE DATOS .....	7
i. Datos de calidad del aire .....	8
ii. Datos meteorológicos.....	9
iii. Datos de tráfico .....	10
iv. Datos de incendios .....	10
5.2. PREPROCESAMIENTO - LIMPIEZA Y TRANSFORMACIÓN.....	11
i. Bloque 1 – Integración y depuración de datos de calidad del aire. ....	11
ii. Bloque 2 – Tratamientos sobre el DataFrame original .....	12
iii. Bloque 3 – Integración y depuración de datos de tráfico.....	15
iv. Bloque 4 – Integración y tratamiento datos incendios forestales.....	16
v. Bloque 5 – Imputación de valores faltantes en calidad del aire.....	17
6. ANÁLISIS EXPLORATORIO DE DATOS (EDA) .....	19
6.1. EDA VARIABLES EXÓGENAS.....	19
i. Variables meteorológicas e incendios forestales .....	19
ii. Variable tráfico - TOTAL_HOR_EXT .....	22
6.2. EDA VARIABLE DEPENDIENTE “MEDICION_UGM3”.....	24
i. Resultados estadísticos de “Las Combinaciones” .....	26
ii. Distribuciones y patrones estadísticos y temporales.....	28
iii. Tendencias .....	33
iv. Correlaciones y variables importantes: .....	40

6.3. CONCLUSIONES GLOBALES .....	45
7. MODELOS DE PREDICCIÓN .....	46
7.1. BLOQUE 1 – PREPARACIÓN INCIAL .....	47
7.2. BLOQUE 2/FASE 1 – MODELOS CLÁSICOS (UNIVARIANTES).....	47
i.    MODELO ARIMA.....	47
ii.   MODELO SARIMA .....	51
7.3. BLOQUE 2/FASE 2 – MODELOS MULTIVARIANTES .....	58
i.    MODELOS MACHINE LEARNIG (ML) .....	58
ii.   MODELOS DEEP LEARNIG (DL) .....	62
7.4. BLOQUE 2/FASE 3 Evaluación por RMSE y Selección del mejor modelo ...	64
7.5. BLOQUE 3/FASE 1 - MODELIZACIÓN REFINADA.....	65
7.6. BLOQUE 3/FASE 2 - NUEVA EVALUACIÓN RMSE .....	67
7.7. BLOQUE 3/FASE 3 – PREDICCIONES FINALES .....	69
8. ESTUDIOS ADICIONALES BASADOS EN LAS CONCLUSIONES .....	72
8.1. PM <sub>10</sub> - SUR.....	72
8.2. NO <sub>2</sub> y O <sub>3</sub> - SUR.....	73
8.3. PM <sub>2.5</sub> – CENTRO.....	75
9. BIBLIOGRAFÍA .....	76

## ÍNDICE TABLAS Y FIGURAS

Tabla 1 - Directiva 2008/50/CE - Valores límites .....	4
Tabla 2 - Guia OMS - Valores recomendados y efectos .....	5
Tabla 3 - Librerías usadas .....	7
Tabla 4 - Magnitudes base de datos calidad del aire .....	9
Tabla 5 - Variables base de datos meteorológicos .....	9
Tabla 6 - Variables base de datos incendios .....	10
Tabla 7 - Bloque 1 – Integración y depuración de datos de calidad del aire .....	12
Tabla 8 - Bloque 2 – Tratamientos sobre el DataFrame original .....	13
Tabla 9 - Coordenadas estaciones calidad del aire .....	14
Tabla 10 - Bloque 3 – Integración y depuración de datos de tráfico .....	15
Tabla 11 - Bloque 4 – Integración y tratamiento datos incendios forestales .....	17
Tabla 12 - Bloque 5 – Imputación de valores faltantes en calidad del aire.....	19
Tabla 13 - Tabla resumen EDA variables meteorológicas e incendios .....	20
Tabla 14 - Magnitudes tras preprocesamiento.....	24
Tabla 15 - Criterios - Valores límite contaminantes .....	26
Tabla 16 - Resultados estadísticos combinaciones MEDICION_ugm3 .....	27
Tabla 17 - Comentarios resultados estacionalidad .....	32
Tabla 18 - Comentarios resultados tendencias.....	40
Tabla 19 - Resultados correlaciones y variables importantes .....	44
Tabla 20 - Series mensuales - Tendencia & Estacionalidad .....	48
Tabla 21 - Parámetro d óptimo .....	48
Tabla 22 - ARIMA candidatos .....	50
Tabla 23 - ARIMAs Óptimos .....	51
Tabla 24 - Parámetros SARIMA .....	51
Tabla 25 - Cumplimiento recomendación OMS año 2030.....	57
Tabla 26 - Resultados SARIMA - Comparación valores anuales 2025 .....	58
Tabla 27 - Comparativa modelos Machine Learning.....	59

Tabla 28 - Variables exógenas para ML .....	59
Tabla 29 - PM <sub>10</sub> - Resultados RMSE ML .....	60
Tabla 30 - O <sub>3</sub> - Resultados RMSE ML .....	60
Tabla 31 - NO <sub>2</sub> - Resultados RMSE ML .....	60
Tabla 32 - PM <sub>2.5</sub> - Resultados RMSE ML .....	60
Tabla 33 - Comparativa modelos Deep Learning .....	62
Tabla 34 - Variables derivadas modelos DL .....	63
Tabla 35 - O <sub>3</sub> - Resultados RMSE DL .....	63
Tabla 36 - % Tendencia anual - O3 - CENTRO - HISTÓRICO .....	69
Tabla 37 - % Tendencia anual - O3 - CENTRO - CNN1D_TRAF .....	69
Tabla 38 - % Tendencia anual - O3 - CENTRO - SARIMÁ .....	69
Tabla 39 - Resumen modelos menor RMSE .....	69
Tabla 40 - Resumen Mejores modelos .....	70
Tabla 41 - PM <sub>10</sub> - SUR - Variabilidad.....	73
 Figura 1 - Valores ausentes en MEDICION_ugm3 por año (2001–2025) .....	14
Figura 2 - % NAs por magnitud - Eliminación magnitudes 42, 43 y 44 .....	14
Figura 3 - Resumen imputación datos tráfico por zona y modelo .....	16
Figura 4 - Mejor modelo por zona imputación datos tráfico .....	16
Figura 5 - Datos anuales incendios tras imputación .....	17
Figura 6 - Datos trimestrales incendios tras imputación .....	17
Figura 7 - Temperatura - Tendencia ascendente.....	21
Figura 8 - Presión atmosférica - Valores mayores en invierno.....	21
Figura 9 - Humedad relativa - Distribución bimodal .....	21
Figura 10 - Nubosidad - Distribución unimodal .....	21
Figura 11 - Radiación solar - Distribución bimodal.....	21
Figura 12 - Tendencia ascendente - Precipitaciones .....	21
Figura 13 - Nieve - Filomena .....	21
Figura 14 - Viento - Predominio viento suroeste (10m).....	21
Figura 15 - Viento - Predominio viento suroeste (100m).....	22
Figura 16 - Incendios Forestales - Ávila ago2021.....	22
Figura 17 - Temperatura - Estacionalidad marcada.....	22
Figura 18 - Radiación Solar - Estacionalidad marcada .....	22
Figura 19 - Velocidad viento - variabilidad intranual .....	22
Figura 20 - Precipitaciones - variabilidad intranual .....	22
Figura 21 - Trafico - Tendencia anual - Zona Centro .....	23
Figura 22 - Trafico - Tendencia anual - Zona Norte .....	23
Figura 23 - Trafico - Tendencia anual - Zona Sur .....	23
Figura 24 - Trafico - Tendencia diaria - Zona Centro .....	23
Figura 25 - Trafico - Tendencia diaria - Zona Norte .....	23
Figura 26 - Trafico - Tendencia diaria - Zona Sur .....	24
Figura 27 - PM <sub>10</sub> - Distribución NORTE.....	28
Figura 28 - PM <sub>10</sub> - Distribución CENTRO .....	28
Figura 29 - PM <sub>10</sub> - Distribución SUR.....	28
Figura 30 - PM <sub>10</sub> - Boxplot mensual NORTE .....	28
Figura 31 - PM <sub>10</sub> - Boxplot mensual CENTRO .....	28
Figura 32 - PM <sub>10</sub> - Boxplot mensual SUR .....	28
Figura 33 - PM <sub>10</sub> - Boxplot semanal NORTE .....	29
Figura 34 - PM <sub>10</sub> - Boxplot semanal CENTRO .....	29
Figura 35 - PM <sub>10</sub> - Boxplot semanal SUR .....	29
Figura 36 - PM <sub>2,5</sub> - Distribución NORTE .....	29
Figura 37 - PM <sub>2,5</sub> - Distribución CENTRO .....	29
Figura 38 - PM <sub>2,5</sub> - Distribución SUR .....	29
Figura 39 - PM <sub>2,5</sub> - Boxplot mensual NORTE .....	29
Figura 40 - PM <sub>2,5</sub> - Boxplot mensual CENTRO .....	29

Figura 41 - PM <sub>2,5</sub> - Boxplot mensual SUR .....	29
Figura 42 - PM <sub>2,5</sub> - Boxplot semanal NORTE .....	29
Figura 43 - PM <sub>2,5</sub> - Boxplot semanal CENTRO .....	29
Figura 44 - PM <sub>2,5</sub> - Boxplot semanal SUR .....	29
Figura 45 - NO <sub>2</sub> - Distribución NORTE .....	30
Figura 46 - NO <sub>2</sub> - Distribución CENTRO .....	30
Figura 47 - NO <sub>2</sub> - Distribución SUR .....	30
Figura 48 - NO <sub>2</sub> - Boxplot mensual NORTE .....	30
Figura 49 - NO <sub>2</sub> - Boxplot mensual CENTRO .....	30
Figura 50 - NO <sub>2</sub> - Boxplot mensual SUR .....	30
Figura 51 - NO <sub>2</sub> - Boxplot semanal NORTE .....	30
Figura 52 - NO <sub>2</sub> - Boxplot semanal CENTRO .....	30
Figura 53 - NO <sub>2</sub> - Boxplot semanal SUR .....	30
Figura 54 - O <sub>3</sub> - Distribución NORTE .....	31
Figura 55 - O <sub>3</sub> - Distribución CENTRO .....	31
Figura 56 - O <sub>3</sub> - Distribución SUR .....	31
Figura 57 - O <sub>3</sub> - Boxplot mensual NORTE .....	31
Figura 58 - O <sub>3</sub> - Boxplot mensual CENTRO .....	31
Figura 59 - O <sub>3</sub> - Boxplot mensual SUR .....	31
Figura 60 - O <sub>3</sub> - Boxplot semanal NORTE .....	31
Figura 61 - O <sub>3</sub> - Boxplot semanal CENTRO .....	31
Figura 62 - O <sub>3</sub> - Boxplot semanal SUR .....	31
Figura 63 - PM <sub>10</sub> - Tendencia diaria NORTE .....	33
Figura 64 - PM <sub>10</sub> - Tendencia mensual NORTE .....	33
Figura 65 - PM <sub>10</sub> - Tendencia trimestral NORTE .....	33
Figura 66 - PM <sub>10</sub> - Tendencia anual NORTE .....	33
Figura 67 - PM <sub>10</sub> - Tendencia diaria CENTRO .....	34
Figura 68 - PM <sub>10</sub> - Tendencia mensual CENTRO .....	34
Figura 69 - PM <sub>10</sub> - Tendencia trimestral CENTRO .....	34
Figura 70 - PM <sub>10</sub> - Tendencia anual CENTRO .....	34
Figura 71 - PM <sub>10</sub> - Tendencia diaria SUR .....	34
Figura 72 - PM <sub>10</sub> - Tendencia mensual SUR .....	34
Figura 73 - PM <sub>10</sub> - Tendencia trimestral SUR .....	34
Figura 74 - PM <sub>10</sub> - Tendencia anual SUR .....	34
Figura 75 - PM <sub>2,5</sub> - Tendencia diaria NORTE .....	35
Figura 76 - PM <sub>2,5</sub> - Tendencia mensual NORTE .....	35
Figura 77 - PM <sub>2,5</sub> - Tendencia trimestral NORTE .....	35
Figura 78 - PM <sub>2,5</sub> - Tendencia anual NORTE .....	35
Figura 79 - PM <sub>2,5</sub> - Tendencia diaria CENTRO .....	35
Figura 80 - PM <sub>2,5</sub> - Tendencia mensual CENTRO .....	35
Figura 81 - PM <sub>2,5</sub> - Tendencia trimestral CENTRO .....	35
Figura 82 - PM <sub>2,5</sub> - Tendencia anual CENTRO .....	35
Figura 83 - PM <sub>2,5</sub> - Tendencia diaria SUR .....	36
Figura 84 - PM <sub>2,5</sub> - Tendencia mensual SUR .....	36
Figura 85 - PM <sub>2,5</sub> - Tendencia trimestral SUR .....	36
Figura 86 - PM <sub>2,5</sub> - Tendencia anual SUR .....	36
Figura 87 - NO <sub>2</sub> - Tendencia diaria NORTE .....	36
Figura 88 - NO <sub>2</sub> - Tendencia mensual NORTE .....	36
Figura 89 - NO <sub>2</sub> - Tendencia trimestral NORTE .....	36
Figura 90 - NO <sub>2</sub> - Tendencia anual NORTE .....	36
Figura 91 - NO <sub>2</sub> - Tendencia diaria CENTRO .....	37
Figura 92 - NO <sub>2</sub> - Tendencia mensual CENTRO .....	37
Figura 93 - NO <sub>2</sub> - Tendencia trimestral CENTRO .....	37
Figura 94 - NO <sub>2</sub> - Tendencia anual CENTRO .....	37
Figura 95 - NO <sub>2</sub> - Tendencia diaria SUR .....	37

Figura 96 - NO <sub>2</sub> - Tendencia mensual SUR .....	37
Figura 97 - NO <sub>2</sub> - Tendencia trimestral SUR .....	37
Figura 98 - NO <sub>2</sub> - Tendencia anual SUR.....	37
Figura 99 - O <sub>3</sub> - Tendencia diaria NORTE .....	38
Figura 100 - O <sub>3</sub> - Tendencia mensual NORTE.....	38
Figura 101 - O <sub>3</sub> - Tendencia trimestral NORTE.....	38
Figura 102 - O <sub>3</sub> - Tendencia anual NORTE.....	38
Figura 103 - O <sub>3</sub> - Tendencia diaria CENTRO .....	38
Figura 104 - O <sub>3</sub> - Tendencia mensual CENTRO .....	38
Figura 105 - O <sub>3</sub> - Tendencia trimestral CENTRO .....	38
Figura 106 - O <sub>3</sub> - Tendencia anual CENTRO .....	38
Figura 107 - O <sub>3</sub> - Tendencia diaria SUR .....	39
Figura 108 - O <sub>3</sub> - Tendencia mensual SUR.....	39
Figura 109 - O <sub>3</sub> - Tendencia trimestral SUR.....	39
Figura 110 - O <sub>3</sub> - Tendencia anual SUR .....	39
Figura 111 - PM <sub>10</sub> - Correlaciones CENTRO.....	41
Figura 112 - PM <sub>10</sub> - Correlaciones NORTE .....	41
Figura 113 - O <sub>3</sub> - Correlaciones CENTRO .....	42
Figura 114 - O <sub>3</sub> - Correlaciones NORTE.....	42
Figura 115 - NO <sub>2</sub> - Correlaciones SUR .....	42
Figura 116 - PM <sub>2,5</sub> - Correlaciones SUR .....	42
Figura 117 - PM <sub>10</sub> - Importancia RMSE NORTE.....	43
Figura 118 - O <sub>3</sub> - Importancia RMSE CENTRO .....	43
Figura 119 - NO <sub>2</sub> - Importancia RMSE CENTRO .....	43
Figura 120 - PM <sub>2,5</sub> - Importancia RMSE SUR .....	43
Figura 121 - Proceso Modelos de predicción.....	47
Figura 122 - PM <sub>10</sub> - ACF NORTE .....	49
Figura 123 - PM <sub>10</sub> - PACF NORTE.....	49
Figura 124 - O <sub>3</sub> - ACF CENTRO .....	49
Figura 125 - O <sub>3</sub> - PACF CENTRO .....	49
Figura 126 - NO <sub>2</sub> - ACF SUR.....	49
Figura 127 - PM <sub>2,5</sub> - PACF SUR .....	49
Figura 128 - PM <sub>10</sub> - SARIMA óptimo NORTE .....	52
Figura 129 - PM <sub>10</sub> - SARIMA óptimo CENTRO .....	52
Figura 130 - PM <sub>10</sub> - SARIMA óptimo SUR .....	52
Figura 131 - O <sub>3</sub> - SARIMA óptimo NORTE .....	53
Figura 132 - O <sub>3</sub> - SARIMA óptimo CENTRO .....	53
Figura 133 - O <sub>3</sub> - SARIMA óptimo SUR .....	53
Figura 134 - NO <sub>2</sub> - SARIMA óptimo NORTE .....	54
Figura 135 - NO <sub>2</sub> - SARIMA óptimo CENTRO .....	54
Figura 136 - NO <sub>2</sub> - SARIMA óptimo SUR .....	54
Figura 137 - PM <sub>2,5</sub> - SARIMA óptimo NORTE.....	55
Figura 138 - PM <sub>2,5</sub> - SARIMA óptimo CENTRO.....	55
Figura 139 . PM <sub>2,5</sub> - SARIMA óptimo SUR .....	55
Figura 140 - PM <sub>10</sub> - Gráfico mensual - Pred 2032.....	56
Figura 141 - O <sub>3</sub> - Gráfico mensual - Pred 2032.....	56
Figura 142 - NO <sub>2</sub> - Gráfico mensual - Pred 2032 .....	56
Figura 143 - PM <sub>2,5</sub> - Gráfico mensual - Pred 2032.....	56
Figura 144 - PM <sub>10</sub> - Gráfico mensual - 2030.....	56
Figura 145 - O <sub>3</sub> - Gráfico mensual - 2030 .....	56
Figura 146 - NO <sub>2</sub> - Gráfico mensual - 2030 .....	56
Figura 147 - PM <sub>2,5</sub> - Gráfico mensual - 2030 .....	56
Figura 148 - PM <sub>10</sub> - Pred 2032 - CatBoost - NORTE .....	61
Figura 149 - PM <sub>10</sub> - Pred 2032 - CatBoost - CENTRO .....	61

Figura 150 - PM <sub>10</sub> - Pred 2032 - XGBoost - SUR .....	61
Figura 151 - O <sub>3</sub> - Pred 2032 - LightGBM - NORTE .....	61
Figura 152 - O <sub>3</sub> - Pred 2032 - CatBoost - CENTRO .....	61
Figura 153 - O <sub>3</sub> - Pred 2032 - LightGBM - SUR.....	61
Figura 154 - NO <sub>2</sub> - Pred 2032 - XGBoost - NORTE .....	61
Figura 155 - NO <sub>2</sub> - Pred 2032 - XGBoost - CENTRO .....	61
Figura 156 - NO <sub>2</sub> - Pred 2032 - XGBoost - SUR.....	61
Figura 157 - PM <sub>2.5</sub> - Pred 2032 - CatBoost - NORTE.....	61
Figura 158 - PM <sub>2.5</sub> - Pred 2032 - LightGBM - CENTRO.....	61
Figura 159 - PM <sub>2.5</sub> - Pred 2032 - LightGBM - SUR .....	61
Figura 160 - O <sub>3</sub> - Pred 2032 - LSTM - NORTE .....	63
Figura 161 - O <sub>3</sub> - Pred 2032 - CNN1D - CENTRO .....	63
Figura 162 - PM <sub>10</sub> - Ranking mejores modelos por zona .....	64
Figura 163 - O <sub>3</sub> - Ranking mejores modelos por zona .....	64
Figura 164 - NO <sub>2</sub> - Ranking mejores modelos por zona.....	64
Figura 165 - PM <sub>2.5</sub> - Ranking mejores modelos por zona .....	64
Figura 166 - PM <sub>10</sub> - Predicciones anuales 2030 mejores modelos .....	65
Figura 167 - O <sub>3</sub> - Predicciones anuales 2030 mejores modelos .....	65
Figura 168 - PM <sub>10</sub> - Predicciones anuales 2030 mejores modelos .....	65
Figura 169 - PM <sub>2.5</sub> - Predicciones anuales 2030 mejores modelos .....	65
Figura 170 - PM <sub>10</sub> - Error Test (+Tráfico) - CatBoost - NORTE .....	66
Figura 171 - PM <sub>10</sub> - Error Test (+Tráfico) - CatBoost - CENTRO .....	66
Figura 172 - PM <sub>10</sub> - Error Test (+Tráfico) - CatBoost - SUR .....	66
Figura 173 - O <sub>3</sub> - pred 2032 (+Tráfico) - LightGBM - NORTE .....	66
Figura 174 - O <sub>3</sub> - pred 2032 (+Tráfico) - CNN1D - CENTRO .....	66
Figura 175 - O <sub>3</sub> - pred 2032 (+Tráfico) - LightGBM - SUR .....	66
Figura 176 - NO <sub>2</sub> - Error Test (+Tráfico) - XGBoost - NORTE.....	66
Figura 177 - NO <sub>2</sub> - Error Test (+Tráfico) - XGBoost - CENTRO.....	66
Figura 178 - NO <sub>2</sub> - Error Test (+Tráfico) - XGBoost - SUR .....	66
Figura 179 - PM <sub>2.5</sub> - Error Test (+Tráfico) - CatBoost - NORTE .....	66
Figura 180 - PM <sub>2.5</sub> - Error Test (+Tráfico) - CatBoost - CENTRO .....	66
Figura 181 - PM <sub>2.5</sub> - Error Test (+Tráfico) - CatBoost - SUR.....	66
Figura 182 - PM <sub>10</sub> - Ranking mejores modelos por zona (+Tráfico) .....	67
Figura 183 - O <sub>3</sub> - Ranking mejores modelos por zona (+Tráfico) .....	67
Figura 184 - NO <sub>2</sub> - Ranking mejores modelos por zona (+Tráfico) .....	67
Figura 185 - PM <sub>2.5</sub> - Ranking mejores modelos por zona (+Tráfico) .....	67
Figura 186 - PM <sub>10</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico) .....	68
Figura 187 - O <sub>3</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico) .....	68
Figura 188 - NO <sub>2</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico).....	68
Figura 189 - PM <sub>2.5</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico).....	68
Figura 190 - O <sub>3</sub> - Predicciones anuales 2030 (+Tráfico) - CENTRO .....	68
Figura 191 - O <sub>3</sub> - Predicciones anuales - CENTRO - CNN1D_TRAF vs CNN1D vs SARIMA.....	68
Figura 192 - O <sub>3</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico) - Definitivo .....	69
Figura 193 - PM <sub>10</sub> - pred 2032 FINAL - CatBoost - NORTE .....	71
Figura 194 - PM <sub>10</sub> - pred 2032 FINAL - CatBoost - CENTRO .....	71
Figura 195 - PM <sub>10</sub> - pred 2032 FINAL - XGBoost - SUR.....	71
Figura 196 - PM <sub>10</sub> - Comparación zonas pred 2030 FINAL .....	71
Figura 197 - PM <sub>10</sub> - Comparación zonas pred 2032 FINAL .....	71
Figura 198 - PM <sub>10</sub> - SUR - Valor límite OMS .....	73
Figura 199 - PM <sub>10</sub> - SUR - Boxplot trimestres.....	73
Figura 200 - NO <sub>2</sub> - SUR - Valor límite OMS .....	74
Figura 201 - Comparativa magnitudes NO <sub>2</sub> y O <sub>3</sub> .....	74

Figura 202 - O <sub>3</sub> - SUR - Valor límite OMS .....	74
Figura 203 - PM <sub>2.5</sub> - SUR - Valor límite OMS .....	75
Figura 204 - PM <sub>2.5</sub> - SUR - Boxplot trimestres .....	75

# 1. INTRODUCCIÓN

## 1.1. CONTEXTO Y MOTIVACIÓN

La calidad del aire es uno de los principales retos ambientales y de salud pública en las áreas metropolitanas. La concentración de población, el tráfico rodado y la actividad industrial han incrementado los niveles de contaminantes atmosféricos que afectan a la salud y a los ecosistemas. Según la OMS, la exposición prolongada a partículas finas y gases nocivos causa millones de muertes prematuras cada año, lo que ha impulsado a las administraciones a fortalecer las políticas de control y planificación ambiental.

El presente documento se centra en la **ciudad de Madrid**, donde las concentraciones de **PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub> y O<sub>3</sub>** superan de forma recurrente los valores recomendados por la OMS. Para evaluar su evolución y estimar escenarios futuros, se analizan series históricas mensuales construidas a partir de datos diarios procedentes de la red de estaciones de calidad del aire, agrupadas en tres zonas de estudio (**Norte, Centro y Sur**) y complementadas con variables meteorológicas, de tráfico y de incendios forestales.

Sobre estas series se aplican distintos enfoques de modelización predictiva: modelos clásicos **univariantes** de series temporales (**ARIMA** y **SARIMA**) y modelos **multivariantes** basados en técnicas de aprendizaje automático (**XGBoost**, **LightGBM** y **CatBoost**) y de aprendizaje profundo (**LSTM** y **CNN**). La comparación entre estos métodos permite evaluar su capacidad e idoneidad para capturar la dinámica temporal de los contaminantes y estimar su comportamiento futuro bajo diferentes configuraciones.

Los resultados, evaluados mediante el **RMSE** y proyecciones hasta 2032, indican una tendencia general decreciente en las concentraciones de contaminantes, coherente con las políticas ambientales vigentes. No obstante, el cumplimiento de los valores recomendados por la OMS en 2030 no está plenamente garantizado. Además, este hallazgo cobra especial relevancia tras el reciente compromiso de la Unión Europea de reducir un **90%** las emisiones de gases de efecto invernadero para **2040** respecto a **1990**.

El documento se estructura en seis bloques divididos en un total de nueve capítulos. El primer bloque expone el contexto, la motivación y el marco institucional. El segundo define los objetivos y describe la metodología empleada. El tercero detalla la base de datos, los lenguajes y librerías utilizadas, junto con el análisis exploratorio realizado. El cuarto presenta los modelos de predicción y sus principales resultados. Finalmente, los capítulos quinto y sexto recogen las conclusiones y las referencias bibliográficas.

## 1.2. MARCO INSTITUCIONAL

La mejora de la calidad del aire forma parte esencial de las políticas europeas de transición ecológica y desarrollo sostenible. La Agenda 2030 de las Naciones Unidas, adoptada por España en 2015, recoge en sus **ODS 11 y 13** la necesidad de reducir los impactos ambientales de las ciudades y de mitigar el cambio climático.

En este marco, la **OMS** actualizó en 2021 sus **Guías de Calidad del Aire**, estableciendo valores recomendados más estrictos para contaminantes como PM<sub>2.5</sub>, PM<sub>10</sub>, NO<sub>2</sub> y O<sub>3</sub>. Estos valores sirven de referencia científica para la legislación ambiental europea y nacional, y los objetivos de calidad del aire para 2030 se inspiran directamente en ellos, con el propósito de acercar los límites legales a los estándares sanitarios.

A nivel europeo, la **Directiva 2008/50/CE** y sus posteriores revisiones fijan los valores límite y los procedimientos de evaluación de la calidad del aire. Tal como se expuso en secciones previas, en 2025 la Comisión Europea reforzó este marco al comprometerse a reducir un 90 % las emisiones de gases de efecto invernadero para 2040 respecto a 1990, avanzando hacia la neutralidad climática.

En España, la **Ley 34/2007**, de calidad del aire y protección de la atmósfera, establece los principios generales y las competencias administrativas, mientras que el **Real Decreto 102/2011** concreta los valores límite y transpone la citada directiva europea. Este Real Decreto fue actualizado por **Real Decreto 34/2023**. A nivel regional, la Comunidad de Madrid aplica planes y estrategias orientados a la reducción de emisiones y la mejora de la salud pública.

En este contexto normativo, el presente estudio se integra en las estrategias de descarbonización y control de la contaminación atmosférica, aportando una visión predictiva basada en datos históricos y técnicas de big data, con el fin de conseguir políticas más eficaces hacia una ciudad más saludable y sostenible.

## 2. OBJETO Y ALCANCE

### 2.1. OBJETIVO GENERAL

El objetivo general de este trabajo es evaluar la viabilidad de cumplir con los objetivos de calidad del aire establecidos en la Agenda 2030 y en la normativa europea, mediante la estimación y predicción de las concentraciones de los principales contaminantes atmosféricos en la ciudad de Madrid. Para ello, se desarrollan y comparan modelos de predicción univariantes y multivariantes.

El análisis tiene como finalidad determinar la tendencia futura de las sustancias más perjudiciales para la salud humana, partículas en suspensión ( $PM_{10}$  y  $PM_{2.5}$ ), dióxido de nitrógeno ( $NO_2$ ) y ozono troposférico ( $O_3$ ), y **evaluar si, bajo las políticas y condiciones actuales, resulta plausible alcanzar los valores recomendados por la OMS en el horizonte 2030**.

### 2.2. OBJETIVOS ESPECÍFICOS

Para alcanzar este objetivo general, se plantean los siguientes objetivos específicos:

1. Integrar y depurar bases de datos procedentes de estaciones de medición atmosférica, así como de diversas fuentes de variables exógenas (meteorológicas, de tráfico y ambientales).
2. Analizar la evolución temporal de los principales contaminantes ( $PM_{10}$ ,  $PM_{2.5}$  y  $NO_2$ ), identificando tendencias, patrones estacionales y correlaciones entre variables, así como determinar las variables más relevantes para la predicción de los valores futuros de la variable objetivo.
3. Aplicar y comparar distintos enfoques de modelización predictiva, incluyendo:
  - 3.1. Modelos clásicos de series temporales univariantes (ARIMA, SARIMA).
  - 3.2. Algoritmos de ML (XGBoost, LightGBM, CatBoost).
  - 3.3. Modelos de DL (LSTM y CNN).
4. Evaluar el rendimiento de los modelos mediante métricas de error (RMSE) y seleccionar el enfoque con mayor precisión y estabilidad predictiva, en función de los resultados obtenidos.
5. Proyectar las concentraciones futuras de contaminantes hasta el horizonte 2032, prestando especial atención al año 2030, con el fin de valorar la coherencia con los límites establecidos por la OMS.
6. Interpretar los resultados.

## 2.3. ALCANCE Y LIMITACIONES

El alcance de este estudio se circunscribe a la **ciudad de Madrid**, tomando como referencia datos históricos comprendidos entre los años 2001 y 2025. No obstante, tal y como se detallará a lo largo del documento, se empleará únicamente una parte de este rango temporal para el análisis y la modelización. El estudio se estructura por zonas, Norte, Centro y Sur, dividiendo la ciudad en función de su latitud, con el objetivo de identificar posibles diferencias espaciales en la evolución de los contaminantes atmosféricos.

En la fase de predicción, el análisis se centrará en cuatro contaminantes clave: **PM<sub>10</sub>**, **PM<sub>2.5</sub>**, **NO<sub>2</sub>** y **O<sub>3</sub>**, seleccionados por su relevancia sanitaria, su disponibilidad de datos y por ser los que presentan mayores dificultades de cumplimiento de la normativa a futuro en la ciudad de Madrid.

Entre las principales limitaciones del estudio se reconocen las siguientes:

1. Heterogeneidad de los datos, que obliga a aplicar procesos intensivos de limpieza y homogeneización.
2. Dependencia de la calidad y continuidad histórica de las series, afectadas en algunos casos por ausencia de registros.
3. Elevada correlación entre variables meteorológicas, que reduce la capacidad explicativa de los modelos multivariantes.
4. Dificultad para proyectar variables exógenas futuras, en particular las meteorológicas, necesarias para la modelización a largo plazo.
5. Falta de variables clave no disponibles, como futuras políticas de movilidad, evolución tecnológica o cambios en el parque automovilístico hacia un modelo más electrificado.
6. Limitaciones de la variable tráfico, afectada por inconsistencias antes de 2019 que reducen su utilidad predictiva. Se explicará más adelante.
7. Restricciones de los modelos aplicados, ya que los modelos univariantes no captan shocks externos y los multivariantes pueden perder coherencia estructural en proyecciones largas.
8. Complejidad del ozono, cuya dependencia regional y fotoquímica supera parcialmente el alcance de un análisis local (usando solo datos de la ciudad de Madrid).

## 3. MARCO NORMATIVO Y REGULADOR

### 3.1. NORMATIVA INTERNACIONAL Y EUROPEA

La calidad del aire constituye una prioridad dentro de las políticas ambientales internacionales. Como se indicó anteriormente, la Agenda 2030 de las Naciones Unidas (ONU, 2015) establece en sus Objetivos de Desarrollo Sostenible (ODS 11 y 13) la necesidad de reducir los impactos negativos de las ciudades sobre el medio ambiente y de mitigar los efectos del cambio climático.

#### i. Normativa europea

A nivel comunitario, la Directiva 2008/50/CE, relativa a la calidad del aire ambiente y a una atmósfera más limpia en Europa, constituye el principal instrumento normativo. Esta directiva fija los valores límite y umbrales de alerta para los contaminantes más relevantes y establece los métodos de evaluación, control y comunicación de la información a la ciudadanía:

Contaminante	Tipo de media	Valor límite ( $\mu\text{g}/\text{m}^3$ )	Frecuencia / Condición
<b>PM<sub>10</sub></b>	Media diaria	50	No más de 35 superaciones/año
	Media anual	40	Valor límite anual
<b>PM<sub>2.5</sub></b>	Media anual	25	Valor objetivo para 2015
	Media anual	20	Valor recomendado para 2020
<b>NO<sub>2</sub></b>	Media horaria	200	No más de 18 superaciones/año
	Media anual	40	Valor límite anual
	Umbral alerta	400	Promedio de 3 horas consecutivas
<b>O<sub>3</sub></b>	Media máxima móvil 8 h	120	No más de 25 superaciones/año, promediado en 3 años
	Umbral de información	180	Valor de 1 hora
	Umbral de alerta	240	Valor de 1 hora
<b>SO<sub>2</sub></b>	Media horaria	350	No más de 24 superaciones/año
	Media diaria	125	No más de 3 superaciones/año

Tabla 1 - Directiva 2008/50/CE - Valores límites

Actualmente, la Comisión Europea trabaja en una revisión de dicha directiva con el objetivo de alinear los límites legales de concentración de contaminantes con los valores recomendados por la OMS para 2030. Además, en noviembre de 2025, la Comisión anunció un nuevo compromiso: reducir en un 90 % las emisiones de gases de efecto invernadero para 2040 respecto a los niveles de 1990.

## ii. Guías de Calidad del Aire de la OMS

En 2021, la OMS publicó la actualización de sus Guías de Calidad del Aire Globales, que constituyen el referente científico internacional en materia de contaminación atmosférica y salud pública. Estas guías definen los niveles de concentración de contaminantes que, según la evidencia epidemiológica disponible, no deberían superarse para evitar efectos adversos sobre la salud humana.

Los valores recomendados por la OMS para los principales contaminantes son los siguientes:

Contaminante	Promedio anual recomendado ( $\mu\text{g}/\text{m}^3$ )	Tipo de efecto principal	Link
<b>PM<sub>2.5</sub></b>	5	Enfermedades respiratorias y cardiovasculares	<a href="#">Link</a>
<b>PM<sub>10</sub></b>	15	Irritación pulmonar, mortalidad prematura	<a href="#">Link</a>
<b>NO<sub>2</sub></b>	10	Daño pulmonar, exacerbación del asma – especialmente en niños	<a href="#">Link</a>
<b>O<sub>3</sub></b>	60 (promedio 8 h)	Estrés oxidativo, reducción de la función pulmonar	<a href="#">Link</a>

Contaminante	Promedio anual recomendado ( $\mu\text{g}/\text{m}^3$ )	Tipo de efecto principal	Link
$\text{SO}_2$	40 (promedio 24 h)	Broncoespasmos, irritación de vías respiratorias	<a href="#">Link</a>

Tabla 2 - Guía OMS - Valores recomendados y efectos

Estas recomendaciones son más estrictas que los valores límite actualmente vigentes en la Directiva 2008/50/CE, lo que evidencia la necesidad de reforzar las políticas de reducción de emisiones en los Estados miembros.

Es por ello que los objetivos europeos de calidad del aire para 2030 se fundamentan en estos valores de referencia, con el propósito de aproximar los límites legales a los niveles de seguridad sanitaria establecidos por la OMS. Por tanto, esta guía constituye un marco técnico esencial para interpretar los resultados obtenidos en este estudio y evaluar la magnitud del reto que afronta la ciudad de Madrid.

### 3.2. NORMATIVA NACIONAL

En el ordenamiento jurídico español, la Ley 34/2007, de calidad del aire y protección de la atmósfera, constituye la norma marco que define los principios generales de actuación y la distribución de competencias entre las administraciones públicas.

Su desarrollo reglamentario se materializa en el Real Decreto 102/2011, relativo a la mejora de la calidad del aire, que transpone íntegramente la Directiva 2008/50/CE al ordenamiento jurídico español. Este Real Decreto establece los valores límite, objetivos de calidad y umbrales de información y alerta para los principales contaminantes atmosféricos, así como los procedimientos para su evaluación, gestión y comunicación pública.

Posteriormente, el Real Decreto 34/2023, de 24 de enero, ha modificado parcialmente el Real Decreto 102/2011 con el propósito de actualizar sus anexos técnicos, modernizar los métodos de medición, mejorar la calidad y trazabilidad de los datos y reforzar la cooperación entre las administraciones competentes. Esta actualización se alinea con las Guías de Calidad del Aire de la OMS (2021) y los objetivos del Pacto Verde Europeo, incorporando una mayor consideración de los efectos sobre la salud y los grupos vulnerables.

No obstante, el Real Decreto 34/2023 no modifica los valores límite ni los umbrales de información y alerta establecidos por el Real Decreto 102/2011, que continúan plenamente vigentes.

Como instrumento de planificación, el Plan Nacional de Calidad del Aire y Protección de la Atmósfera (Planes Aire), elaborado por el Ministerio para la Transición Ecológica y el Reto Demográfico (MITECO), tiene como finalidad reducir las emisiones de los contaminantes atmosféricos más relevantes, garantizar el cumplimiento de los valores legales y reducir la exposición de la población. Los Planes Aire incluyen medidas sectoriales en transporte, industria, agricultura, ganadería y energía, y se actualizan periódicamente para integrar las políticas de descarbonización, movilidad sostenible y transición energética, en coherencia con los compromisos de la Agenda 2030 y las estrategias europeas de neutralidad climática.

### 3.3. NORMATIVA AUTONÓMICA Y MUNICIPAL

En el ámbito regional, la Comunidad de Madrid ha desarrollado distintos instrumentos estratégicos orientados a la mejora de la calidad del aire y la mitigación del cambio climático.

El más reciente es el Plan de Calidad del Aire y Cambio Climático de la Comunidad de Madrid (2024–2030), aprobado mediante la Orden 2322/2023, de 29 de diciembre (BOCM nº 12, de 15 de enero de 2024), que sustituye al anterior Plan Azul+ (2018–2030). Este plan tiene como finalidad reducir las emisiones de contaminantes atmosféricos —principalmente NO<sub>2</sub>, PM<sub>10</sub> y PM<sub>2.5</sub>—, mejorar la eficiencia energética y promover una movilidad más limpia, en consonancia con los objetivos nacionales y europeos de calidad del aire.

Entre sus principales ejes de actuación destacan la implantación de zonas de bajas emisiones, la electrificación progresiva del transporte, el impulso de las energías renovables en el sector industrial y residencial y el seguimiento de los indicadores de calidad del aire mediante sistemas de evaluación continuos.

A nivel municipal, el Ayuntamiento de Madrid mantiene su propio Plan de Calidad del Aire y Cambio Climático “Madrid 360”, centrado en el ámbito urbano y complementario a la estrategia regional, con medidas específicas de restricción al tráfico, promoción del transporte público y renovación del parque automovilístico.

### 3.4. IMPLICACIONES DEL ESTUDIO

Los valores guía de la OMS y los límites fijados por la Directiva 2008/50/CE constituyen referencias fundamentales para analizar los datos históricos de la variable objetivo y contextualizar los resultados obtenidos en términos de salud pública y sostenibilidad ambiental. Estos valores permiten cuantificar la brecha existente entre la situación actual y los objetivos internacionales, aportando una perspectiva comparativa que refuerza la interpretación de las proyecciones.

Asimismo, el marco normativo y regulador descrito constituye la base de referencia para la evaluación futura de los resultados de este estudio. La comparación entre las predicciones obtenidas mediante modelos de predicción y los valores límite legales establecidos en la normativa europea, nacional y autonómica permitirá determinar el grado de cumplimiento de los objetivos de calidad del aire previstos para 2030 en la ciudad de Madrid.

De esta forma, los resultados del presente estudio no solo poseen relevancia técnica y científica, sino que pueden servir como herramienta de apoyo a la planificación ambiental y a la toma de decisiones públicas, en coherencia con los planes de mejora de la calidad del aire promovidos a escala nacional, autonómica y local.

## 4. LENGUAJE DE PROGRAMACIÓN Y LIBRERIAS

El desarrollo del presente estudio se ha realizado íntegramente en el lenguaje de programación Python, empleando un amplio conjunto de librerías científicas orientadas a la manipulación, análisis y modelización de datos a gran escala.

El trabajo se ha ejecutado en el entorno Jupyter Notebook, que facilita la integración del código con su documentación y la visualización interactiva de resultados.

Las librerías utilizadas se agrupan a continuación según su funcionalidad:

Categoría	Librerías principales	Función en el estudio
Procesamiento y gestión de datos	pandas, numpy, openpyxl, os, pathlib, re, json, unicodedata, itertools, skimpy, ydata_profiling	Lectura, limpieza, transformación y estructuración de bases de datos; análisis exploratorio automático; tipificación de

Categoría	Librerías principales	Función en el estudio
		variables y generación de perfiles descriptivos.
<b>Preprocesamiento y preparación de variables</b>	scikit-learn (StandardScaler, PCA), CategoricalDtype, random	Escalado y normalización de datos, reducción de dimensionalidad, tipificación de variables categóricas y control de aleatoriedad en los procesos de muestreo y entrenamiento para garantizar la reproducibilidad.
<b>Análisis estadístico y series temporales</b>	statsmodels (ARIMA, SARIMAX, adfuller, plot_acf, plot_pacf, ExponentialSmoothing), prophet, itertools	Modelización estadística clásica de series temporales, identificación de parámetros, análisis de estacionariedad y detección de patrones estacionales y de tendencia.
<b>Modelos predictivos de Machine Learning</b>	xgboost, lightgbm, catboost, sklearn (KNeighborsRegressor)	Entrenamiento de modelos supervisados basados en gradient boosting y k-nearest neighbors; comparación con modelos clásicos y optimización de hiperparámetros.
<b>Modelos de Deep Learning</b>	tensorflow, keras (capas: Sequential, LSTM, Conv1D, Dense, Dropout, Flatten, GlobalAveragePooling1D, EarlyStopping, Adam)	Diseño e implementación de redes neuronales recurrentes (LSTM) y convolucionales (CNN) para la predicción no lineal y dependiente del tiempo.
<b>Evaluación y validación de modelos</b>	scikit-learn (mean_absolute_error, mean_squared_error), permutation_importance	Cálculo de métricas de precisión (MAE, RMSE) aplicadas a todos los modelos; estimación de la importancia de variables mediante perturbación de características.

Tabla 3 - Librerías usadas

## 5. BASE DE DATOS Y PREPROCESADO

### 5.1. FUENTE DE DATOS

El conjunto de datos empleado en este estudio procede de distintas fuentes, que garantizan la fiabilidad y comparabilidad de la información. Los datos utilizados se estructuran en cuatro bloques principales:

## i. Datos de calidad del aire

Los datos de calidad del aire empleados en este estudio proceden del *Portal de Datos Abiertos del Ayuntamiento de Madrid*, disponible en la siguiente dirección oficial: [Link](#)

En este portal, el Ayuntamiento pone a disposición pública las mediciones automáticas y validadas de contaminantes atmosféricos realizadas por la Red de Vigilancia de la Calidad del Aire, con estaciones distribuidas por toda la ciudad.

Los registros están disponibles con periodicidad horaria y diaria; sin embargo, en **este estudio se han empleado los datos diarios**, por constituir una base más adecuada para el manejo de la información, la interpretabilidad del problema, el análisis temporal y la modelización predictiva.

Los archivos descargados se presentan en formato CSV, y en el mismo portal se incluye el documento técnico titulado “Intérprete de ficheros de datos horarios, diarios y tiempo real”, donde se detalla la estructura de los registros y el significado de cada campo, lo que permite reproducir con precisión el proceso de lectura y decodificación de los datos.

Según dicha documentación, cada registro de la base de datos contiene los siguientes campos:

- **PROVINCIA**: Código de la provincia de Madrid. No aporta información puesto que todas las instancias contienen el mismo valor.
- **MUNICIPIO**: Código del municipio de Madrid (capital de provincia). No aporta información puesto que todas las instancias contienen el mismo valor.
- **ESTACIÓN**: Código de la estación en la que se lleva a cabo la medida correspondiente.
- **MAGNITUD**: Código de la magnitud, o mejor dicho, del componente que se mide.
- **PUNTO MUESTREO**: El código aporta información adicional respecto a las variables anteriores, indicando la técnica empleada para medir el componente correspondiente. Este integra la información de provincia (28 para Madrid), municipio (079), estación y magnitud, lo que permite identificar de forma única cada punto de medición dentro de la red.
- **ANÓ**: Año en el que se toma la medición.
- **MES**: Mes en el que se toma la medición.
- **Dxx**: Variables que van desde xx=01 hasta xx=31, representando el valor de la medición diaria de cada mes.
- **Vxx**: Variables que van desde xx=01 hasta xx=31, representando la validación de la medida correspondiente. Por ejemplo, si V01 = V, el valor representado en D01 será válido; por el contrario, si V01 = N, el valor representado en D01 será inválido.

En cuanto a los diferentes contaminantes disponibles en la medición y su codificación se recogen en el Anexo II del documento técnico anteriormente señalado:

Cod.	Magnitud	Abrevia.	Unidad
01	Dióxido de Azufre	SO <sub>2</sub>	µg/m <sup>3</sup>
06	Monóxido de Carbono	CO	mg/m <sup>3</sup>
07	Monóxido de Nitrógeno	NO	µg/m <sup>3</sup>
08	Dióxido de Nitrógeno	NO <sub>2</sub>	µg/m <sup>3</sup>
09	Partículas < 2.5 µm	PM2.5	µg/m <sup>3</sup>
10	Partículas < 10 µm	PM10	µg/m <sup>3</sup>
12	Óxidos de Nitrógeno	NO <sub>x</sub>	µg/m <sup>3</sup>
14	Ozono	O <sub>3</sub>	µg/m <sup>3</sup>
20	Tolueno	TOL	µg/m <sup>3</sup>
30	Benceno	BEN	µg/m <sup>3</sup>
35	Etilbenceno	EBE	µg/m <sup>3</sup>
37	Metaxileno	MXY	µg/m <sup>3</sup>
38	Paraxileno	PXY	µg/m <sup>3</sup>
39	Ortoxileno	OXY	µg/m <sup>3</sup>
42	Hidrocarburos totales (hexano)	TCH	mg/m <sup>3</sup>
43	Metano	CH <sub>4</sub>	mg/m <sup>3</sup>
44	Hidrocarburos no metánicos (hexano)	NMHC	mg/m <sup>3</sup>
431	Metaparaxileno	MPX	mg/m <sup>3</sup>

Tabla 4 - Magnitudes base de datos calidad del aire

El periodo analizado comprende desde el 1 de enero de 2001 hasta el 30 de junio de 2025, cubriendo un rango temporal suficientemente amplio para estudiar la evolución histórica de los niveles de contaminación y desarrollar modelos predictivos robustos.

## ii. Datos meteorológicos

Los **datos meteorológicos** empleados en este estudio se han obtenido a través del portal **Meteoblue – Historical Weather Data** ([Link](#)), una plataforma reconocida que proporciona información meteorológica.

El archivo descargado, en formato CSV, incluye una amplia gama de parámetros que describen las condiciones atmosféricas generales sobre el municipio de **Madrid**. En total, se incorporaron 22 variables, que se detallan a continuación:

Descripción	Unidad
Temperaturas media, mínima y máxima diarias	°C
Humedad relativa media, mínima y máxima diarias	%
Precipitación total	mm
Espesor nieve	cm
Velocidad viento media, mínima rachas máximas y dirección del viento, tanto a 10 metros como a 100.	km/h / °
Presión atmosférica media al nivel del mar: media, mínima y máxima	hPa
Radiación solar	W/m <sup>2</sup>
Cobertura nubosa total y horas de sol	% / min

Tabla 5 - Variables base de datos meteorológicos

El conjunto de datos cubre el periodo comprendido entre el 1 de enero de 2001 y el 30 de junio de 2025, permitiendo analizar la relación entre las condiciones meteorológicas y los niveles de contaminación atmosférica registrados en la ciudad.

### iii. Datos de tráfico

Los **datos de tráfico** utilizados en este estudio se han obtenido del **Portal de Datos Abiertos del Ayuntamiento de Madrid**, disponible en la siguiente dirección oficial: [Link](#)

Estos registros proceden del Sistema de Aforos Permanentes de la **Ciudad de Madrid**, que recopila datos automáticos sobre el flujo vehicular en distintos puntos de la red viaria.

La estructura y descripción de los campos se detallan en el documento técnico complementario *Estructura del conjunto de datos de Aforos Permanentes*, disponible en el mismo portal anteriormente mencionado

Cada fichero, en formato CSV, contiene los registros horarios de circulación vehicular capturados por las estaciones de aforo.

Según la documentación oficial, los principales campos incluidos son los siguientes:

- Día.
- Estación
- Sentido de circulación y franja horaria (madrugada-día: de 00:00 a 12:00 / tarde-noche: de 12:00 a 00:00)
- Intensidad de tráfico horario (de HOR1 a HOR12).

Las estaciones de aforo, distribuidas por toda la ciudad, operan de forma continua, lo que permite obtener una representación precisa de la intensidad media diaria de tráfico.

El periodo disponible para este conjunto de datos abarca desde enero de 2018 hasta marzo de 2025, ofreciendo una cobertura temporal adecuada para evaluar la influencia del tráfico urbano sobre las concentraciones de contaminantes atmosféricos.

### iv. Datos de incendios

Los **datos de incendios forestales** empleados en este estudio se han obtenido del portal de **datos abiertos de Civio**, disponible en: [Link](#)

Este conjunto de datos recopila información oficial del Ministerio para la Transición Ecológica y el Reto Demográfico (MITERD), procedente del **Inventario Español de Incendios Forestales (EGIF)**.

El portal Civio ofrece una versión consolidada y actualizada de los registros históricos desde 1968 hasta 2023, con un formato accesible y estructurado en archivos CSV.

Cada registro del fichero incluye, entre otros, los siguientes campos principales:

Campo	Descripción
<b>Id</b>	Identificador individual del incendio
<b>fecha</b>	Fecha del incendio (DD-MM-YYYY)
<b>idprovincia</b>	Código de la provincia donde se originó el incendio
<b>idcomunidad</b>	Código de la comunidad autónoma afectada
<b>superficie</b>	Superficie total afectada (hectáreas)

Tabla 6 - Variables base de datos incendios

Para su utilización en este estudio, los registros se filtraron con el fin de seleccionar únicamente los **incendios ocurridos en la Comunidad de Madrid y provincias limítrofes, durante el periodo comprendido entre 2001 y 2023**.

## 5.2. PREPROCESAMIENTO - LIMPIEZA Y TRANSFORMACIÓN

El preprocesamiento de los datos se estructuró en cinco fases sucesivas, cada una documentada en archivos independientes en formato **.ipynb**. Estos archivos contienen el código desarrollado en **lenguaje Python** y se encuentran recopilados en el **Anexo A1**, incluido como carpeta adicional adjunta al presente documento.

Cada bloque aborda una parte específica del tratamiento. A continuación, se describen todos y cada uno de los bloques.

### i. Bloque 1 – Integración y depuración de datos de calidad del aire.

El primer bloque del preprocesado se centró en la lectura, **unificación y limpieza de los registros de calidad del aire** procedentes del Portal de Datos Abiertos del Ayuntamiento de Madrid, correspondientes al período 2001–2025 **y la unificación con los datos meteorológicos**.

Las principales operaciones realizadas fueron:

Etapa del proceso	Descripción de las acciones realizadas
<b>1. Lectura y normalización de ficheros anuales – DATOS CALIDAD DE MADRID</b>	Carga de los archivos anuales de medición y estandarización de sus columnas principales mediante iteración (PROVINCIA, MUNICIPIO, ESTACION, MAGNITUD, PUNTO_MUESTREO, AÑO, MES, D01–D31, V01–V31). Transformación del formato original a un esquema de tipo largo. <i>Véase función process_dataset en archivo 01_Integración y depuración de datos de calidad del aire.ipynb contenido dentro del Anexo A1</i>
<b>2. Conversión temporal y estructuración de fechas – DATOS CALIDAD DE MADRID</b>	Generación de la columna DIA/MES/AÑO con formato YYYY-MM-DD y conversión al tipo datetime64 para su integración con otras fuentes (meteorología, tráfico e incendios). <i>Véase apartado ITERACIÓN CSV del mismo archivo del punto anterior</i>
<b>3. Completitud temporal y creación de combinaciones posibles – DATOS CALIDAD DE MADRID</b>	Se <b>forzó</b> que para cada día existieran mediciones correspondientes a todas las combinaciones posibles de Estación–Magnitud–Técnica (24.049.536 instancias totales). En los casos donde no se registraron datos (por ejemplo, días sin medición en diciembre de 2024), se incorporaron valores NA en la variable MEDICIÓN. <b>95,78% de las instancias NAS</b> . <i>Véase apartado DATAFRAME COMPLETO del mismo archivo del punto anterior</i>
<b>4. Depuración de valores nulos y duplicados – DATOS CALIDAD DE MADRID</b>	Se observa muchas combinaciones que no tienen valores MEDICION en todo el dataframe final. Se eliminan estas combinaciones. Pasamos a <b>30.35% de las instancias NAs</b> . <i>Véase apartado ANÁLISIS COMBINACIONES ESTACION, MAGNITUD, TECNICA del mismo archivo del punto anterior</i>
<b>5. Normalización variable Magnitud – DATOS CALIDAD DE MADRID</b>	En la variable MAGNITUD, algunos valores vienen expresados en mg/m3 y otros en µg/m3, por ello, normalizo la variable para que todos sus valores queden expresados en la misma unidad. <i>Véase</i>

Etapa del proceso	Descripción de las acciones realizadas
	apartado NORMALIZACIÓN VARIABLE MAGNITUD del mismo archivo del punto anterior
<b>6. Lectura datos, renombres variables y conversiones variables numéricas – DATOS METEOROLÓGICOS</b>	Se importaron el archivo meteorológico y se renombraron las variables al español para una mejor identificación. Se ajustaron los tipos de datos (numéricos y de fecha). Véase apartado DATOS METEOROLÓGICOS del mismo archivo del punto anterior
<b>7. Combinación ambas bases de datos</b>	Se integraron ambos conjuntos mediante una unión temporal por fecha, generando un único dataset consolidado con concentraciones de contaminantes y variables meteorológicas. Véase mismo apartado del punto anterior.

Tabla 7 - Bloque 1 – Integración y depuración de datos de calidad del aire.

El resultado final de esta fase fue la obtención de un dataset consolidado, que integra los datos meteorológicos y de calidad del aire registrados en las distintas estaciones de la ciudad de Madrid, con una resolución diaria y cobertura continua desde enero de 2001 hasta diciembre de 2025.

## ii. Bloque 2 – Tratamientos sobre el DataFrame original

El segundo bloque del proceso de preprocesamiento se centró en la depuración y transformación avanzada del DataFrame consolidado obtenido en la fase anterior, con el objetivo de garantizar la completitud temporal, reducir el número de valores ausentes y mejorar la coherencia espacial de las series.

Durante esta fase, se realizaron ajustes sobre las variables meteorológicas y de calidad del aire, abordando la imputación de datos faltantes, la eliminación de registros incompletos o redundantes y la redefinición espacial de las estaciones mediante un sistema de zonificación geográfica.

Asimismo, se descartaron magnitudes y estaciones con altos porcentajes de valores nulos, se integraron las variables meteorológicas corregidas y se generó un nuevo dataset estructurado por zonas.

Las principales operaciones realizadas fueron:

Etapa del proceso	Descripción de las acciones realizadas
<b>1. Registros ausentes variables meteorológicas</b>	Se observó que todos los datos ausentes de las variables meteorológica se concentraban en el <b>30 de junio de 2025</b> . Véase apartado NAN DATA POR CADA VARIABLE en archivo <i>02_Tratamientos sobre el DataFrame original.ipynb</i> contenido dentro del Anexo A1
<b>2. Imputación valores variables meteorológicas</b>	Se procedió con la imputación de los datos meteorológicos del <b>30 de junio de 2025</b> utilizando el promedio de los valores registrados el 30 de junio en años anteriores. Véase apartado IMPUTACIÓN DATOS METEOROLÓGICOS del mismo archivo del punto anterior
<b>3. Análisis valores ausentes variable MEDICION_ugm3</b>	Como se observó del bloque anterior, <b>un 30,35% de las instancias eran valores ausentes</b> . Se analizó dónde se concentraban esos valores ausentes. Dos conclusiones:

Etapa del proceso	Descripción de las acciones realizadas
	<p>1. Gran parte de los valores ausente de la variable se encuentran anterior al año 2010.</p> <p>2. En 2020 hubo un gran número de datos ausentes. Véase <a href="#">Figura 1</a>.</p>
<b>4. Eliminación instancias anteriores a 2010</b>	<p>Se procedió a <b>eliminar las instancias anteriores al 01 de enero de 2010</b>. Se pasó a un <b>17,63%</b> de valores ausentes de la variable <b>MEDICION_ugm3</b>. Véase apartado <b>IMPUTACIÓN DATOS CALIDAD DEL AIRE</b> del mismo archivo de los puntos 1 y 2 anteriores.</p>
<b>5. Valores ausentes magnitud 9 (PM<sub>2.5</sub>)</b>	<p>Se observa que una de las magnitudes que mayor número de valores ausentes aportaba era la magnitud número 9 (<b>PM<sub>2.5</sub></b>). Al tratarse de una variable tan relevante no se pudo presidir de sus instancias por lo que <b>se decidió prescindir de las estaciones que mayor número de NAs aportaban a la magnitud 9</b>. Estas fueron las siguientes: 54, 4, 60, 57 y 56. Véase apartado <b>VALORES AUSENTES - ESTACIÓN 9</b> del mismo archivo del punto anterior.</p>
<b>6. Eliminación magnitudes 42, 43 y 44</b>	<p>Una vez ejecutado el punto anterior se prescindió de las magnitudes que, ahora sí, ya sin la magnitud 9, presentaban mayor número de datos ausentes. Véase <a href="#">Figura 2</a>. Tras esta operación, se pasó a un <b>7.54%</b> de valores ausentes de la variable <b>MEDICION_ugm3</b>.</p>
<b>7. Zonificación de cada estación</b>	<p>A partir de las coordenadas geográficas (latitud y longitud) de cada estación de medición, se clasificó cada registro en una de las nueve zonas establecidas: <b>Noreste, Norte, Noroeste, Este, Centro, Oeste, Sureste, Sur y Suroeste</b>. Los límites que delimitan cada zona se definieron utilizando los percentiles 30 y 60 de las distribuciones de latitud y longitud de todas las estaciones, siguiendo un criterio espacial de representatividad. Véase <a href="#">Tabla 9</a> más abajo. Posteriormente, las instancias fueron agrupadas por zona. Tras esta operación, el porcentaje de <b>valores ausentes en la variable MEDICION_ugm3 se redujo a un 4,87 %</b>.</p>
<b>8. Dataframe final bloque 2</b>	<p>Tras las actuaciones llevadas a cabo en los puntos anteriores, quedó un dataframe con siguientes variables: 1) DIA/MES/ANO, 2) ZONA_ESTACION (NO, NO, NE, O, C, E, SO, S y SE), 3) MAGNITUD (todas las anteriores menos 44, 43 y 42), 4) MEDICION_ugm3 (la variable objetivo) y 5) Las variables meteorológicas)</p>

Tabla 8 - Bloque 2 – Tratamientos sobre el DataFrame original

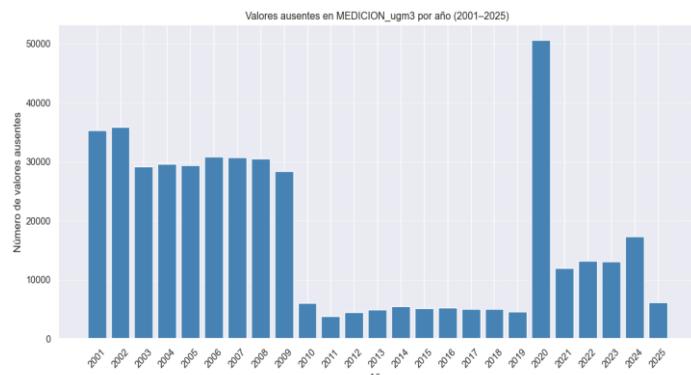


Figura 1 - Valores ausentes en MEDICION\_ugm3 por año (2001–2025)

MAGNITUD	NA_count	total_count	Missing_porcentaje
0	44	10973	48.47
1	43	10973	48.47
2	42	10973	48.47
3	6	10492	26.48
4	1	11457	25.30
5	35	3993	11.76
6	20	3980	11.72
7	30	3947	11.62
8	9	3292	9.69
9	10	4857	8.58
10	14	4557	7.32
11	8	7692	7.15
12	7	7692	7.15
13	12	7692	7.15

Figura 2 - % NAs por magnitud - Eliminación magnitudes 42, 43 y 44

CODIGO	CODIGO_CORTO	ESTACION	LONGITUD	LATITUD
28079004	4	Plaza de España	-3.7122567	40.4238823
28079008	8	Escuelas Aguirre	-3.6823158	40.4215533
28079011	11	Ramón y Cajal	-3.6773491	40.4514734
28079016	16	Arturo Soria	-3.6392422	40.4400457
28079017	17	Villaverde	-3.7133167	40.347147
28079018	18	Farolillo	-3.7318356	40.3947825
28079024	24	Casa de Campo	-3.7473445	40.4193577
28079027	27	Barajas Pueblo	-3.5800258	40.4769179
28079035	35	Plaza del Carmen	-3.7031662	40.4192091
28079036	36	Moratalaz	-3.6453104	40.4079517
28079038	38	Cuatro Caminos	-3.7071303	40.4455439
28079039	39	Barrio del Pilar	-3.7115364	40.4782322
28079040	40	Vallecas	-3.6515286	40.3881478
28079047	47	Méndez Álvaro	-3.6868138	40.3980991
28079048	48	Castellana	-3.6903729	40.4398904
28079049	49	Parque del Retiro	-3.6824999	40.4144444
28079050	50	Plaza Castilla	-3.6887449	40.4655841
28079054	54	Ensanche de Vallecas	-3.6121394	40.3730118
28079055	55	Urb. Embajada	-3.5805649	40.4623628
28079056	56	Plaza Elíptica	-3.7187679	40.3850336
28079057	57	Sanchinarro	-3.6605173	40.4942012
28079058	58	El Pardo	-3.7746101	40.5180701
28079059	59	Juan Carlos I	-3.609031	40.465144
28079060	60	Tres Olivos	-3.6897308	40.5005477

Tabla 9 - Coordenadas estaciones calidad del aire

### iii. Bloque 3 – Integración y depuración de datos de tráfico

El tercer bloque del proceso de preprocesamiento **se centró en la lectura, limpieza e integración de los datos de tráfico urbano de Madrid**, procedentes del Sistema de Aforos Permanentes del *Portal de Datos Abiertos del Ayuntamiento de Madrid*.

El objetivo fue construir una serie temporal homogénea y representativa de la intensidad media diaria de tráfico para su integración posterior con los datos de contaminación y meteorología.

Las principales operaciones realizadas fueron:

Etapa del proceso	Descripción de las acciones realizadas
<b>1. Lectura y unificación de archivos</b>	Se importaron los ficheros mensuales de tráfico en formato .csv, con datos provistos por estaciones, y se unificaron en un único DataFrame mediante pd.concat(). Véase apartado <b>LECTURA DE DATOS DE TRÁFICO</b> en archivo <i>03_Integración y depuración de datos de tráfico.ipynb</i> contenido dentro del Anexo A1
<b>2. Agregación diaria</b>	Se calculó la intensidad diaria por estación y día mediante groupby() y sum(), integrando las franjas horarias en una única por estación al día. Véase apartado <b>AGRUPACIÓN DATOS POR ESTACIÓN Y DIA</b> del mismo archivo del punto anterior.
<b>3. Eliminación estaciones sin medición</b>	Se observaron muchas estaciones donde sus registros eran 0 en muchos días, por lo que se decidió prescindir de ellas para el análisis. Véase <b>IMPUTACIÓN VALORES NULOS DE DATOS DE TRÁFICO</b> del mismo archivo del punto anterior.
<b>4. Zonificación de cada estación</b>	A partir de las coordenadas geográficas (latitud y longitud) de cada estación de medición, se clasificó cada registro en una de las nueve zonas establecidas: <b>Noreste, Norte, Noroeste, Este, Centro, Oeste, Sureste, Sur y Suroeste</b> . Los límites que delimitan cada zona son los mismos que los definidos en el bloque anterior. Véase apartados <b>DATOS ESTACIONES TRÁFICO</b> y <b>CLASIFICACIÓN ZONA A LOS DATOS DE TRÁFICO</b> del mismo archivo del punto anterior. Posteriormente, las instancias fueron agrupadas por zona.
<b>4. Imputación valores fuera de rango</b>	Dado que solo se disponía de registros de tráfico entre noviembre de 2019 y marzo de 2025, se estimaron los valores faltantes para completar la serie histórica desde enero de 2010 hasta junio de 2025. La estimación se abordó mediante dos estrategias complementarias: 1) un modelo determinista con <b>series de Fourier</b> , que capta las estacionalidades semanales y anuales, y 2) un <b>modelo Prophet</b> , que incorpora tendencias no lineales y efectos estacionales flexibles. Ambos modelos se evaluaron, <b>para cada una de las zonas</b> , mediante validación cruzada con ventana móvil, empleando <b>RMSE</b> y <b>MAE</b> como métricas de error, seleccionándose el modelo con mayor precisión y menor varianza de predicción. Para <b>todas las zonas</b> , se observó que <b>el modelo Prophet era el más adecuado</b> . Véase apartado <b>IMPUTACION VALORES FUERA DE RANGO</b> del mismo archivo del punto anterior y figuras <i>Figura 3</i> y <i>Figura 4</i> con los resultados obtenidos para cada uno de los modelos.

Tabla 10 - Bloque 3 – Integración y depuración de datos de tráfico

> Resumen por zona y modelo:
ZONA model rmse_mean rmse_std mae_mean mae_std n_splits
Centro fourier 58518.856671 16229.480048 41957.177598 13711.204973 20
Centro prophet 33455.535658 17012.441508 28953.537469 16846.918460 20
Este fourier 43212.046908 14674.312087 34927.008479 13391.546783 20
Este prophet 28000.240284 11221.070157 15710.281162 9441.895568 20
Noreste fourier 44283.334097 17365.234558 34917.233316 16686.642856 20
Noreste prophet 22079.674774 12384.734692 18139.350891 11438.070889 20
Noroeste fourier 15410.681566 5919.477926 12388.194897 5758.133328 20
Noroeste prophet 7697.741284 3532.286608 6465.379061 3214.426124 20
Norte fourier 35886.758560 13837.188649 28772.560478 12563.008878 20
Norte prophet 19082.519124 9802.170936 15325.280215 8628.617377 20
Oeste fourier 32195.276681 13493.504178 26896.327298 11868.713782 20
Oeste prophet 23488.001487 12982.515858 20697.108659 12943.211942 20
Sur fourier 54001.192714 12969.687859 46872.465638 10638.736581 20
Sur prophet 36977.078343 19160.166230 32332.374421 19445.229679 20
Sureste fourier 16246.752733 5589.842494 13198.586229 5851.749720 20
Sureste prophet 9218.061162 4284.982259 7803.723898 3954.891426 20
Suroeste fourier 27760.170945 9180.156238 22110.650236 8423.693648 20
Suroeste prophet 16821.201193 9149.861074 14820.504398 8355.387250 20

Figura 3 - Resumen imputación datos tráfico por zona y modelo

> Mejor modelo por zona (con sugerencia de ensemble si RMSE ~empata):
ZONA best_model_by_rmse rmse_mean mae_mean n_splits ensemble_suggested
Centro prophet 33455.535658 28953.537469 20 False
Este prophet 28000.240284 15710.281162 20 False
Noreste prophet 22079.674774 18139.350891 20 False
Noroeste prophet 7697.741284 6465.379061 20 False
Norte prophet 19082.519124 15325.280215 20 False
Oeste prophet 23488.001487 20697.108659 20 False
Sur prophet 36977.078343 32332.374421 20 False
Sureste prophet 9218.061162 7803.723898 20 False
Suroeste prophet 16821.201193 14820.504399 20 False

Figura 4 - Mejor modelo por zona imputación datos tráfico

#### iv. Bloque 4 – Integración y tratamiento datos incendios forestales

El cuarto bloque del proceso de preprocesamiento se **centró en la lectura, limpieza y preparación de los datos de incendios forestales** para integrarlos dentro del dataframe final.

El objetivo fue obtener una serie diaria consolidada de incendios en la Comunidad de Madrid y provincias limítrofes (Ávila, Segovia, Guadalajara, Toledo y Cuenca), útil como variable exógena en los modelos predictivos de contaminación atmosférica.

Las principales operaciones realizadas fueron:

Etapa del proceso	Descripción de las acciones realizadas
1. Lectura y elección de variables	Importación del archivo df_incendios.csv y filtrado de los campos relevantes (fecha, idprovincia, idcomunidad y superficie). Eliminación el resto de variables. Véase apartado LECTURA DE DATOS DE INCENDIOS en archivo 04_Integración y tratamiento de los datos de incendios forestales.ipynb contenido dentro del Anexo A1
2. Filtro por provincias cercanas y ajusto fechas	Se filtra el dataframe por la provincia de Madrid y las provincias limítrofes. Además, se toman los valores dentro del rango del dataframe original, esto es, superiores al 01 de enero de 2010. Véase mismo apartado del punto anterior.
3. Agrupación dataframe por DIA/MES/ANO	Se seleccionaron únicamente las columnas DIA/MES/ANO y superficie, generándose nuevas instancias para completar la secuencia temporal diaria entre el 1 de enero de 2010 y la última fecha disponible del dataset. A las fechas sin registro se les asignó un valor cero en la variable superficie, asumiendo ausencia de incendios. Posteriormente, el conjunto se agrupó por fecha, sumando la superficie afectada en los casos con múltiples incendios ocurridos el mismo día. Véase mismo apartado del punto anterior.
4. Estimación datos faltantes	Para estimar los datos faltantes de superficie quemada entre el 10 de noviembre de 2022 y el 30 de junio de 2025, se aplicó un <b>modelo Hurdle</b> , adecuado para series con una alta proporción de valores nulos. Este modelo combina dos componentes: una fase binaria ( <b>modelo Logit</b> ) que estima la probabilidad de ocurrencia de incendios, y una fase continua ( <b>modelo Gamma con enlace log</b> ) que estima la superficie afectada condicionada a que ocurra un incendio.

Etapa del proceso	Descripción de las acciones realizadas
	A partir de variables meteorológicas integradas mediante un merge entre ambos conjuntos de datos, se estimó la superficie diaria esperada, completando así los registros ausentes en la serie de incendios. Véase apartado ESTIMACIÓN DATOS FALTANTES del archivo del punto anterior y las figuras <i>Figura 5</i> y <i>Figura 6</i> tras completar las imputaciones.

Tabla 11 - Bloque 4 – Integración y tratamiento datos incendios forestales

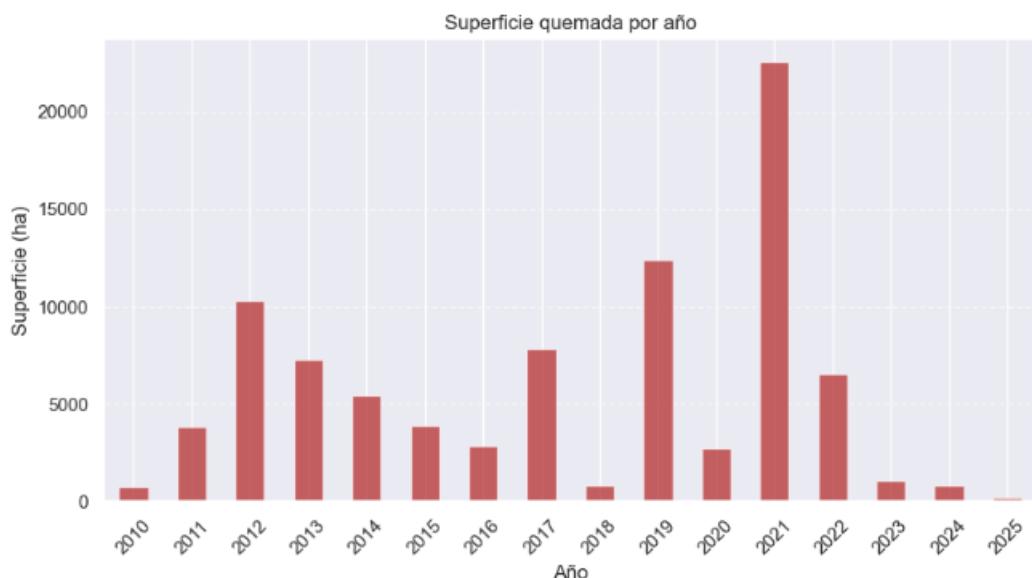


Figura 5 - Datos anuales incendios tras imputación

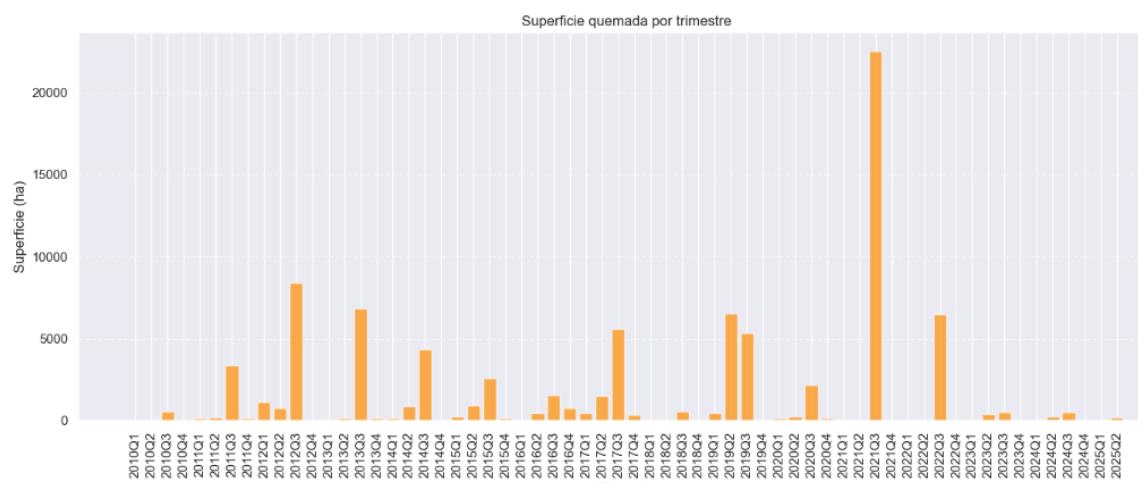


Figura 6 - Datos trimestrales incendios tras imputación

## v. Bloque 5 – Imputación de valores faltantes en calidad del aire

El quinto y último bloque del preprocesamiento se centró **en integrar, depurar y completar la base conjunta de calidad del aire, meteorología, tráfico e incendios, con el fin de obtener un dataset unificado, coherente y sin discontinuidades temporales**. Para ello, se verificó la completitud diaria por magnitud y zona, se incorporaron las instancias faltantes necesarias para garantizar la estructura espacial del conjunto de datos y se fusionaron todas las fuentes disponibles. Posteriormente, se

aplicaron **técnicas de aprendizaje no supervisado** para redefinir las zonas de estudio y, finalmente, se llevaron a cabo distintos procedimientos de imputación, medias zonales y modelos predictivos (**Prophet** y **modelos ML**), hasta obtener una serie íntegra y apta para las fases posteriores de análisis y modelización.

Las principales operaciones realizadas fueron:

Etapa del proceso	Descripción de las acciones realizadas
<b>1. Lectura datos calidad del aire, meteorológicos, tráfico e incendios forestales</b>	Se leen todos los dataframes que se han preparado hasta ahora. Se comprueba que todos ellos tienen el mismo rango temporal (desde el 01 de enero de 2010 al 30 de junio de 2025). Véase apartado <b>LECTURA DATOS DE CALIDAD DEL AIRE, DE TRÁFICO Y DE INCENDIOS</b> en archivo <b>05_Imputación de valores faltantes en calidad del aire.ipynb</b> contenido dentro del Anexo A1
<b>2. Comprobación – Todos los días, todas las magnitudes</b>	Del dataframe de la calidad del aire, se comprueba que, por cada nivel de magnitud, existe, al menos, una medición al día. Véase apartado <b>TODOS LOS DIAS, TODAS LAS MAGNITUDES</b> del archivo del punto anterior.
<b>3. Comprobación – Todos los días, todas las zonas en todas las magnitudes</b>	Se comprueba si, en cada magnitud, existe una medición por día y por zona, es decir, 9 mediciones diferentes al día. Se observa que las magnitudes codificadas como 1, 6, 14, 9, 30, 35 y 20 no cumplen esta condición. Véase apartado <b>TODOS LOS DÍAS, TODAS LAS ZONAS EN TODAS LAS MAGNITUDES</b> del archivo del punto anterior.
<b>4. Inclusión instancias faltantes como NAs</b>	Se incorporaron al dataframe las mediciones ausentes de las zonas que no contaban con registro para las magnitudes indicadas, asignando valores NA con el fin de mantener la estructura temporal y espacial completa del conjunto de datos. Véase apartado <b>ZONAS FALTANTES --&gt; MEDICION NA</b> del archivo del punto anterior.
<b>5. Merge dataframes</b>	Preparados los 3 dataframes, se agrupan en uno solo (df_completo_CA_TRAF_INC) en base a la fecha. Se comprueba que solo existen valores ausentes en la variable MEDICION_ugm3. Véase apartado <b>MERGE DATAFRAMES</b> del archivo del punto anterior.
<b>6. Agrupación zonas - Modelo no supervisado</b>	Se aplicó un Análisis de Componentes Principales (PCA) sobre el dataframe df_completo_CA_TRAF_INC con el fin de identificar similitudes entre zonas de Madrid a partir de variables de calidad del aire, tráfico y meteorología. Se analizó cada magnitud contaminante por separado mediante PCA, observándose patrones espaciales consistentes. Los resultados mostraron una alta similitud entre las zonas Noreste y Noroeste, lo que motivó una reagrupación territorial en tres grandes áreas: NORTE, CENTRO y SUR. Véase apartado <b>MODELO NO SUPERVISADO - ANÁLISIS ZONAS PARECIDAS</b> del archivo del punto anterior.
<b>7. Imputación valores ausentes – por medias</b>	Tras la última agrupación, se imputaron los valores ausente, de modo que: Si hay 1 NA en un día (en una zona), en una magnitud concreta → reemplazar por la media de las otras dos; Si hay 2 NAs → ambos se igualan al valor conocido y Si hay 3 NAs → se dejan como están (NaN). Véase apartado <b>ANÁLISIS VALORES AUSENTES POR COMBINACIÓN ZONA y MAGNITUD</b> del archivo del punto anterior

<b>8. Imputación valores ausentes – por modelos Prophet y ML</b>	<p>Tras la imputación realizada en el apartado anterior, aún persistían valores faltantes en aquellos días en los que no existían mediciones de MEDICION_ugm3 en ninguna de las tres zonas para determinadas magnitudes. Para completar estos huecos se compararon tres modelos: Prophet y modelos de ML (LightGBM y XGBoost). Los resultados mostraron que LightGBM ofrecía el mejor rendimiento y la menor variabilidad, por lo que fue seleccionado como modelo final. Posteriormente, se aplicó un proceso automatizado de imputación por niveles de MAGNITUD, ajustando un modelo LGBM específico para cada contaminante y sustituyendo los valores ausentes por las predicciones generadas. Finalmente, todos los subconjuntos se recombinaron en un único dataset limpio, continuo y libre de valores faltantes. Véase apartado <b>IMPUTACION - COMPARATIVA MODELOS</b> del archivo del punto anterior</p>
--	---

Tabla 12 - Bloque 5 – *Imputación de valores faltantes en calidad del aire*

## 6. ANÁLISIS EXPLORATORIO DE DATOS (EDA)

El análisis exploratorio de datos constituye una fase esencial del estudio, orientada a comprender el comportamiento temporal, estacional y estructural de todas las series analizadas, tanto de calidad del aire como de las variables exógenas, incluyendo los registros meteorológicos, de tráfico y de incendios forestales. A partir del dataset depurado e imputado en fases previas, se examinan las tendencias históricas, los patrones estacionales, las correlaciones entre variables y la presencia de posibles irregularidades o valores anómalos. Este análisis preliminar permite identificar diferencias entre zonas, validar la coherencia global del conjunto de datos y fundamentar la selección de variables y la calibración de los modelos predictivos desarrollados en los capítulos posteriores.

Todo este análisis, tanto para los contaminantes como para las variables exógenas, se recoge de forma detallada en el archivo **06\_EDA.ipynb**, incluido íntegramente dentro del **Anexo A1** del presente documento.

### 6.1. EDA VARIABLES EXÓGENAS

Se diferencian, en este apartado las variables meteorológicas y las relativas a incendios forestales de la variable de tráfico, puesto que esta última, a diferencia del resto, presenta valores diarios específicos para cada una de las tres zonas de estudio. De modo que:

#### i. Variables meteorológicas e incendios forestales

Este subapartado presenta el Análisis Exploratorio de Datos (EDA) de las **variables meteorológicas y de la superficie provocada por incendios forestales**, con el objetivo de comprender su comportamiento general y su posible influencia sobre la calidad del aire. El estudio se centró en analizar la distribución, la estacionalidad y la evolución temporal de estas variables, permitiendo identificar patrones propios del clima de Madrid y del régimen de incendios en su entorno.

Por cada una de estas variables se generaron los siguientes gráficos:

- Histogramas
- Box-plots:
  - En función del día de la semana (lunes-domingo)
  - En función del mes del año (enero-diciembre)
- Gráficos temporales con:
  - Frecuencia diaria
  - Frecuencia mensual
  - Frecuencia trimestral
  - Frecuencia anual

Todos estos gráficos se pueden localizar en la carpeta 01 figs variables meteo inc contenida del Anexo A2.

A partir de estos gráficos se obtiene la siguiente tabla resumen:

CATEGORÍA	VARIABLES CLAVES	TENDENCIA	CONCLUSIÓN
Temperatura	Máxima, media, mínima	↑ Ascenso suave	Ligero aumento de las temperaturas
Presión atmosférica	Máxima, media, mínima	→ Estabilidad general	Estabilidad barométrica; altas presiones en meses de invierno, bajas en verano
Humedad relativa	Mínima, media, máxima	→ Estabilidad general	Valores estables; humedad media con distribución claramente bimodal (concentración de dos rangos de valores)
Nubosidad	% de cobertura nubosa	→ Estabilidad general	Distribución unimodal, sesgada a la derecha, con predominio de cielos despejados.
Radiación solar / insolación	Radiación solar (Wh/m <sup>2</sup> ), horas de sol (min)	→ Estabilidad general	Distribución bimodal, con dos regímenes de radiación solar claramente diferenciados.
Precipitaciones	Precipitaciones	↑ Incremento desde 2021, aunque irregular	Alternancia entre sequías y lluvias torrenciales; Mayor frecuencia en meses de marzo, abril y noviembre
Nieve	Nieve	= Difícil conclusión, picos extremos ocasionales	Destaca un evento en enero 2021 --> Filomena
Viento (10 m)	Velocidad máx, min, media y dirección	→ Estable; eventos extremos puntuales; predominio dirección sur	Circulación atmosférica constante; ráfagas >35 km/h ocasionales (valores máximos); Viento del suroeste predominante
Viento (100 m)	Velocidad máx, min, media y dirección	→ Estable; predominio dirección sur	Viento del suroeste predominante
Superficie incendiada	Superficie incendiada	= Difícil conclusión, picos extremos ocasionales	Menor recurrencia de incendios extensos; riesgo asociado a condiciones locales secas o ventosas; Destaca evento en 2021

Tabla 13 - Tabla resumen EDA variables meteorológicas e incendios

Se muestran a continuación algunos gráficos generados en este apartado que merecen una mención especial en el presente documento:

- Se observa una **tendencia ascendente** de las temperaturas a lo largo de los últimos 15 años. Veáse *Figura 7*.
- Se observa que la **presión atmosférica** presenta mayores valores en invierno y menores en verano. Veáse *Figura 8*.
- Se observa que la **humedad relativa** presenta una distribución bimodal. Veáse *Figura 9*.
- Se observa que la **nubosidad** presenta una distribución unimodal. Predominio de cielos despejados. Veáse *Figura 10*.
- Se observa que la **radiación solar** presenta una distribución bimodal. Veáse *Figura 11*.
- Se observa que existe un incremento de las **precipitaciones**. Veáse *Figura 12*.
- Se aprecia un evento significativo el día 09 de enero de 2021, en la variable **Nieve** → [FILOMENA](#). Veáse *Figura 13*.
- Se observa un predominio de la dirección del viento en sentido suroeste (tanto en 10 como en 100 metros). Vease *Figura 14* y *Figura 15*.
- Se aprecia un evento significativo el día 14 de agosto de 2021, en la variable **Superficie Incendiada** → [INCENDIO](#) EN ÁVILA. Veáse *Figura 16*.

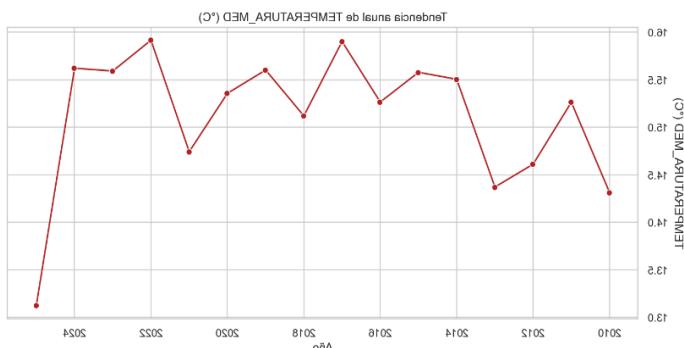


Figura 7 - Temperatura - Tendencia ascendente

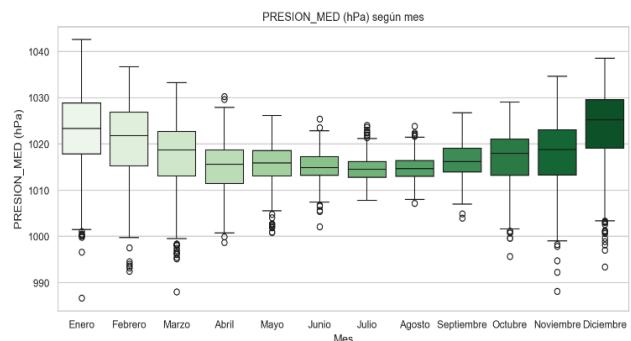


Figura 8 - Presión atmosférica - Valores mayores en invierno

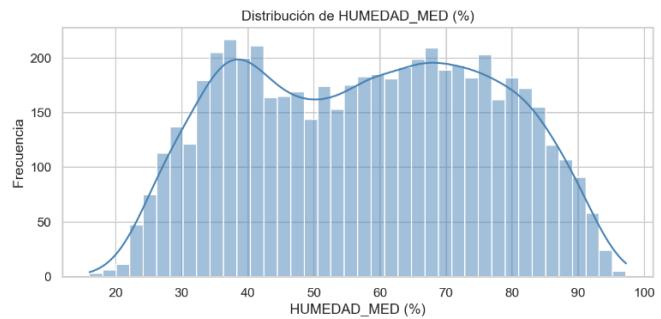


Figura 9 - Humedad relativa - Distribución bimodal

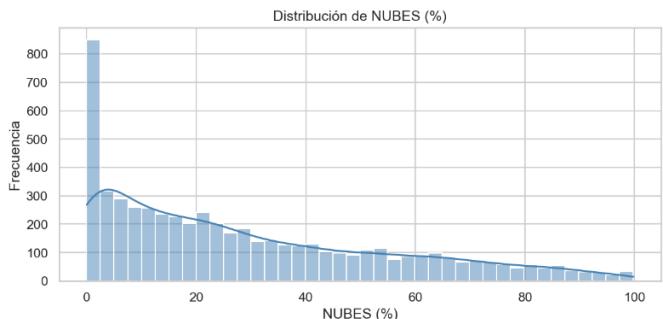


Figura 10 - Nubosidad - Distribución unimodal

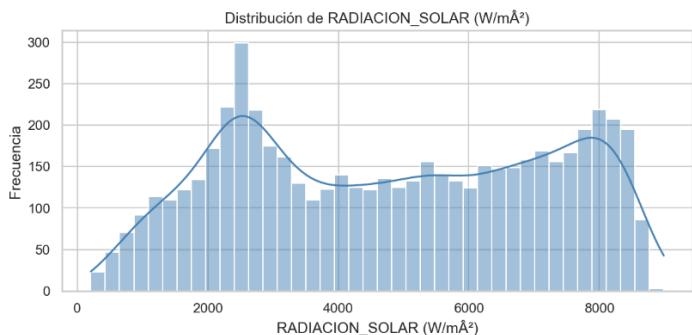


Figura 11 - Radiación solar - Distribución bimodal

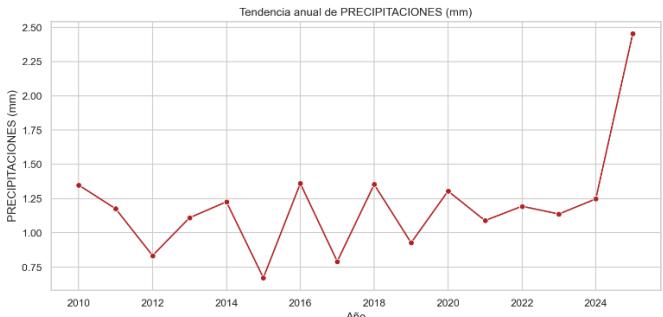


Figura 12 - Tendencia ascendente - Precipitaciones

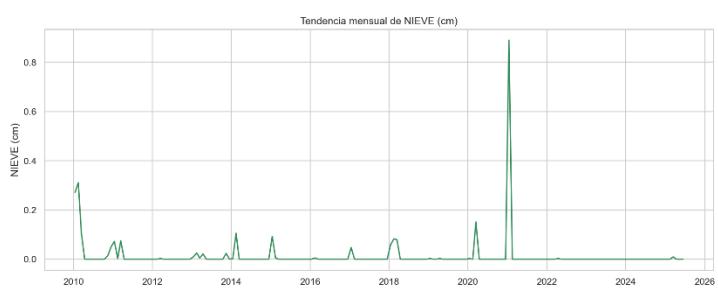


Figura 13 - Nieve - Filomena

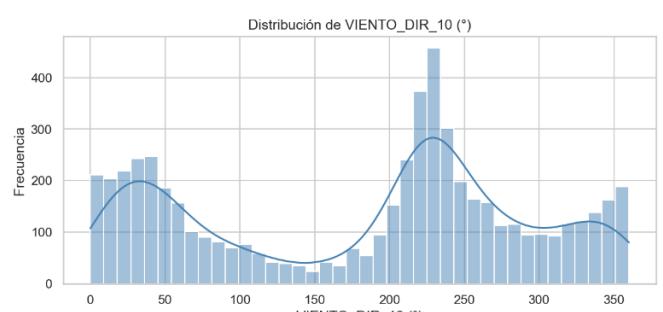


Figura 14 - Viento - Predominio viento suroeste (10m)

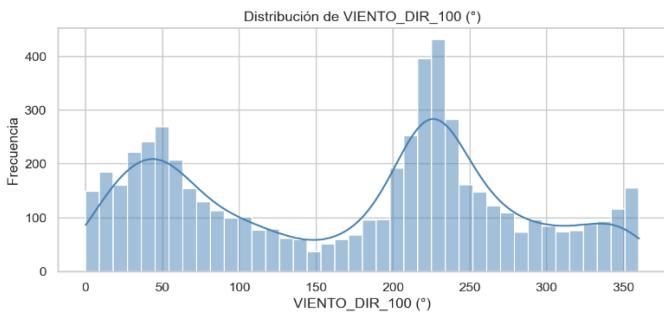


Figura 15 - Viento - Predominio viento suroeste (100m)

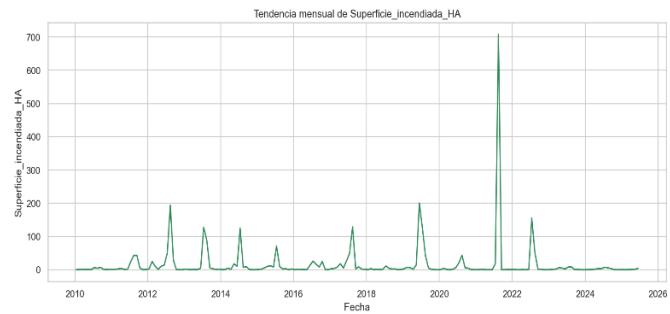


Figura 16 - Incendios Forestales - Ávila ago2021

En cuanto a la estacionalidad se observó que variables como la temperatura o la radiación solar siguen ciclos estacionales bien definidos, mientras que la velocidad del viento o las precipitaciones presentan una mayor variabilidad intranual. Asimismo, se observó que la ocurrencia de incendios forestales se concentra principalmente en los meses de verano, con una marcada estacionalidad y valores prácticamente nulos durante el resto del año.

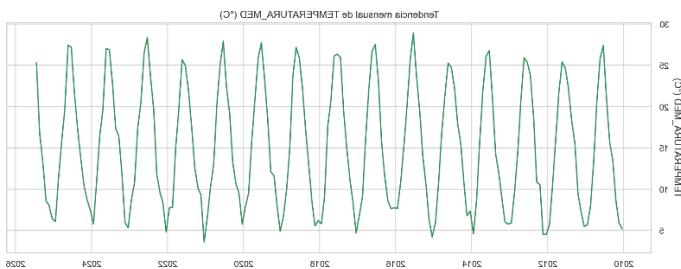


Figura 17 - Temperatura - Estacionalidad marcada

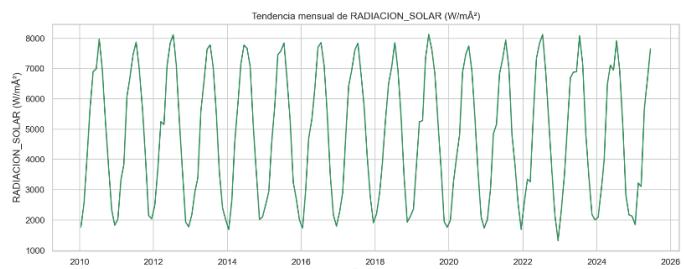


Figura 18 - Radiación Solar - Estacionalidad marcada

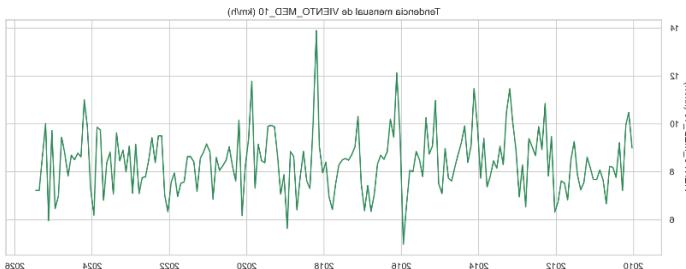


Figura 19 - Velocidad viento - variabilidad intranual

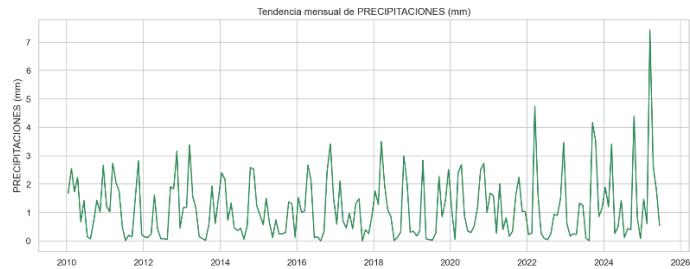


Figura 20 - Precipitaciones - variabilidad intranual

## ii. Variable tráfico - TOTAL\_HOR\_EXT

Este subapartado presenta el Análisis Exploratorio de Datos (EDA) de la variable "TOTAL\_HOR\_EXT", indicador que representa la densidad de tráfico registrada en las diferentes estaciones distribuidas a lo largo de la ciudad de Madrid.

A diferencia de las variables tratadas en el subapartado anterior, "TOTAL\_HOR\_EXT" presenta valores diarios que dependen de la zona geográfica asociada a cada observación (NORTE, CENTRO y SUR). Esta particularidad introduce una dimensión espacial adicional que resulta esencial para interpretar adecuadamente las variaciones de la intensidad del tráfico y su relación con el comportamiento urbano.

Desde un enfoque exploratorio, este análisis permite identificar patrones diferenciales entre zonas, detectar posibles anomalías y evaluar la consistencia temporal de los datos, lo que constituye un paso previo imprescindible antes de abordar cualquier modelización posterior.

Para cada zona geográfica se generaron los mismos gráficos que los del subapartado anterior. Tras la generación de estos gráficos, se observó que la estimación de los datos que se llevó a cabo por medio del modelo **Prophet** en el subapartado 5.2.iii. punto 4º de la *Tabla 10* (Imputación valores fuera de rango) no era la adecuada, puesto que como veremos en los gráficos temporales siguientes, los datos previos a 2019-2020, presentan un comportamiento extraño:

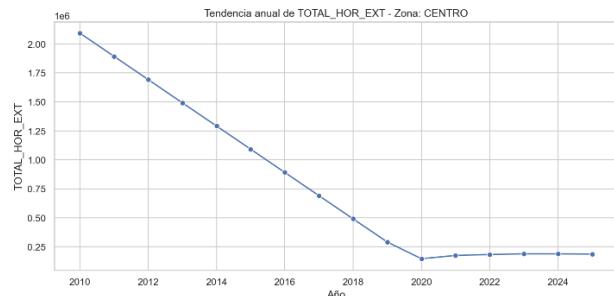


Figura 21 - Trafico - Tendencia anual - Zona Centro

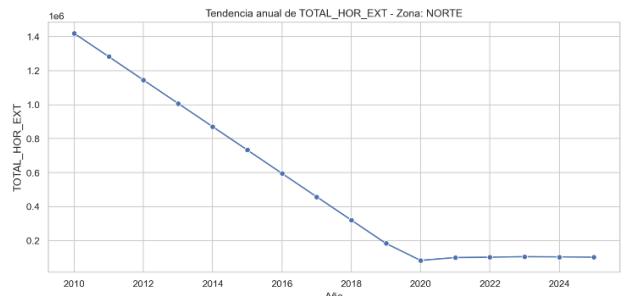


Figura 22 - Trafico - Tendencia anual - Zona Norte

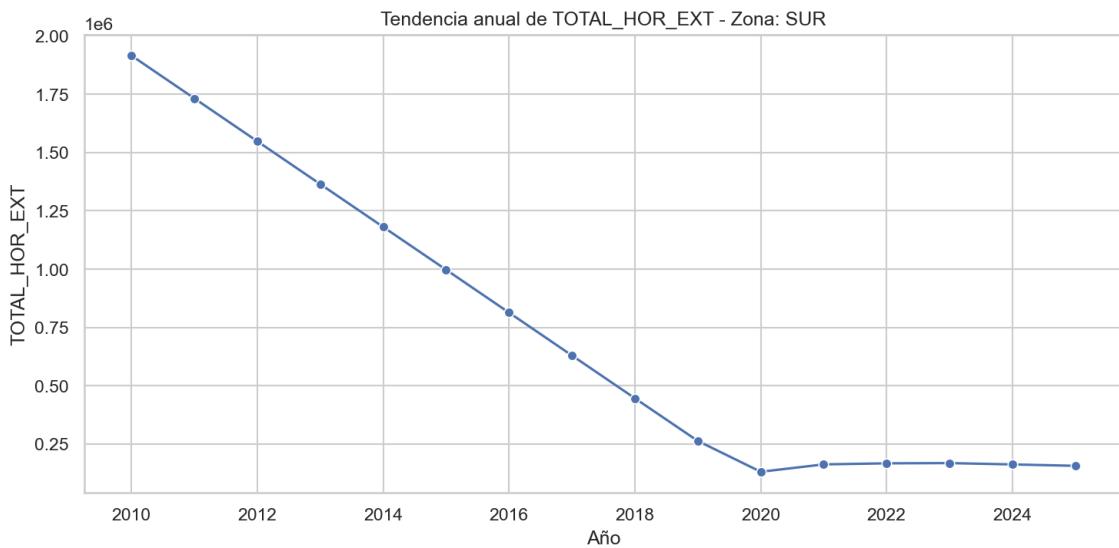


Figura 23 - Trafico - Tendencia anual - Zona Sur

No obstante, las gráficas temporales diarias sí muestran con claridad el efecto que tuvo la pandemia de la COVID-19 sobre la variable “TOTAL\_HOR\_EXT” en el primer y segundo trimestre del año 2020.

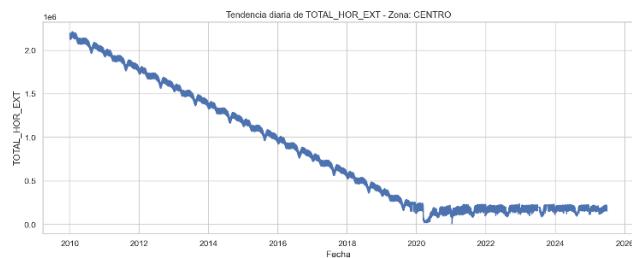


Figura 24 - Trafico - Tendencia diaria - Zona Centro

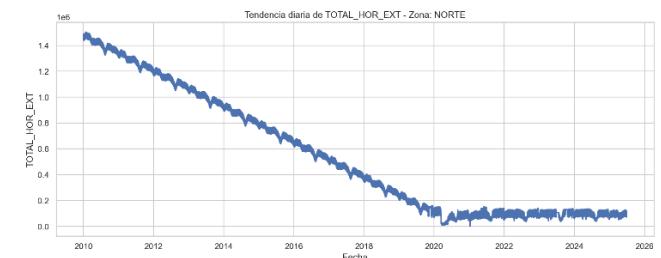


Figura 25 - Trafico - Tendencia diaria - Zona Norte

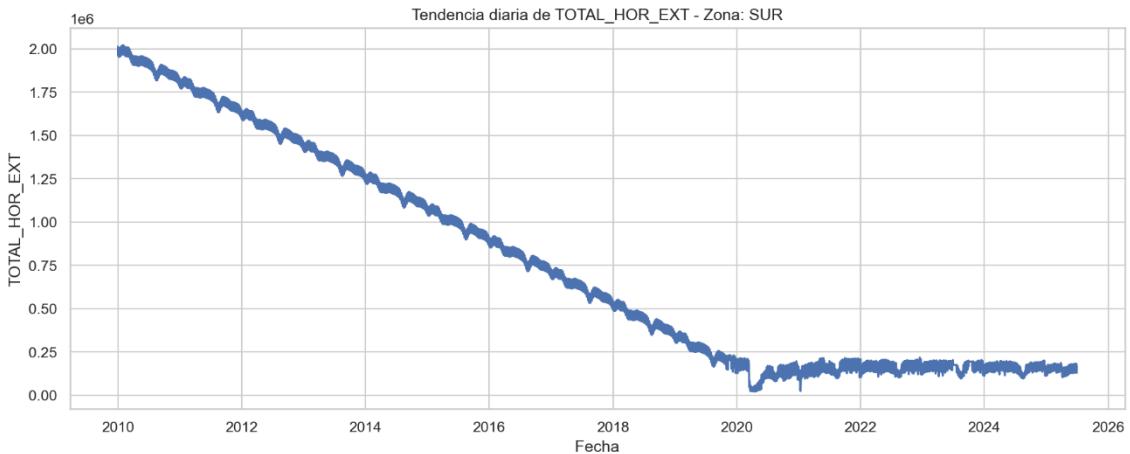


Figura 26 - Trafico - Tendencia diaria - Zona Sur

Todos estos gráficos se pueden localizar en la carpeta 02 figs variable tráfico contenida del Anexo A2.

**CONCLUSIÓN:** Los datos que las gráficas de evolución temporal evidencian una falta de coherencia en los datos correspondientes a los años previos a 2019-2020, lo que justifica la decisión de excluir esta variable de los análisis posteriores, al menos de manera, temporal.

## 6.2. EDA VARIABLE DEPENDIENTE “MEDICION\_UGM3”

En el conjunto de datos analizado, la variable **MEDICION\_ugm3** no constituye una única serie temporal, sino que presenta múltiples valores en un mismo día en función de dos dimensiones clave: la magnitud del contaminante (MAGNITUD) y la estación de medida (ZONA\_ESTACION). Cada combinación MAGNITUD × ZONA\_ESTACION genera una medición independiente, lo que se traduce en un total de **33 valores diarios diferentes**. Esta estructura multicomponente implica que el análisis exploratorio debe abordar estas **series de forma diferenciada**, dado que cada una refleja un comportamiento específico condicionado tanto por la tipología del contaminante como por la localización espacial de la estación.

Los 33 valores diarios (en adelante, “**Las Combinaciones**”) se originan al cruzar las 11 magnitudes finalmente conservadas (tras aplicar los criterios de depuración descritos en el subapartado 5.2.ii, punto 6º de la *Tabla 8*) con las tres zonas espaciales consideradas (NORTE, CENTRO y SUR). Dichas magnitudes son las siguientes:

MAGNITUD	CONTAMINANTE / SUSTANCIA
1	Dióxido de Azufre
6	Monóxido de Carbono
7	Monóxido de Nitrógeno
8	Dióxido de Nitrógeno
9	Partículas < 2,5 µm
10	Partículas < 10 µm
12	Óxidos de Nitrógeno
14	Ozono
20	Tolueno
30	Benceno
35	Etilbenceno

Tabla 14 - Magnitudes tras preprocesamiento

Señalar que todos los gráficos y resultados generados en este apartado, de todas “**Las Combinaciones**”, se pueden localizar en la carpeta **03 figs variable MEDICION ugm3 contenida del Anexo A2.**

Por último, indicar que todas las conclusiones acerca de los resultados obtenidos en este apartado se han basado en la siguiente tabla de criterios que, como se observó anteriormente (*Tabla 1* y *Tabla 2*), se han basado principalmente en la [Directiva 2008/50/CE](#), en el [Real Decreto 102/2011](#) y en la [Guía OMS](#).

MAGNITUD - NUM	MAGNITUD - NOM	NORMATIVA	SALUD HUMANA ( $\mu\text{g}/\text{m}^3$ )	VEGETACIÓN ( $\mu\text{g}/\text{m}^3$ )	UMBRAL ALERTA ( $\mu\text{g}/\text{m}^3$ )
10	Partículas < 10 $\mu\text{m}$	Directiva 2008/50/CE ANEXO XI	<b>50</b> (diario) <b>40</b> (anual)	-	-
	Partículas < 10 $\mu\text{m}$	Real Decreto 102/2011	<b>50</b> (diario) <b>40</b> (anual)	-	<b>80</b> (diario)
	Partículas < 10 $\mu\text{m}$	Guía OMS	<b>45</b> (diario) <b>15</b> (anual)	-	-
12	Óxidos de Nitrógeno	Directiva 2008/50/CE ANEXO XI	-	<b>30</b>	-
	Óxidos de Nitrógeno	Real Decreto 102/2011	-	<b>30</b>	-
	Óxidos de Nitrógeno	Guía OMS	-	-	-
14	Ozono	Directiva 2008/50/CE ANEXO VII	<b>120</b>	-	<b>240</b>
	Ozono	Real Decreto 102/2011	<b>120</b>	-	<b>240</b>
	Ozono	Guía OMS	<b>60</b>	-	-
1	Dióxido de Azufre	Directiva 2008/50/CE ANEXO XI	<b>125</b> (diario)	<b>20</b> (año civil e invierno - 01/10 - 31/03-)	<b>500</b>
	Dióxido de Azufre	Real Decreto 102/2011	<b>125</b> (diario)	<b>20</b> (año civil e invierno - 01/10 - 31/03-)	-
	Dióxido de Azufre	Guía OMS	<b>40</b> (diario)	-	-
20	Tolueno	<a href="#">Murcia Salud</a>	<b>188000</b>	-	-
30	Benceno	Directiva 2008/50/CE ANEXO XI	<b>5</b>	-	-
	Benceno	Real Decreto 102/2011	<b>5</b>	-	-
	Benceno	Guía OMS	-	-	-
35	Etilbenceno	<a href="#">Recomendación ATSDR</a>	<b>434000</b>	-	-
6	Monóxido de Carbono	Directiva 2008/50/CE ANEXO XI	<b>10000</b> (media móvil 8h)	-	-
	Monóxido de Carbono	Real Decreto 102/2011	<b>10000</b>	-	-
	Monóxido de Carbono	Guía OMS	<b>4000</b>	-	-
7	Monóxido de Nitrógeno	Directiva 2008/50/CE ANEXO XI	-	<b>30</b>	-
	Monóxido de Nitrógeno	Real Decreto 102/2011	-	<b>30</b>	-
	Monóxido de Nitrógeno	Guía OMS	-	-	-
8	Dióxido de Nitrógeno	Directiva 2008/50/CE ANEXO XI	<b>40</b> (año civil)	-	<b>400</b>

	Dióxido de Nitrógeno	Real Decreto 102/2011	<b>40</b> (año civil)	-	<b>400</b>
	Dióxido de Nitrógeno	Guía OMS	<b>25</b> (diario) <b>10</b> (anual)	-	-
9	Partículas < 2,5 µm	Directiva 2008/50/CE ANEXO XIV	<b>20</b> (IEM 2015) ~ <b>10,35</b> (IEM 2020) <b>25</b> (a partir de 2015) <b>20</b> (a partir de 2020)	-	-
	Partículas < 2,5 µm	Real Decreto 102/2011	<b>20</b> (IEM 2015) ~ <b>10,35</b> (IEM 2020) <b>25</b> (a partir de 2015) <b>20</b> (a partir de 2020)	-	<b>50</b>
	Partículas < 2,5 µm	Guía OMS	<b>15</b> (diario) <b>5</b> (anual)	-	-

Tabla 15 - Criterios - Valores límite contaminantes

### i. Resultados estadísticos de “Las Combinaciones”

En este subapartado se presentan los resultados estadísticos correspondientes a cada una de “**Las Combinaciones**” analizadas. El objetivo es caracterizar su comportamiento mediante indicadores descriptivos que permitan comparar niveles, variabilidad y distribución entre ellas, identificando patrones relevantes para el análisis posterior. Este estudio ofrece una primera evaluación cuantitativa de las diferencias entre combinaciones y proporciona una base sólida para extraer conclusiones iniciales que faciliten la interpretación de las etapas posteriores del EDA.

Los estadísticos de cada combinación se resumen en la siguiente tabla (todos los valores expresados en  $\mu\text{g}/\text{m}^3$ ):

ID COMB	ZONA	MAGNITUD	MEDIA	DESV. TIP.	MÍN	MÁX	MODA	P25	P75
1	CENTRO	10	19,74	12,50	2,67	299,33	17,33	11,67	24,67
2	NORTE	10	20,64	13,07	1,67	381,67	18,33	12,67	25,67
3	SUR	10	20,23	13,00	3,00	373,00	18	12	25,49
4	CENTRO	12	55,70	45,16	6,00	381,67	41,83	27	66,5
5	NORTE	12	62,28	48,68	6,28	434,69	48,60	30,44	75,75
6	SUR	12	61,39	57,05	5,83	418,50	40,37	25,83	73,17
7	CENTRO	14	50,02	23,05	1,83	114,00	51,67	32,67	67,67
8	NORTE	14	53,00	23,59	2,75	118,50	55,23	34,75	71,52
9	SUR	14	48,84	23,48	1,75	109,25	51	30,75	67
10	CENTRO	1	5,77	3,04	0,00	23,50	5	3,67	7,33
11	NORTE	1	5,04	3,51	1,00	24,00	4	3	6
12	SUR	1	5,40	3,24	1,00	19,75	4,59	3	7,5
13	CENTRO	20	2,34	1,78	0,10	18,30	1,85	1,1	3,05
14	NORTE	20	2,63	2,12	0,12	18,90	2,02	1,2	3,4
15	SUR	20	1,94	1,81	0,10	18,60	1,4	0,8	2,4

16	CENTRO	30	0,52	0,38	0,10	3,10	0,4	0,25	0,65
17	NORTE	30	0,62	0,48	0,10	4,40	0,49	0,3	0,8
18	SUR	30	0,38	0,37	0,00	4,70	0,22	0,2	0,4
19	CENTRO	35	0,59	0,60	0,01	5,70	0,35	0,2	0,8
20	NORTE	35	0,57	0,64	0,00	8,62	0,4	0,2	0,72
21	SUR	35	0,38	0,36	0,00	5,00	0,3	0,1	0,5
22	CENTRO	6	329,31	118,38	100	1050	300	250	383,33
23	NORTE	6	355,97	171,54	100	1500	300	250	400
24	SUR	6	356,80	143,24	100	1150	300	250	400
25	CENTRO	7	14,69	20,04	0,26	188,00	7,5	3,67	15,65
26	NORTE	7	18,08	22,07	0,26	219,64	10,72	5,14	20,67
27	SUR	7	17,68	26,81	0,26	210,83	6,67	3,17	18,83
28	CENTRO	8	33,20	16,39	4,83	108,00	30,17	20,83	42,33
29	NORTE	8	34,60	16,66	4,36	118,72	31,85	22,05	44,2
30	SUR	8	34,28	17,94	4,50	120,17	30,33	20,68	44,33
31	CENTRO	9	10,12	5,41	2,00	65,67	9	6,33	12,67
32	NORTE	9	10,42	5,79	1,00	72,00	9,5	6,5	13
33	SUR	9	10,56	6,12	1,00	78,00	9	6	13,79

Tabla 16 - Resultados estadísticos combinaciones MEDICION\_ugm3

Se han sombreado en **naranja** aquellos valores que superan algunos de los límites señalados en la *Tabla 15*.

#### CONCLUSIONES:

- **PM10 y PM2.5** (magnitudes 10 y 9 respectivamente) **presentan los niveles más problemáticos**, superando de forma reiterada los valores recomendados por la OMS e incluso alcanzando episodios que rebasan umbrales de alerta, como puede apreciarse en los valores máximos referidos a estas magnitudes.
- El Dióxido de Nitrógeno (**NO<sub>2</sub>**) presenta **niveles medios elevados y máximos que superan los valores recomendados por la OMS para la protección de la salud humana**, reflejando su fuerte vinculación a las emisiones urbanas y especialmente al tráfico rodado.
- El Monóxido de Nitrógeno (**NO**) y los Óxidos de Nitrógeno totales (**NOx**) muestran una variabilidad muy alta y episodios máximos que superan los valores límite para la protección de la vegetación.
- Para **O<sub>3</sub>** (ozono), aunque los valores medios se mantienen en rangos saludables, los máximos y el percentil 75, alcanzan niveles que superan las recomendaciones de la OMS, reflejando un **comportamiento típico de contaminante secundario<sup>2</sup> con picos estacionales**.
- El **SO<sub>2</sub>** y el **CO** mantienen concentraciones bajas y estables en todas las zonas, situándose en rangos saludables y sin episodios de superación relevantes; únicamente el SO<sub>2</sub> presenta algunos picos puntuales cercanos al valor límite para vegetación.
- Los contaminantes de tipo orgánico volátil (**Tolueno, Benceno y Etilbenceno**) presentan niveles medios y máximos bajos en todas las zonas, situándose en rangos saludables y sin episodios relevantes.
- En general, se observa un patrón claro de **variabilidad alta** para los contaminantes más críticos (**PM10 y PM2.5**), indicando una influencia

<sup>2</sup> Se considera un contaminante secundario ya que no se emite directamente: se forman en la atmósfera por reacciones químicas

significativa de las condiciones meteorológicas y de los picos de emisiones urbanos.

## ii. Distribuciones y patrones estadísticos y temporales

En este subapartado se examinan los patrones de distribución y la variabilidad temporal de “**Las Combinaciones**” mediante la revisión de histogramas y de los boxplots mensuales y semanales. Los histogramas permiten identificar la forma de la distribución de cada serie, evidenciando su grado de asimetría, la presencia de colas largas asociadas a episodios de alta concentración y la posible existencia de valores atípicos o extremos. Además, facilitan la comparación visual entre magnitudes y zonas, mostrando cómo ciertos contaminantes presentan una mayor dispersión o una acumulación más frecuente en niveles bajos. Por su parte, los boxplots permiten analizar cómo evolucionan “**Las Combinaciones**” a lo largo del año y de la semana, poniendo de manifiesto las diferencias en estacionalidad.

A continuación, se presentan los gráficos correspondientes a los contaminantes que, a juicio del autor y a la luz de los resultados anteriores, resultan más relevantes y potencialmente perjudiciales para la salud humana: **PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub> y O<sub>3</sub>**.

### PM<sub>10</sub> - Distribuciones:

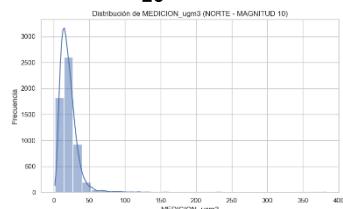


Figura 27 - PM<sub>10</sub> - Distribución NORTE

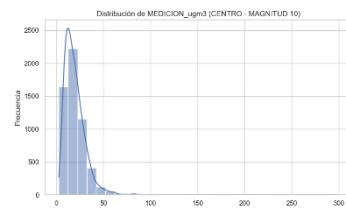


Figura 28 - PM<sub>10</sub> - Distribución CENTRO

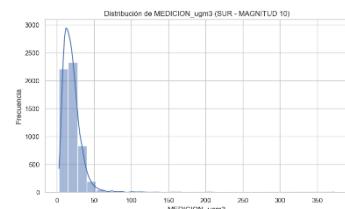


Figura 29 - PM<sub>10</sub> - Distribución SUR

### PM<sub>10</sub> - Boxplot - Mensual:

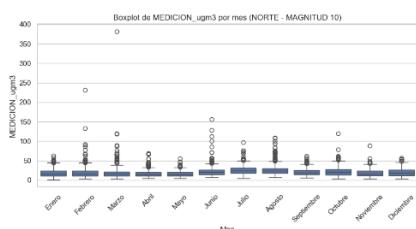


Figura 30 - PM<sub>10</sub> - Boxplot mensual NORTE

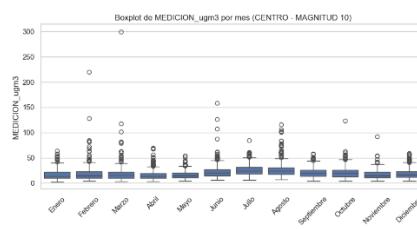


Figura 31 - PM<sub>10</sub> - Boxplot mensual CENTRO

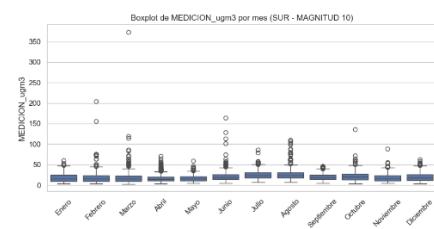


Figura 32 - PM<sub>10</sub> - Boxplot mensual SUR

## PM<sub>10</sub> - Boxplot - Semanal:

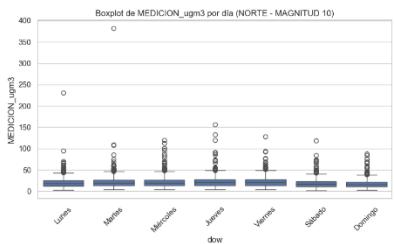


Figura 33 - PM<sub>10</sub> - Boxplot semanal NORTE

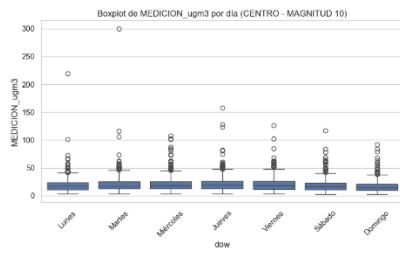


Figura 34 - PM<sub>10</sub> - Boxplot semanal CENTRO

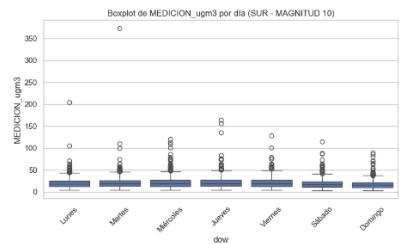


Figura 35 - PM<sub>10</sub> - Boxplot semanal SUR

## PM<sub>2.5</sub> - Distribuciones:

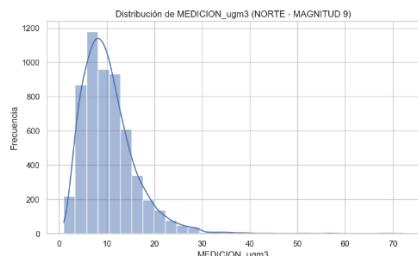


Figura 36 - PM<sub>2.5</sub> - Distribución NORTE

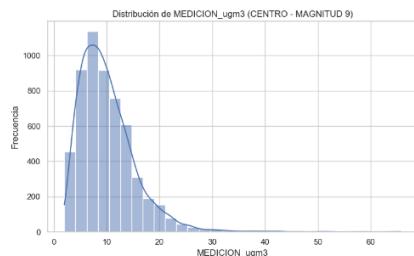


Figura 37 - PM<sub>2.5</sub> - Distribución CENTRO

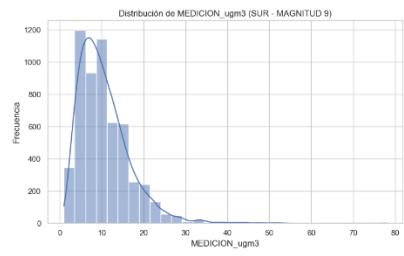


Figura 38 - PM<sub>2.5</sub> - Distribución SUR

## PM<sub>2.5</sub> - Boxplot - Mensual:

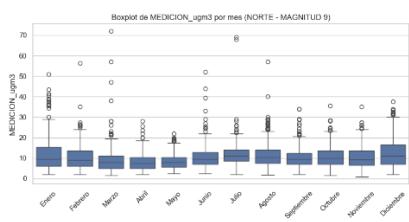


Figura 39 - PM<sub>2.5</sub> - Boxplot mensual NORTE

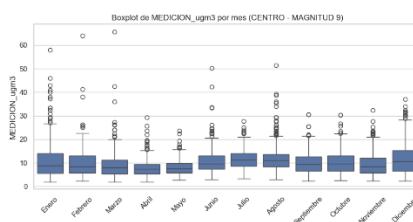


Figura 40 - PM<sub>2.5</sub> - Boxplot mensual CENTRO

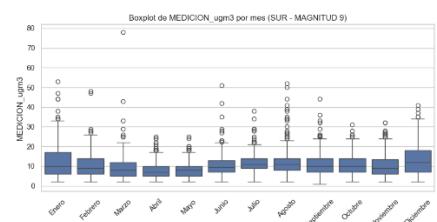


Figura 41 - PM<sub>2.5</sub> - Boxplot mensual SUR

## PM<sub>2.5</sub> - Boxplot - Semanal:

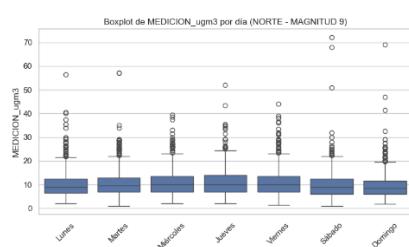


Figura 42 - PM<sub>2.5</sub> - Boxplot semanal NORTE

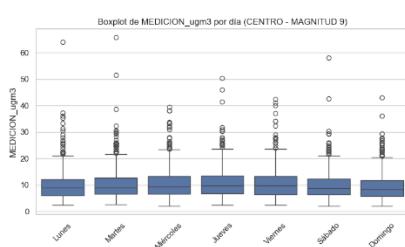


Figura 43 - PM<sub>2.5</sub> - Boxplot semanal CENTRO

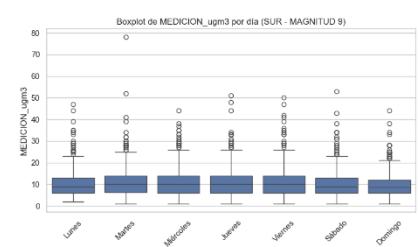


Figura 44 - PM<sub>2.5</sub> - Boxplot semanal SUR

## NO<sub>2</sub> - Distribuciones:

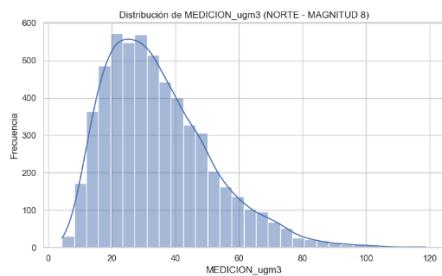


Figura 45 - NO<sub>2</sub> - Distribución NORTE

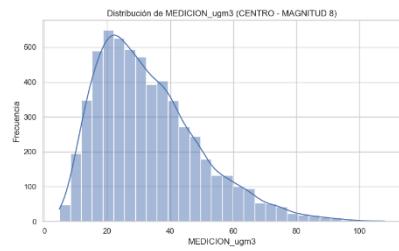


Figura 46 - NO<sub>2</sub> - Distribución CENTRO

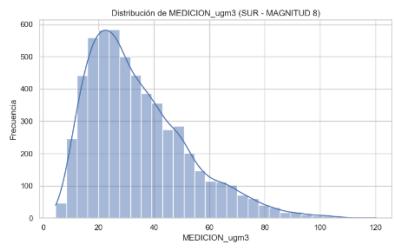


Figura 47 - NO<sub>2</sub> - Distribución SUR

## NO<sub>2</sub> - Boxplot - Mensual:

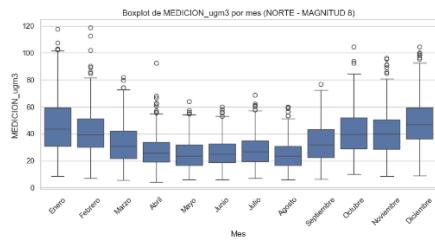


Figura 48 - NO<sub>2</sub> - Boxplot mensual NORTE

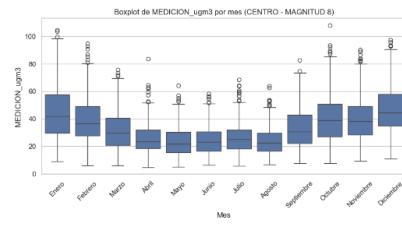


Figura 49 - NO<sub>2</sub> - Boxplot mensual CENTRO

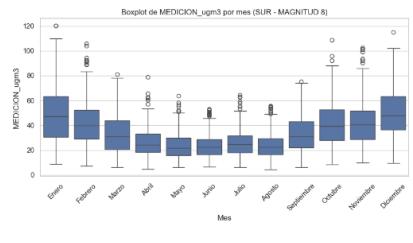


Figura 50 - NO<sub>2</sub> - Boxplot mensual SUR

## NO<sub>2</sub> - Boxplot - Semanal:

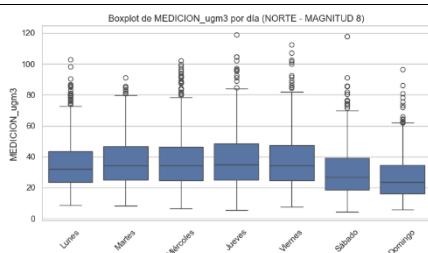


Figura 51 - NO<sub>2</sub> - Boxplot semanal NORTE

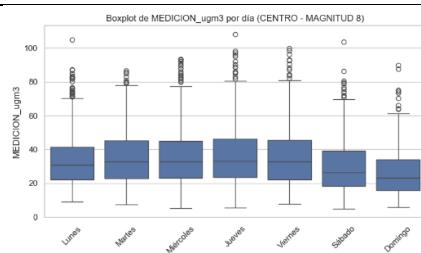


Figura 52 - NO<sub>2</sub> - Boxplot semanal CENTRO

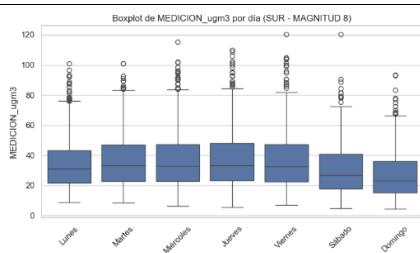


Figura 53 - NO<sub>2</sub> - Boxplot semanal SUR

### O<sub>3</sub> - Distribuciones:

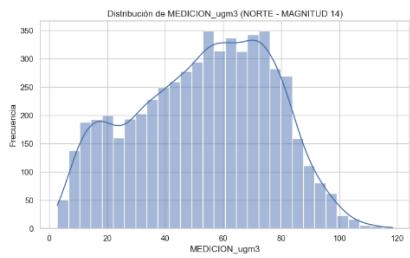


Figura 54 - O<sub>3</sub> - Distribución NORTE

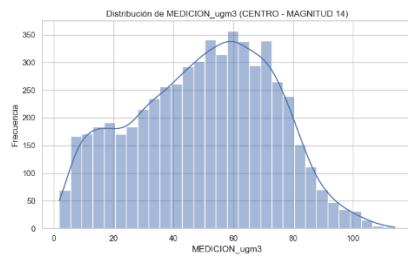


Figura 55 - O<sub>3</sub> - Distribución CENTRO

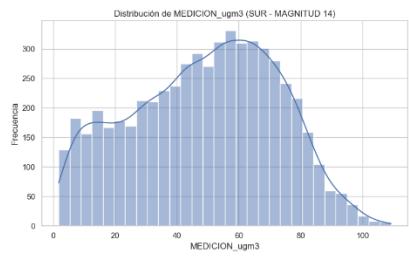


Figura 56 - O<sub>3</sub> - Distribución SUR

### O<sub>3</sub> - Boxplot - Mensual:

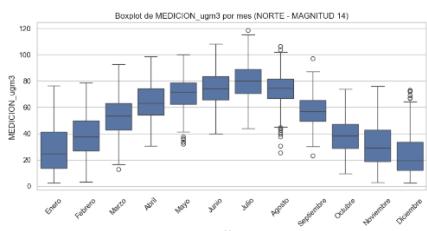


Figura 57 - O<sub>3</sub> - Boxplot mensual NORTE

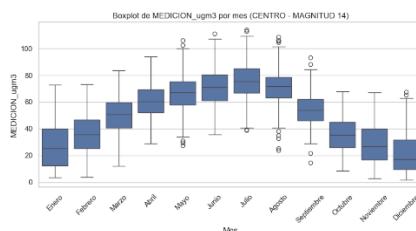


Figura 58 - O<sub>3</sub> - Boxplot mensual CENTRO

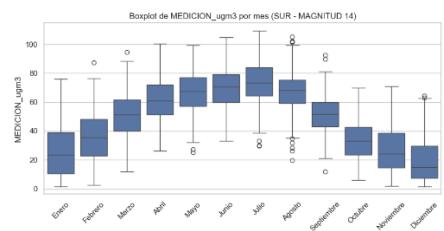


Figura 59 - O<sub>3</sub> - Boxplot mensual SUR

### O<sub>3</sub> - Boxplot - Semanal:

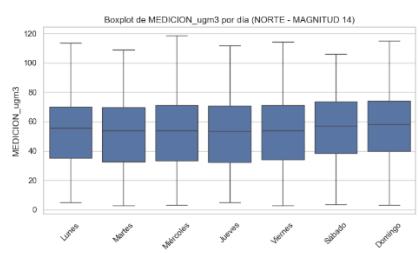


Figura 60 - O<sub>3</sub> - Boxplot semanal NORTE

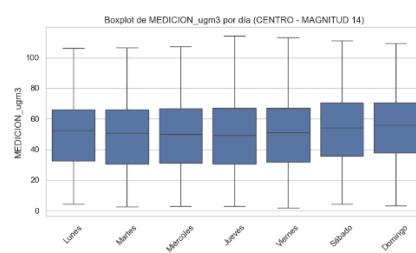


Figura 61 - O<sub>3</sub> - Boxplot semanal CENTRO

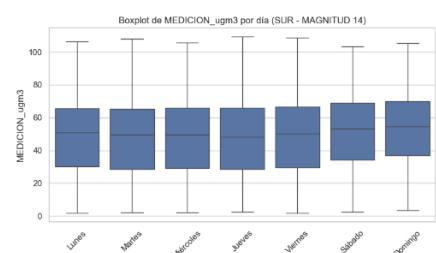


Figura 62 - O<sub>3</sub> - Boxplot semanal SUR

### Comentarios a las gráficas anteriores:

ID COMB	ZONA	MAGNITD	DISTRIBUCIÓN	BOXPLOT - MENSUAL	BOXPLOT - DIARIO
1	CENTRO	Partículas < 10 $\mu\text{m}$	Asimétrica Pos	Verano alto, resto bajo	Ausencia de patrón semanal
2	NORTE	Partículas < 10 $\mu\text{m}$	Asimétrica Pos	Verano alto, resto bajo	Ligera disminución fin de semana

3	SUR	Partículas < 10 µm	Asimétrica Pos	Verano alto, resto bajo	Ligera disminución fin de semana
7	CENTRO	Ozono	Asimétrica Pos	Verano alto, invierno bajo	Aumento fin de semana
8	NORTE	Ozono	Asimétrica Pos	Verano alto, invierno bajo	Ligero aumento fin de semana
9	SUR	Ozono	Asimétrica Pos	Verano alto, invierno bajo	Ligero aumento fin de semana
28	CENTRO	Dióxido de Nitrógeno	Asimétrica Pos	Invierno alto, verano bajo	Disminución domingo
29	NORTE	Dióxido de Nitrógeno	Asimétrica Pos	Invierno alto, verano bajo	Disminución domingo
30	SUR	Dióxido de Nitrógeno	Asimétrica Pos	Invierno alto, verano bajo	Disminución domingo
31	CENTRO	Partículas < 2,5 µm	Asimétrica Pos	Inviernos y veranos altos, resto bajos	Ligera disminución fin de semana
32	NORTE	Partículas < 2,5 µm	Asimétrica Pos	Inviernos y veranos altos, resto bajos	Ligera disminución fin de semana
33	SUR	Partículas < 2,5 µm	Asimétrica Pos	Inviernos y veranos altos, resto bajos	Ligera disminución fin de semana

Tabla 17 - Comentarios resultados estacionalidad

## CONCLUSIONES:

- **PM<sub>10</sub> y PM<sub>2.5</sub>** muestran distribuciones asimétricas positivas y un comportamiento estacional marcado, con valores más altos en verano en el caso de PM<sub>10</sub> y picos tanto en invierno como en verano para PM<sub>2.5</sub>. A nivel semanal, ambas fracciones presentan una ligera disminución durante el fin de semana, coherente con la reducción del tráfico.
- El **NO<sub>2</sub>** presenta concentraciones significativamente más elevadas en invierno, con mínimos en verano, lo que hace presagiar que tenga una relación directa con las horas solares. Su patrón semanal revela una disminución clara los domingos, reflejando la caída de la actividad urbana (tráfico).
- El **O<sub>3</sub>** muestra el patrón inverso al **NO<sub>2</sub>**: alcanza sus máximos en verano y presenta niveles bajos en invierno. En el análisis semanal se observa un ligero aumento durante el fin de semana, un comportamiento asociado a la menor emisión de **NO** procedente del tráfico urbano. Al reducirse este contaminante primario, disminuye también el proceso de destrucción del ozono, permitiendo

que sus concentraciones sean ligeramente superiores respecto a los días laborables.

- En general, todos los contaminantes mantienen distribuciones asimétricas positivas, con predominancia de valores bajos y presencia de episodios puntuales que generan colas largas.
- Parece haber una correlación inversa entre el contaminante NO<sub>2</sub> y el O<sub>3</sub>, que se aprecia, claramente en los gráficos boxplots semanales y mensuales.

### iii. Tendencias

En este subapartado se analiza la evolución temporal de todas “**Las Combinaciones**” con el fin de identificar sus tendencias generales y estimar los niveles medios previstos para final de año 2025. El estudio de estas trayectorias permite evaluar si los contaminantes presentan patrones de descenso, estabilización o incremento a lo largo del periodo analizado, así como determinar su situación actual en relación con los valores objetivo de calidad del aire.

Al igual que en el subapartado anterior, se presentan los gráficos temporales correspondientes a los contaminantes que resultan más relevantes y potencialmente perjudiciales para la salud humana.

#### PM<sub>10</sub> - NORTE

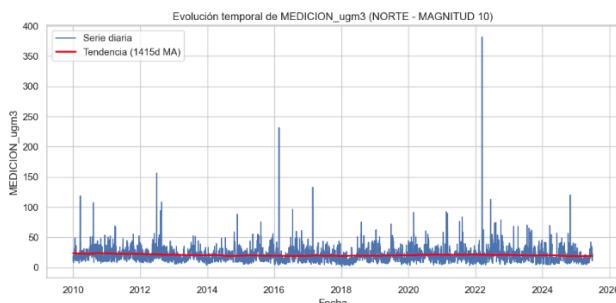


Figura 63 - PM<sub>10</sub> - Tendencia diaria NORTE

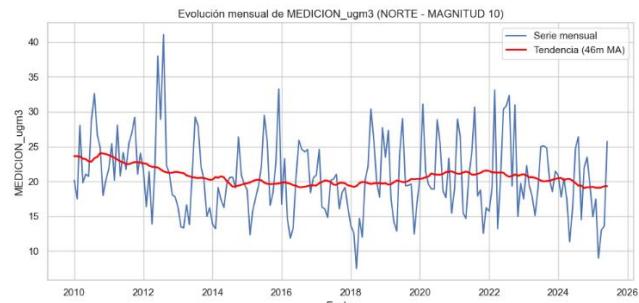


Figura 64 - PM<sub>10</sub> - Tendencia mensual NORTE

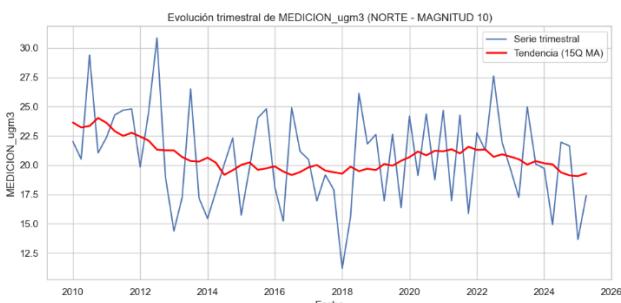


Figura 65 - PM<sub>10</sub> - Tendencia trimestral NORTE

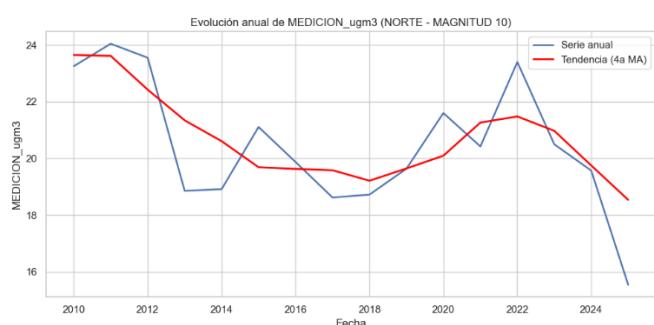


Figura 66 - PM<sub>10</sub> - Tendencia anual NORTE

## PM<sub>10</sub> - CENTRO

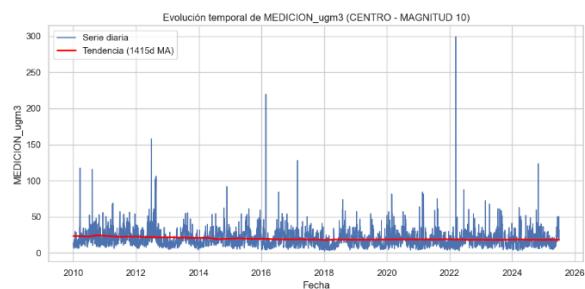


Figura 67 - PM<sub>10</sub> - Tendencia diaria CENTRO

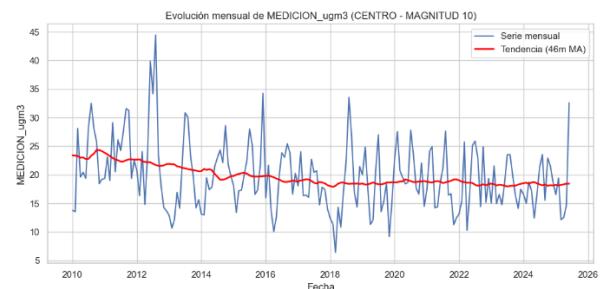


Figura 68 - PM<sub>10</sub> - Tendencia mensual CENTRO

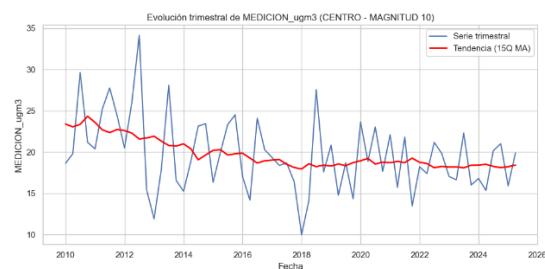


Figura 69 - PM<sub>10</sub> - Tendencia trimestral CENTRO

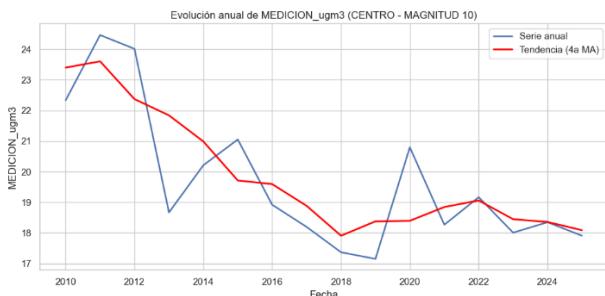


Figura 70 - PM<sub>10</sub> - Tendencia anual CENTRO

## PM<sub>10</sub> - SUR

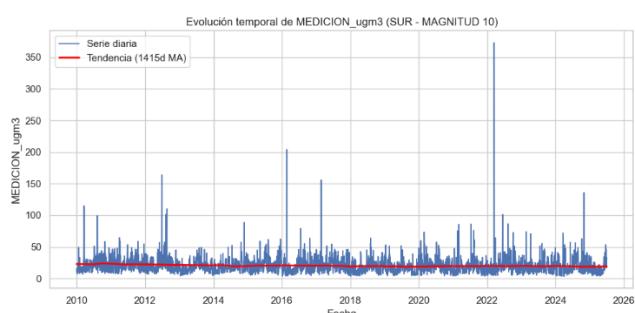


Figura 71 - PM<sub>10</sub> - Tendencia diaria SUR

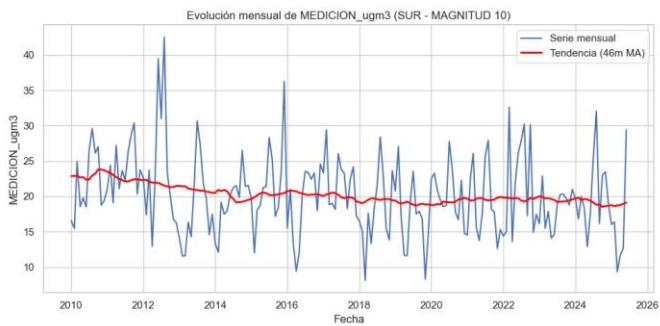


Figura 72 - PM<sub>10</sub> - Tendencia mensual SUR

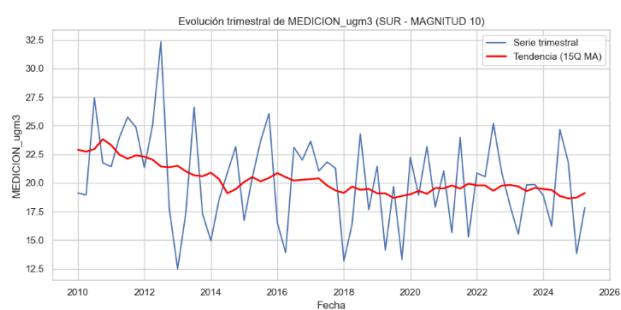


Figura 73 - PM<sub>10</sub> - Tendencia trimestral SUR

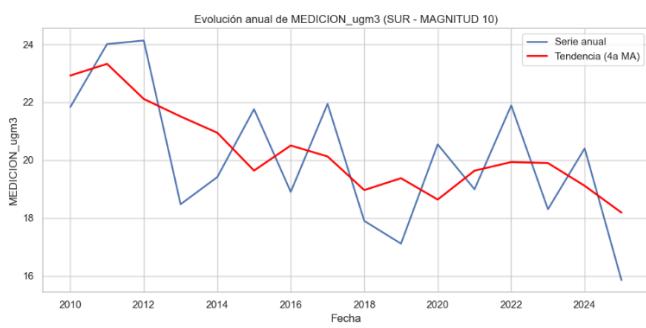


Figura 74 - PM<sub>10</sub> - Tendencia anual SUR

## PM<sub>2.5</sub> - NORTE

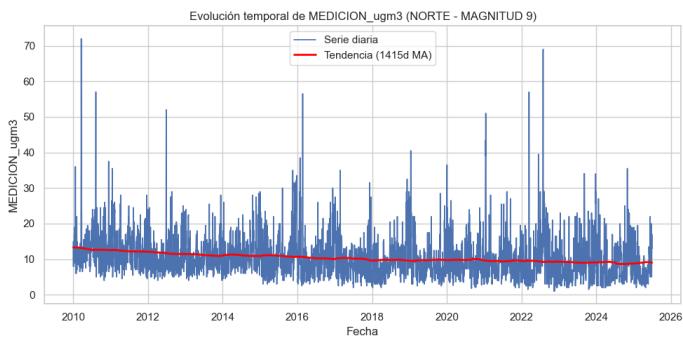


Figura 75 - PM<sub>2.5</sub> - Tendencia diaria NORTE

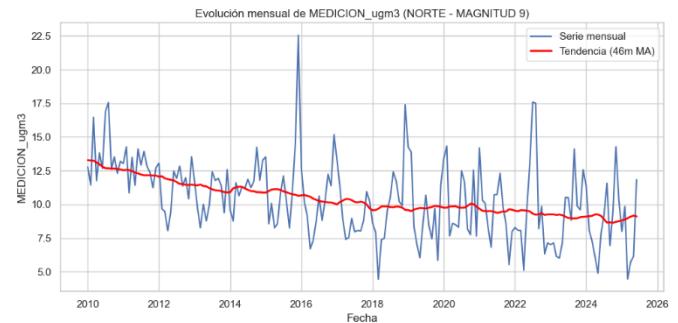


Figura 76 - PM<sub>2.5</sub> - Tendencia mensual NORTE

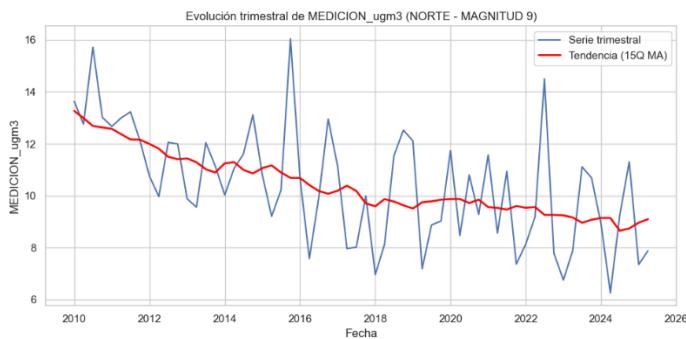


Figura 77 - PM<sub>2.5</sub> - Tendencia trimestral NORTE

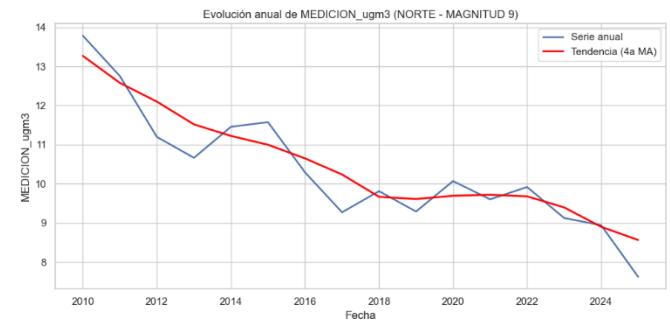


Figura 78 - PM<sub>2.5</sub> - Tendencia anual NORTE

## PM<sub>2.5</sub> - CENTRO

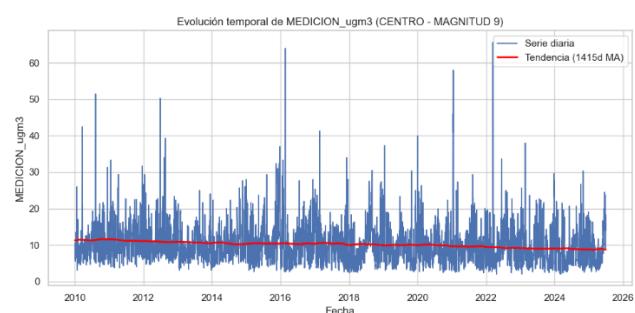


Figura 79 - PM<sub>2.5</sub> - Tendencia diaria CENTRO

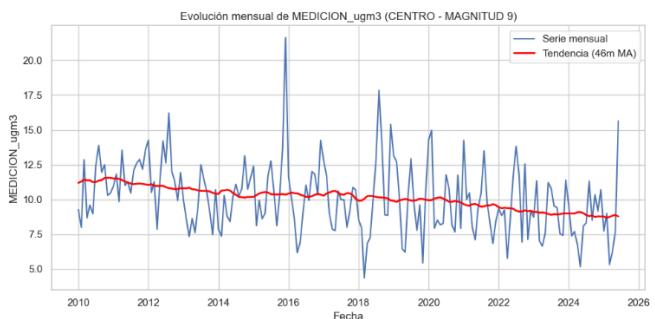


Figura 80 - PM<sub>2.5</sub> - Tendencia mensual CENTRO

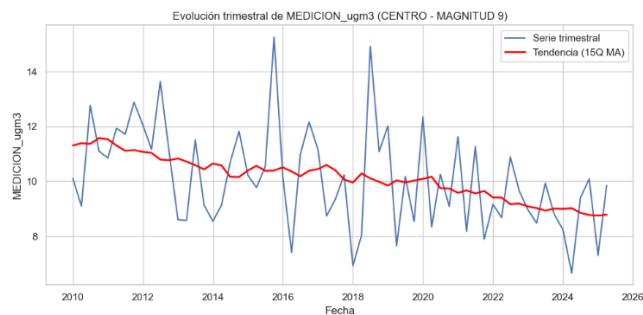


Figura 81 - PM<sub>2.5</sub> - Tendencia trimestral CENTRO

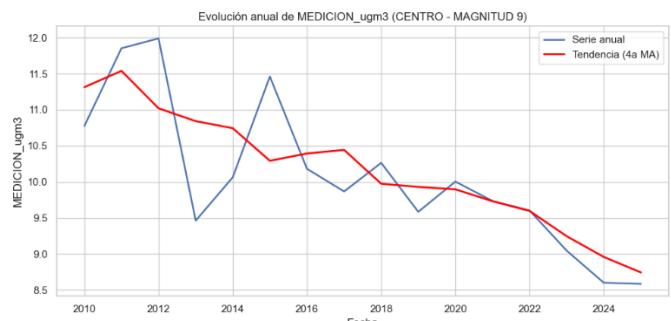


Figura 82 - PM<sub>2.5</sub> - Tendencia anual CENTRO

## PM<sub>2.5</sub> - SUR

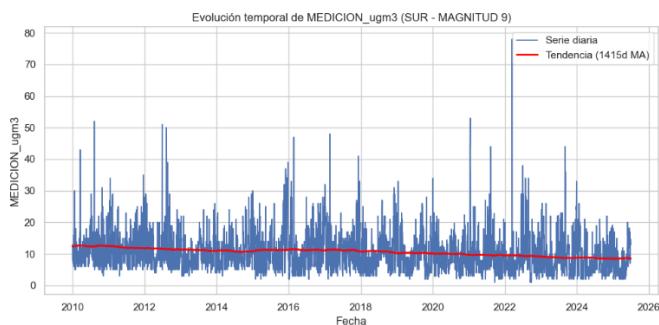


Figura 83 - PM<sub>2.5</sub> - Tendencia diaria SUR

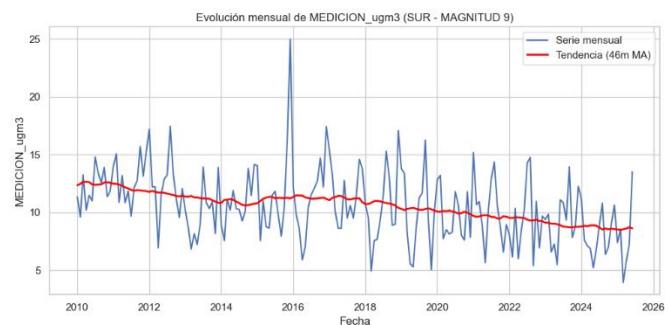


Figura 84 - PM<sub>2.5</sub> - Tendencia mensual SUR

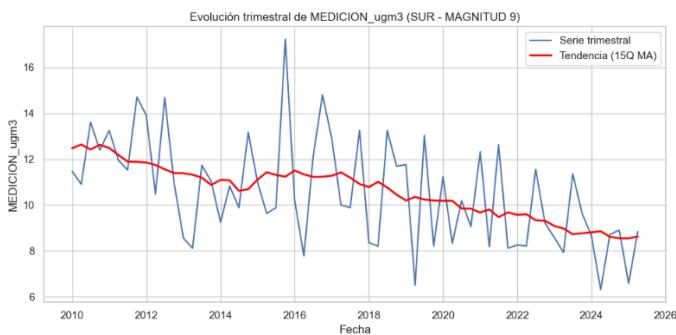


Figura 85 - PM<sub>2.5</sub> - Tendencia trimestral SUR

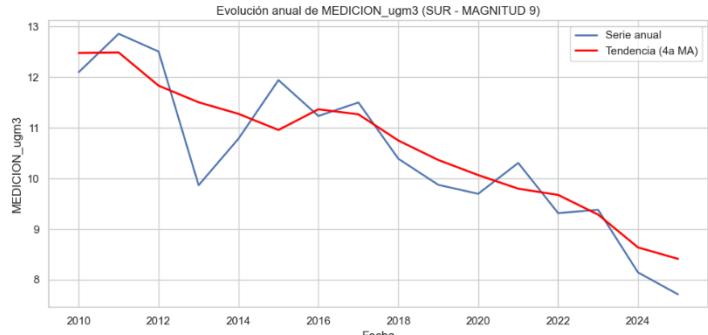


Figura 86 - PM<sub>2.5</sub> - Tendencia anual SUR

## NO<sub>2</sub> - NORTE

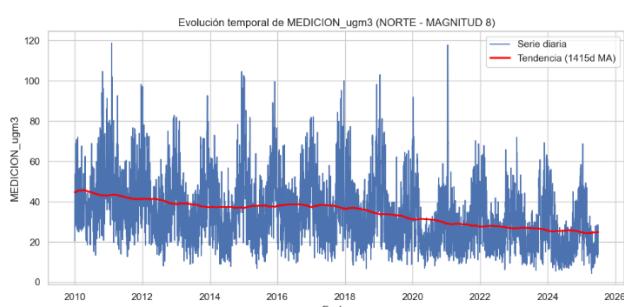


Figura 87 - NO<sub>2</sub> - Tendencia diaria NORTE

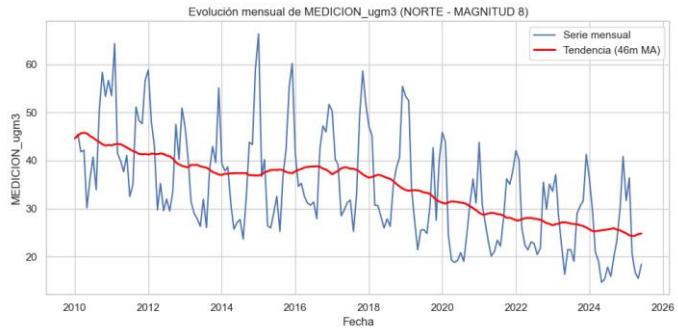


Figura 88 - NO<sub>2</sub> - Tendencia mensual NORTE

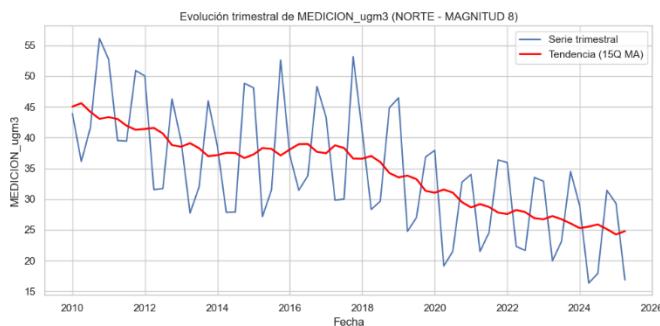


Figura 89 - NO<sub>2</sub> - Tendencia trimestral NORTE

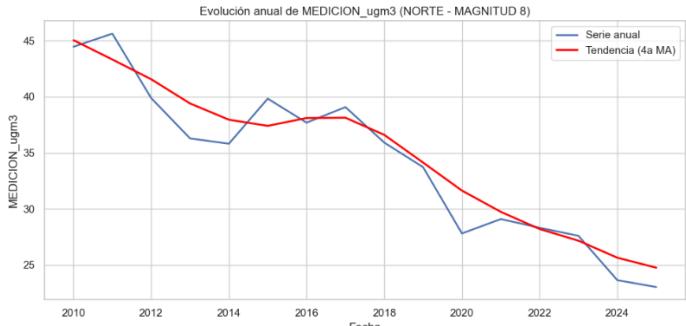


Figura 90 - NO<sub>2</sub> - Tendencia anual NORTE

## NO<sub>2</sub> - CENTRO

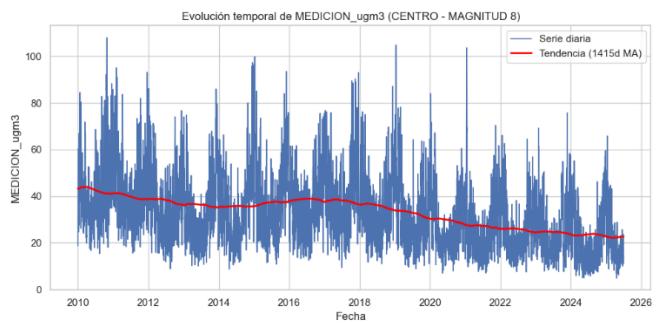


Figura 91 - NO<sub>2</sub> - Tendencia diaria CENTRO

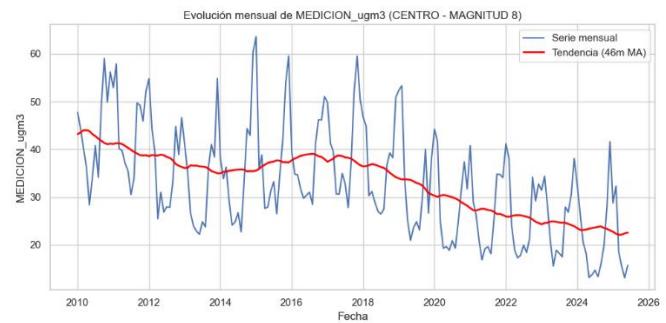


Figura 92 - NO<sub>2</sub> - Tendencia mensual CENTRO

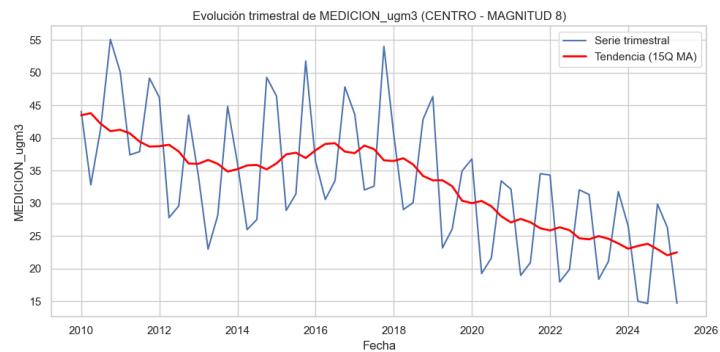


Figura 93 - NO<sub>2</sub> - Tendencia trimestral CENTRO

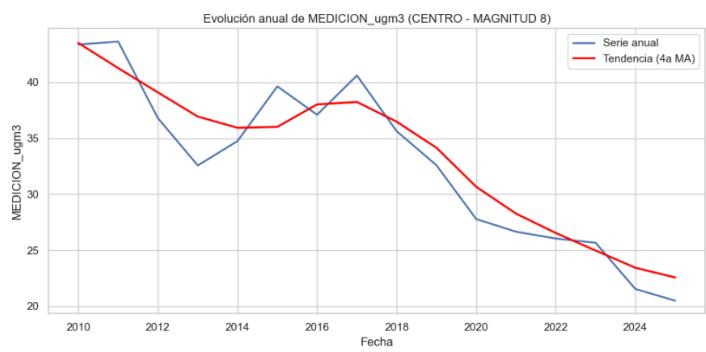


Figura 94 - NO<sub>2</sub> - Tendencia anual CENTRO

## NO<sub>2</sub> - SUR

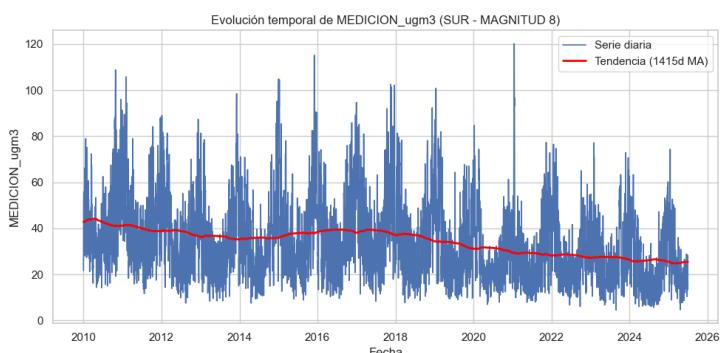


Figura 95 - NO<sub>2</sub> - Tendencia diaria SUR

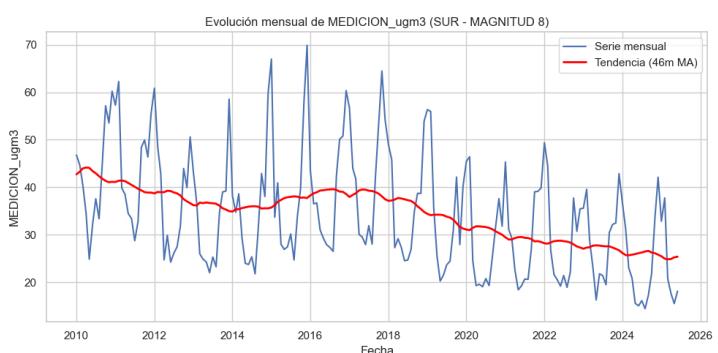


Figura 96 - NO<sub>2</sub> - Tendencia mensual SUR

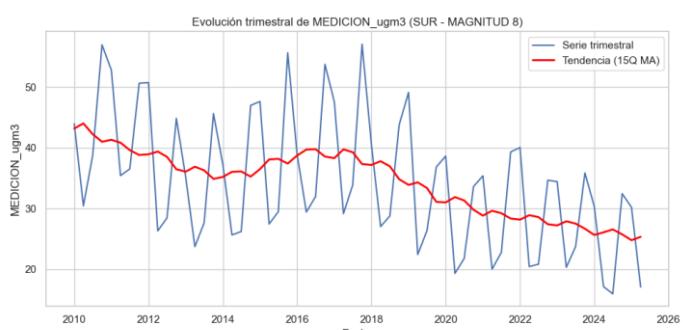


Figura 97 - NO<sub>2</sub> - Tendencia trimestral SUR

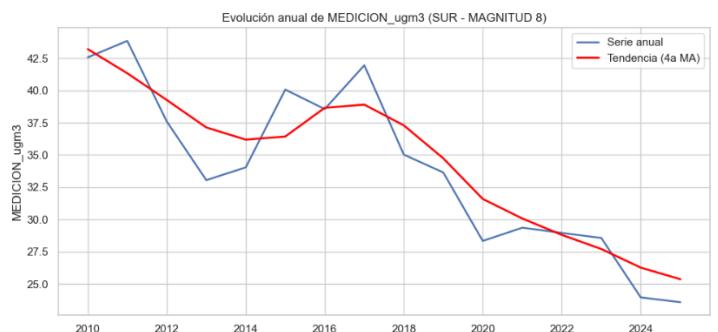


Figura 98 - NO<sub>2</sub> - Tendencia anual SUR

## O<sub>3</sub> - NORTE

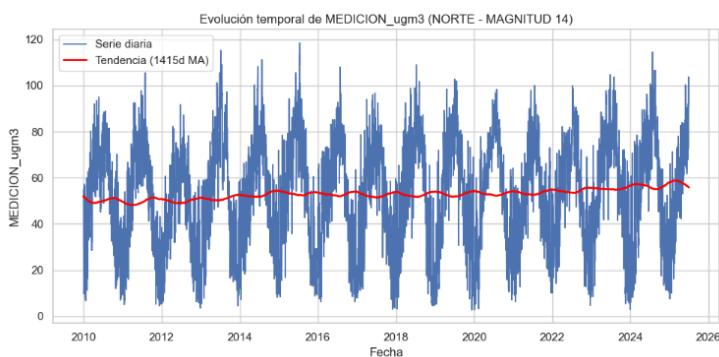


Figura 99 - O3 - Tendencia diaria NORTE

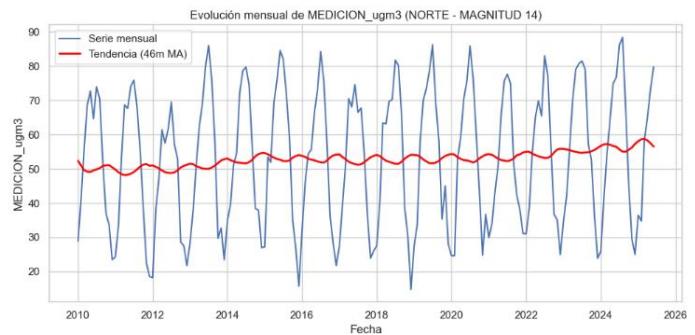


Figura 100 - O3 - Tendencia mensual NORTE

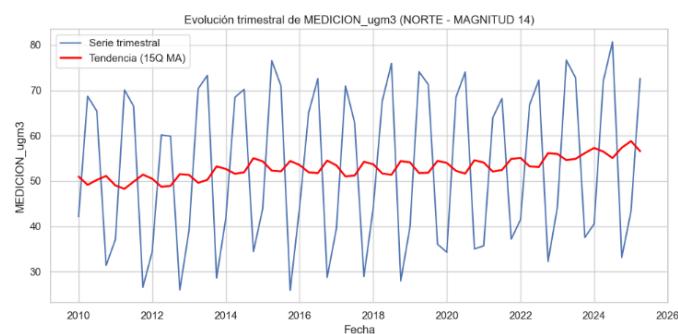


Figura 101 - O3 - Tendencia trimestral NORTE

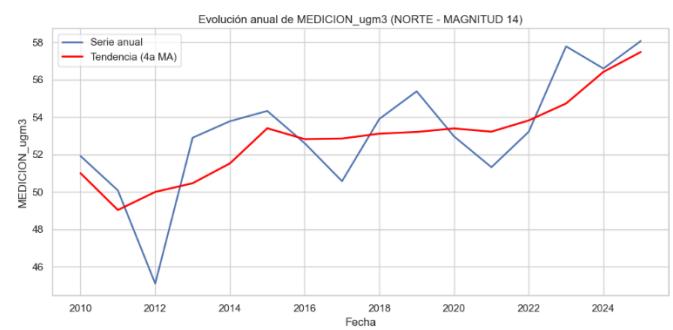


Figura 102 - O3 - Tendencia anual NORTE

## O<sub>3</sub> - CENTRO

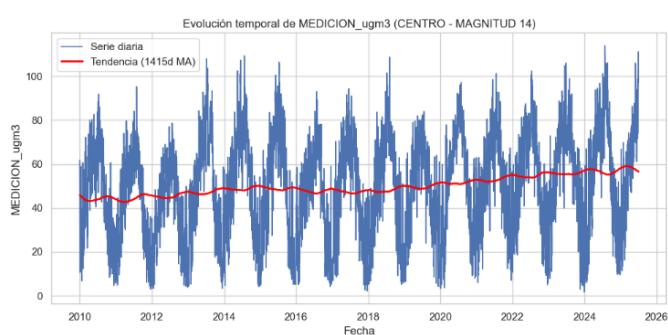


Figura 103 - O3 - Tendencia diaria CENTRO

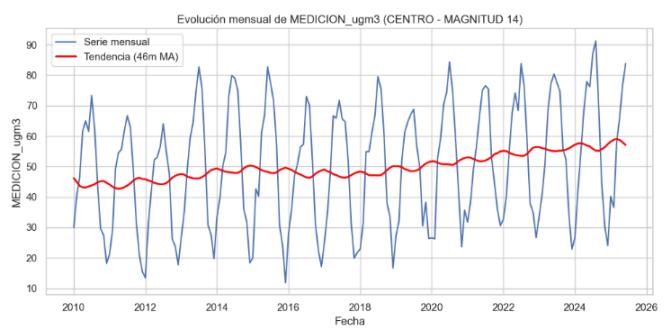


Figura 104 - O3 - Tendencia mensual CENTRO

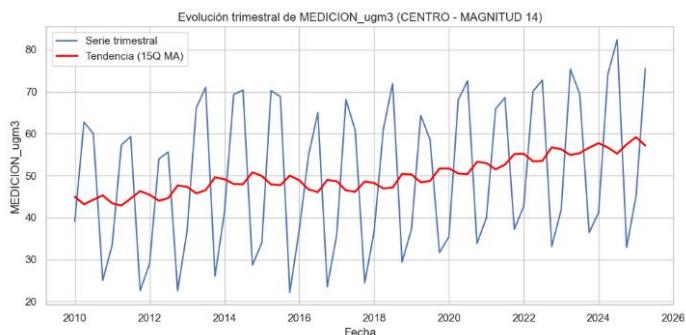


Figura 105 - O3 - Tendencia trimestral CENTRO

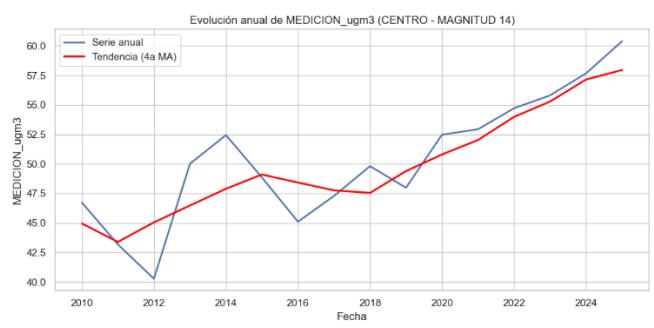


Figura 106 - O3 - Tendencia anual CENTRO

## O<sub>3</sub> - SUR

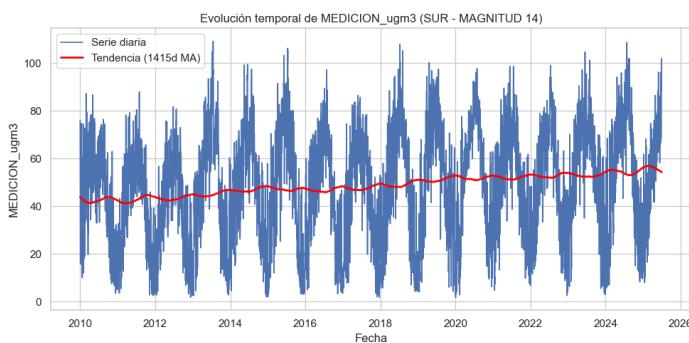


Figura 107 - O3 - Tendencia diaria SUR

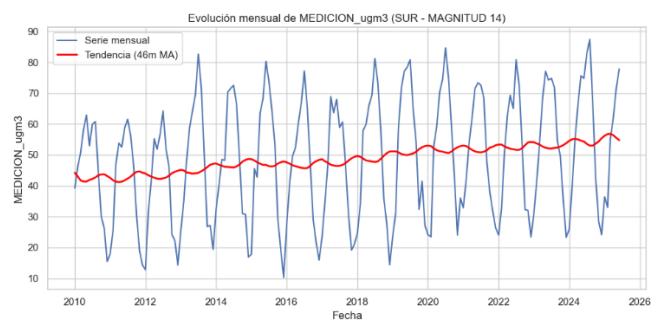


Figura 108 - O3 - Tendencia mensual SUR

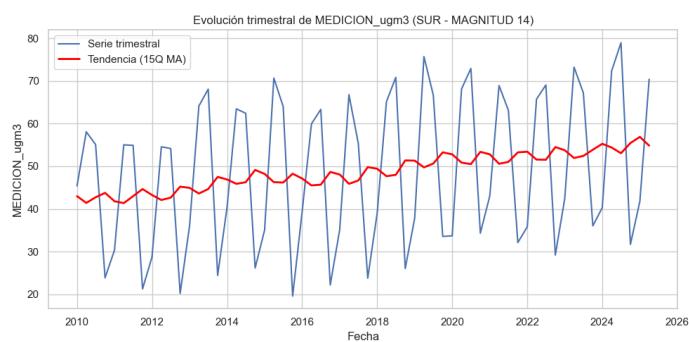


Figura 109 - O3 - Tendencia trimestral SUR

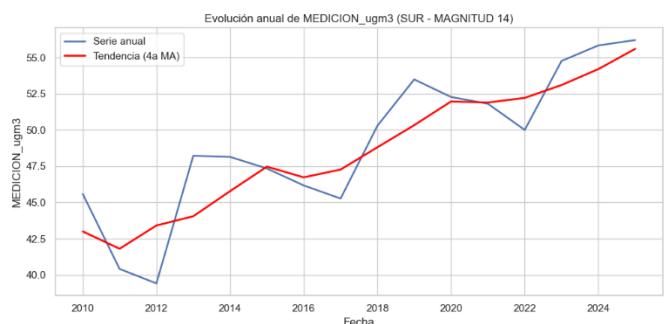


Figura 110 - O3 - Tendencia anual SUR

### Comentarios a las gráficas anteriores:

ID COMB	ZONA	MAGNITD	TENDENCIA GENERAL	VALORES 2025 (µg/m <sup>3</sup> )
1	CENTRO	Partículas < 10 µm	Descenso hasta 2017/Establecimiento posterior	<b>18,08</b>
2	NORTE	Partículas < 10 µm	Descenso hasta 2014/Establecimiento posterior	<b>18,53</b>
3	SUR	Partículas < 10 µm	Descenso hasta 2014/Establecimiento posterior	<b>18,18</b>
7	CENTRO	Ozono	<b>Ligero ascenso sostenido</b>	57,96
8	NORTE	Ozono	<b>Ligero ascenso sostenido</b>	57,5
9	SUR	Ozono	<b>Ligero ascenso sostenido</b>	55,61
28	CENTRO	Dióxido de Nitrógeno	Descenso sostenido	<b>22,57</b>
29	NORTE	Dióxido de Nitrógeno	Descenso sostenido	<b>24,75</b>
30	SUR	Dióxido de Nitrógeno	Descenso sostenido	<b>25,38</b>

31	CENTRO	Partículas < 2,5 µm	Clara tendencia descendente	<b>8,74</b>
32	NORTE	Partículas < 2,5 µm	Clara tendencia descendente	<b>8,56</b>
33	SUR	Partículas < 2,5 µm	Clara tendencia descendente	<b>8,41</b>

Tabla 18 - Comentarios resultados tendencias

## CONCLUSIONES:

- **PM<sub>10</sub>** muestra un descenso notable en la década anterior, seguido de una estabilización en los últimos años en las tres zonas, situándose en 2025 en torno a 18–18,5 µg/m<sup>3</sup> (valor medio anual), valores aún por encima de la recomendación de la OMS.
- **PM<sub>2,5</sub>** presenta una clara tendencia descendente en todas las zonas, alcanzando niveles en 2025 próximos a 8,5–8,75 µg/m<sup>3</sup> (valor medio anual), lo que supone una mejora significativa respecto a años anteriores, aunque todavía lejos del valor guía de la OMS (5 µg/m<sup>3</sup>).
- **NO<sub>2</sub>** mantiene un descenso sostenido en las tres zonas urbanas, convergiendo en valores medios anuales para 2025 entre 22 y 26 µg/m<sup>3</sup>, lo que refleja el impacto positivo de las restricciones al tráfico y la renovación del parque móvil. Sin embargo, estos valores permanecen aún lejos de cumplir con las recomendaciones de la OMS (10 µg/m<sup>3</sup>).
- **O<sub>3</sub>** muestra un comportamiento opuesto al resto de contaminantes: registra un ascenso sostenido, especialmente acusado en la zona Centro, alcanzando valores entre 55 y 58 µg/m<sup>3</sup> en 2025. Este incremento confirma la persistencia del ozono como contaminante crítico por su origen secundario. No obstante, su valor medio anual previsto para 2025 se mantendría por debajo del valor límite recomendado por la OMS.
- **Relación NO<sub>2</sub>–O<sub>3</sub> y efecto de las políticas de reducción del tráfico:** La evolución divergente entre NO<sub>2</sub> (descenso) y O<sub>3</sub> (ascenso) responde en parte a la relación química entre ambos. En entornos urbanos, el NO procedente del tráfico destruye parte del ozono. Al reducirse las emisiones de NOx —como resultado de las restricciones al tráfico y la mejora tecnológica del parque móvil— disminuye también este proceso de eliminación del O<sub>3</sub>, favoreciendo que sus concentraciones se mantengan o incluso aumenten en determinados períodos de alta radiación. Esta situación pone de manifiesto la necesidad de políticas integradas que aborden no solo las emisiones de NOx, sino también las de compuestos orgánicos volátiles (COV) y otros precursores fotoquímicos.
- En conjunto, las series evidencian mejoras claras en los contaminantes primarios asociados al tráfico (PM y NO<sub>2</sub>), mientras que el ozono continúa mostrando una tendencia ascendente, lo que subraya la necesidad de estrategias específicas orientadas a la reducción de sus precursores.

## iv. Correlaciones y variables importantes:

Una vez estudiados los resultados estadísticos, la estacionalidad y la tendencia de todas “**Las Combinaciones**”, se procede a analizar cuáles son las variables exógenas —meteorológicas e incendios— que presentan una mayor influencia sobre su comportamiento. Para ello se combinan tres aproximaciones complementarias: (i) **la correlación estadística**, que identifica asociaciones lineales directas; (ii) **la importancia relativa basada en teoría de juegos** (valores Shapley), que permite estimar el peso marginal de cada variable dentro del modelo; y (iii) **la importancia**

**según contribución al error (RMSE)**, que revela qué predictores resultan críticos para mejorar la capacidad predictiva. Este análisis integrado no solo permite comprender los factores que condicionan la dinámica de cada combinación, sino que resulta esencial para la selección posterior de las variables exógenas que se incorporarán al proceso de predicción de las distintas combinaciones.

## CORRELACIÓN ESTADÍSTICA

La primera aproximación empleada para identificar las variables exógenas más influyentes se basa en el **análisis de correlación estadística**. Mediante la construcción de matrices de correlación específicas para cada una de “**Las Combinaciones**”, se evalúa la relación lineal entre las series de concentración y el conjunto de variables independientes. Para garantizar la validez del análisis, se excluyen las variables no numéricas, así como aquellas que presentan una correlación estructuralmente redundante, como ocurre con las temperaturas máxima y mínima respecto a la temperatura media, o con los valores máximos/mínimos de viento y presión.

La visualización mediante heatmaps permite identificar de manera rápida qué predictores muestran asociaciones —positivas o negativas— más estrechas con cada contaminante. Este enfoque ofrece una primera visión descriptiva de los factores externos que podrían condicionar la evolución de “**Las Combinaciones**”. Aunque la correlación no implica causalidad, sí permite detectar patrones consistentes y orientar los análisis posteriores hacia aquellas variables con mayor potencial explicativo.

Como se indicaba previamente, todos los gráficos generados se encuentran recogidos en la **carpeta 03\_figs\_variable\_MEDICION\_ugm3** del **Anexo A2**. A continuación, se presentan algunos de los gráficos de correlación correspondientes a las combinaciones magnitud–zonas más relevantes.

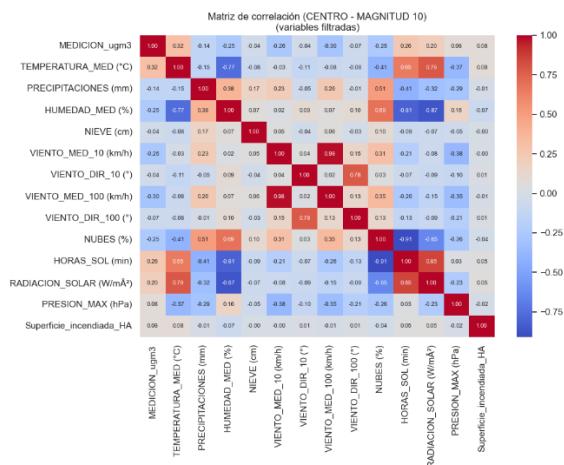


Figura 111 - PM<sub>10</sub> - Correlaciones CENTRO

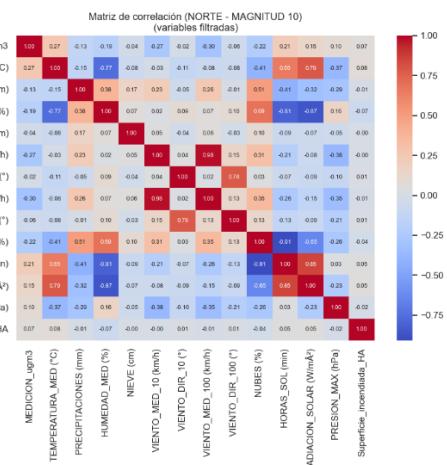


Figura 112 - PM<sub>10</sub> - Correlaciones NORTE

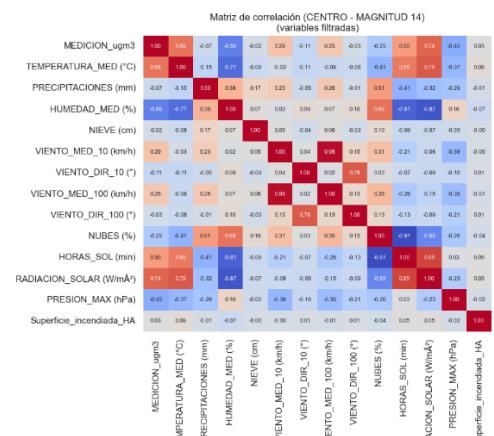


Figura 113 - O3 - Correlaciones CENTRO

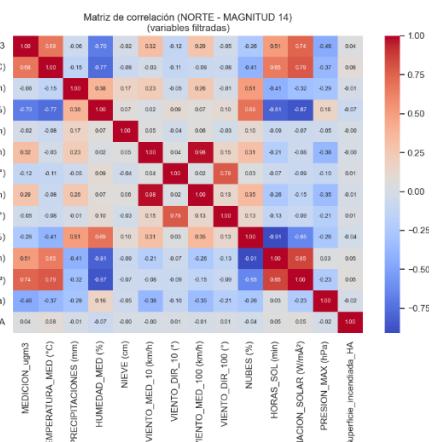


Figura 114 - O3 - Correlaciones NORTE

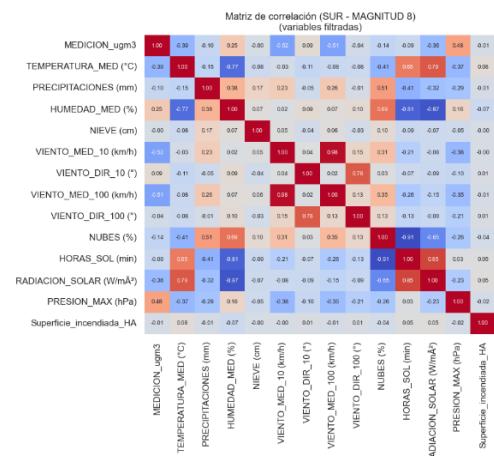


Figura 115 - NO2 - Correlaciones SUR

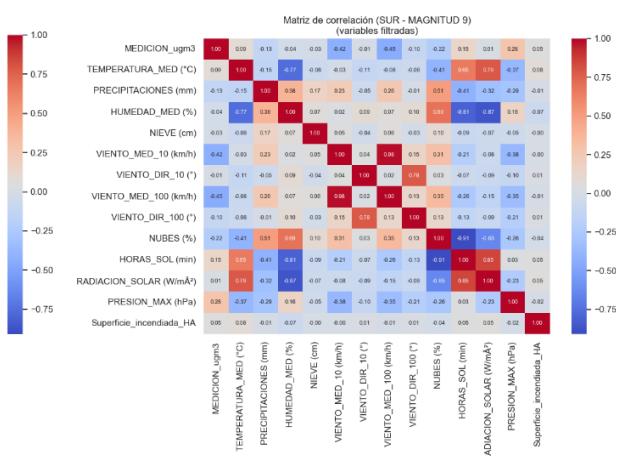


Figura 116 - PM2.5 - Correlaciones SUR

## TEORÍA DE JUEGOS (VALORES SHAPLEY)

La segunda aproximación utilizada para identificar las variables exógenas más influyentes se fundamenta en la **importancia relativa derivada de la teoría de juegos, concretamente mediante los valores Shapley**. Esta metodología permite cuantificar la contribución marginal de cada predictor al resultado del modelo, evaluando su impacto de forma equitativa a través de todas las combinaciones posibles de variables. A diferencia de la simple correlación, los valores Shapley ofrecen una perspectiva completamente multivariante, ya que consideran cómo cambia la predicción cuando una variable se incorpora o se elimina del conjunto. Aplicados sobre **modelos tipo LightGBM**, estos valores permiten identificar qué variables independientes —tanto numéricas como categóricas— aportan mayor capacidad explicativa. Esta aproximación proporciona así una medida robusta, coherente y matemáticamente bien fundamentada de la relevancia individual de cada variable dentro de las dinámicas que condicionan el comportamiento de “**Las Combinaciones**”.

## IMPORTANCIA SEGÚN CONTRIBUCIÓN AL ERROR (RMSE)

La tercera aproximación empleada para determinar la importancia de las variables exógenas se basa en la **contribución al error de predicción, evaluada mediante el**

**RMSE (Root Mean Squared Error).** Para ello se utiliza el procedimiento de permutation importance aplicado sobre un **modelo LightGBM** entrenado para cada combinación, lo que permite medir el impacto real de cada predictor en el rendimiento del modelo. En este enfoque, cada variable independiente —tanto numérica como categórica— se permuta de forma aleatoria en el conjunto de test mientras el resto permanecen sin alterar. Si la variable es relevante, esta permutación provoca un aumento del RMSE, indicando hasta qué punto el modelo depende de dicha información para generar predicciones precisas. El incremento medio del error tras la permutación actúa, así como una medida directa y multivariante de la pérdida de información atribuible a cada predictor. Esta metodología complementa la correlación y los valores Shapley, proporcionando una perspectiva adicional orientada al rendimiento, y contribuye a una selección más sólida de las variables exógenas que se incorporarán al proceso de predicción de todas y cada una de “**Las Combinaciones**”.

Se muestran algunas de las figuras más relevantes:

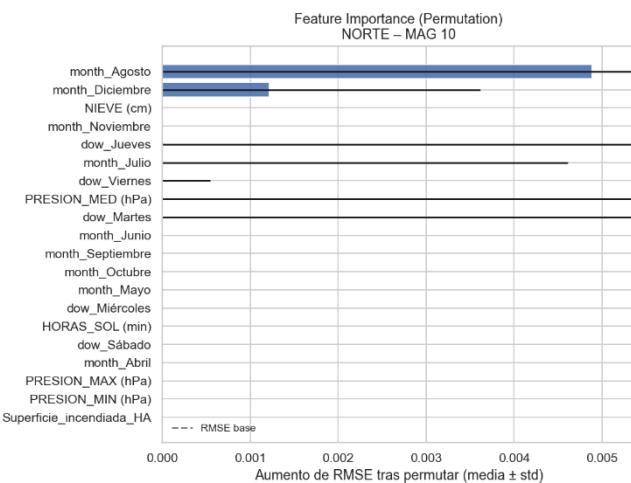


Figura 117 - PM<sub>10</sub> - Importancia RMSE NORTE

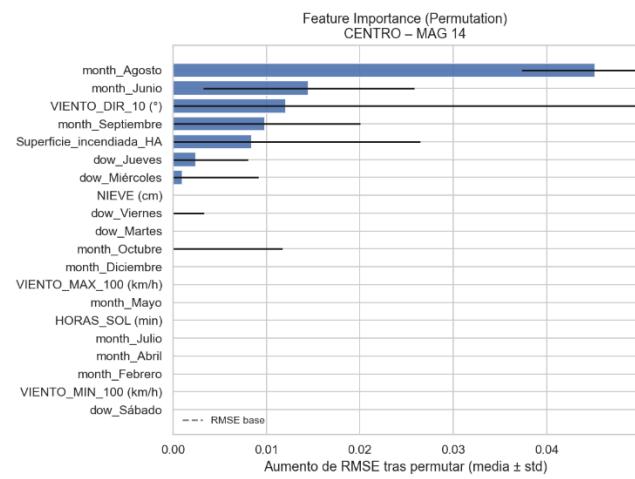


Figura 118 - O3 - Importancia RMSE CENTRO

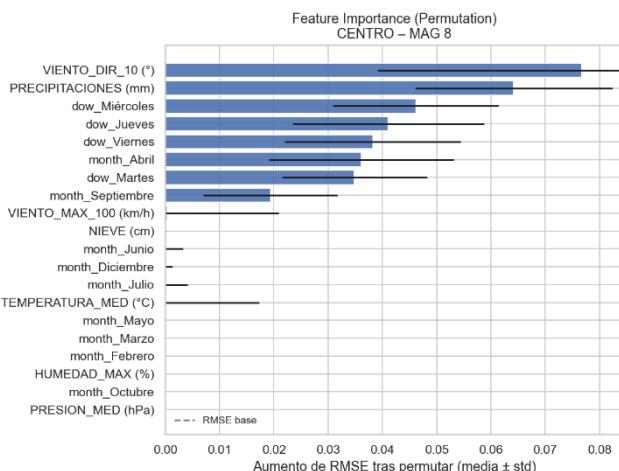


Figura 119 - NO<sub>2</sub> - Importancia RMSE CENTRO

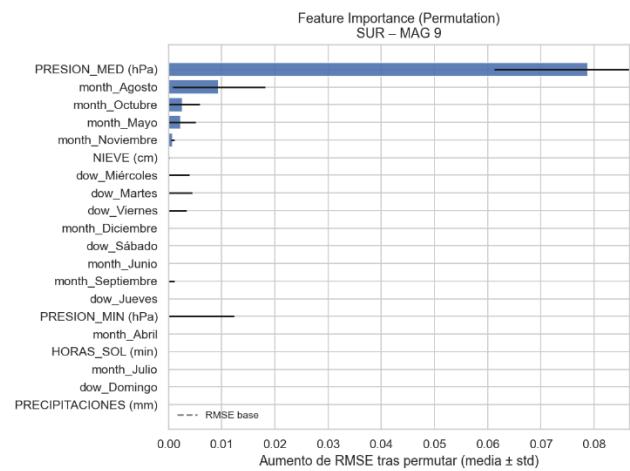


Figura 120 - PM<sub>2,5</sub> - Importancia RMSE SUR

## Comentarios a los resultados anteriores:

ID COMB	ZONA_ESTACIÓN	MAGNITUD - NUM	MAGNITUD - NOM	TOP 3 - VARIABLE MAYOR CORRELACIÓN	TOP 3 - VARIABLE MÁS IMPORTANTE - TEORÍA DE JUEGOS	TOP 3 - VARIABLE MÁS IMPORTANTE - RMSE
1	CENTRO	10	Partículas < 10 µm	1. Temperatura 2. Viento - 100m 3. Horas de sol / Viento 10m	1. Mes - marzo 2. Mes - febrero 3. Día semana - domingo	1. Presión media 2. Mes - diciembre 2. Mes - junio
2	NORTE	10	Partículas < 10 µm	1. Viento 100m 2. Viento 10m / Temperatura 3. %Nubes	1. Mes - marzo 2. Día semana - domingo 3. Mes - febrero	1. Mes - agosto 2. Mes - diciembre 3. Nieve
3	SUR	10	Partículas < 10 µm	1. Viento - 100m 2. Viento 10m 3. Temperatura	1. Mes - marzo 2. Día semana - domingo 3. Mes - febrero	1. Mes - agosto 2. Mes - octubre 3. Mes - diciembre
7	CENTRO	14	Ozono	1. Radiación Solar 2. Temperatura / Humedad 3. Horas de Sol	1. Día semana - domingo 2. Día semana - abril 3. Mes - abril	1. Mes - agosto 2. Mes - junio 3. Dirección viento - 10 m
8	NORTE	14	Ozono	1. Radiación Solar 2. Humedad 3. Temperatura	1. Día semana - domingo 2. Día semana - abril 3. Mes - abril	1. Viento máximo 100m 2. Mes - agosto 3. Presión media
9	SUR	14	Ozono	1. Radiación Solar 2. Humedad 3. Temperatura	1. Día semana - domingo 2. Día semana - abril 3. Mes - abril	1. Mes - agosto 2. Viento máximo - 100m 3. Día semana - miércoles
28	CENTRO	8	Dióxido de Nitrógeno	1. Viento - 10m 2. Viento - 100m 3. Presión	1. Día semana - domingo 2. Día semana - sábado 3. Mes - agosto	1. Dirección viento - 10 m 2. Precipitaciones 3. Día semana - miércoles
29	NORTE	8	Dióxido de Nitrógeno	1. Viento - 10m 2. Viento - 100m 3. Presión	1. Día semana - domingo 2. Día semana - sábado 3. Mes - agosto	1. Temperatura media 2. Horas sol 3. Precipitaciones
30	SUR	8	Dióxido de Nitrógeno	1. Viento - 10m 2. Viento - 100m 3. Presión	1. Día semana - domingo 2. Día semana - sábado 3. Mes - agosto	1. Dirección viento - 10 m 2. Mes - julio 3. Día semana - miércoles
31	CENTRO	9	Partículas < 2,5 µm	1. Viento - 100m 2. Viento - 10m 3. Presión	1. Mes - marzo 2. Día semana - domingo 3. Mes - febrero	1. Presión media 2. Presión mínima 3. Mes - octubre
32	NORTE	9	Partículas < 2,5 µm	1. Viento - 100m 2. Viento - 10m 3. Presión	1. Día semana - domingo 2. Mes - marzo 3. Superficie incendiada	1. Mes - agosto 2. Presión mínima 3. Mes - mayo
33	SUR	9	Partículas < 2,5 µm	1. Viento - 100m 2. Viento - 10m 3. Presión	1. Día semana - domingo 2. Superficie incendiada 3. Mes - marzo	1. Presión media 2. Mes - agosto 3. Mes - octubre

Tabla 19 - Resultados correlaciones y variables importantes

## CONCLUSIONES:

- Los contaminantes primarios (**PM<sub>10</sub>, PM<sub>2,5</sub> y NO<sub>2</sub>**) están fuertemente condicionados por la ventilación atmosférica. En todas las zonas analizadas, el viento —tanto a 10 m como a 100 m— aparece sistemáticamente como la variable con mayor correlación, acompañado por la presión atmosférica.
- El **ozono troposférico** presenta un comportamiento diferenciado, estrechamente ligado a procesos fotoquímicos. La radiación solar, la temperatura, la humedad y las horas de sol constituyen las variables con mayor correlación en las tres zonas, evidenciando el carácter secundario del O<sub>3</sub> y su dependencia directa de la intensidad de la radiación.
- Las variables temporales (**mes del año y día de la semana**) emergen como los factores más relevantes en términos de importancia para los modelos predictivos. Tanto por teoría de juegos como por análisis de contribución al error (RMSE), las variables temporales dominan en todos los contaminantes y en todas las zonas, reflejando la **fuerte estacionalidad, periodicidad semanal y patrones de actividad urbana que modulan las concentraciones**. El **domingo** aparece como el día más distintivo por la fuerte reducción del tráfico.
- La **presión atmosférica y el mes del año** explican una parte sustancial del error de predicción. La presión media y mínima, junto con los meses críticos (agosto, diciembre, octubre), concentran la mayor contribución al RMSE. Estos factores suelen asociarse a episodios de estabilidad atmosférica, inversiones térmicas o transiciones estacionales que los modelos capturan con mayor dificultad.
- Eventos meteorológicos menos frecuentes —como nieve, precipitaciones o viento máximo— tienen un impacto localizado en el error del modelo. No son variables dominantes en correlación ni en importancia estructural, pero sí condicionan significativamente la precisión del modelo en episodios concretos, lo que sugiere la necesidad de estrategias específicas para la gestión de extremos meteorológicos.

- El análisis revela que la dinámica de los contaminantes primarios está determinada por los procesos de dispersión y por los ciclos temporales, mientras que el ozono sigue un patrón fotoquímico propio. Este contraste subraya la importancia de estrategias diferenciadas.

### 6.3. CONCLUSIONES GLOBALES

El análisis exploratorio muestra que las variables meteorológicas y los incendios forestales presentan comportamientos coherentes, con tendencias y ciclos estacionales bien definidos. Se identifican patrones climatológicos característicos de Madrid —como el incremento de las temperaturas, la estacionalidad marcada de la radiación solar y la concentración de incendios en verano— junto con eventos singulares como Filomena o el gran incendio de Ávila, todos ellos relevantes para comprender su influencia sobre la calidad del aire.

En contraste, **la variable de tráfico TOTAL\_HOR\_EXT presenta inconsistencias significativas en los datos previos a 2019–2020, lo que impide su utilización inmediata en modelos explicativos o predictivos.** Aunque las series diarias reflejan con claridad el impacto de la COVID-19, la falta de coherencia histórica justifica su exclusión de manera temporal.

Por otro lado, las conclusiones del EDA de la variable objetivo se apoyan en el conjunto de análisis realizados —estadísticos, estacionales, de tendencia y de correlaciones/variables importantes— a partir del cual puede extraerse una síntesis clara. En primer lugar, los resultados muestran que **PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub> y O<sub>3</sub> son las magnitudes más preocupantes para la salud humana**, superando de forma recurrente los valores límite o recomendados por la normativa europea y por las guías de la OMS, tanto en medias anuales como en episodios de máximos y percentiles elevados. En cuanto a patrones temporales, las partículas (PM<sub>10</sub> y PM<sub>2.5</sub>) presentan distribuciones asimétricas, **fuerte estacionalidad** y una dinámica vinculada a invierno y verano, con descensos ligeros en fin de semana coherentes con la reducción del tráfico. El NO<sub>2</sub> registra máximos en invierno y disminuciones acusadas los domingos, lo que refuerza su relación con la actividad urbana. Por el contrario, el O<sub>3</sub> alcanza sus máximos en verano y aumenta ligeramente en fin de semana, reflejo de su carácter fotoquímico y de la relación inversa con el NO<sub>2</sub>.

En cuanto a las tendencias, se constata una mejora apreciable en los contaminantes primarios asociados al tráfico: PM<sub>10</sub> y PM<sub>2.5</sub> descienden de forma progresiva, aunque se mantienen por encima de las recomendaciones de la OMS, y el NO<sub>2</sub> presenta una reducción sostenida hasta estabilizarse en valores intermedios (22–26 µg/m<sup>3</sup>), aún insuficientes para alcanzar los objetivos más exigentes. **El ozono, por el contrario, muestra una tendencia ascendente, especialmente en la zona Centro,** convirtiéndose en un contaminante crítico pese a que su media anual se mantiene cerca del límite recomendado.

El análisis de correlaciones y de variables explicativas confirma que la dinámica de los contaminantes primarios está dominada por la ventilación atmosférica (viento, presión) y por los ciclos temporales (mes y día de la semana), mientras que el ozono depende principalmente de la radiación solar, la temperatura y otros factores fotoquímicos. Esta dualidad, junto con la relación inversa NO<sub>2</sub>–O<sub>3</sub>, evidencia que **las políticas centradas exclusivamente en la reducción de NOx y partículas no son suficientes para controlar el ozono, siendo necesario intervenir también sobre los COV y el conjunto de precursores.**

En conclusión, a la luz de los límites normativos y de los objetivos del PNIEC 2023–2030 y de la Agenda 2030, las magnitudes verdaderamente prioritarias en el entorno urbano de Madrid son PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub> y O<sub>3</sub>. El resto de contaminantes analizados (SO<sub>2</sub>, CO,tolueno, benceno, etilbenceno y NOx en su vertiente de protección vegetal) presentan niveles bajos o de relevancia limitada en un contexto urbano. Por ello, el

estudio posterior se centra en analizar la evolución futura de estas cuatro magnitudes críticas en las tres zonas de la ciudad, como base para evaluar la viabilidad de cumplir los objetivos de calidad del aire y garantizar una protección efectiva de la salud pública.

## 7. MODELOS DE PREDICCIÓN

El presente apartado desarrolla la fase de modelización predictiva, cuyo objetivo es estimar la evolución futura de las magnitudes **PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub> y O<sub>3</sub>**—como vimos, las más perjudiciales desde el punto de vista de la salud humana— hasta el año 2032, poniendo **especial énfasis en el horizonte 2030** por su relevancia en los compromisos establecidos en el PNIEC y en los objetivos de la Agenda 2030. Para cada contaminante y cada zona de estudio (12 combinaciones en total), se ejecuta un procedimiento de predicción que integra modelos clásicos de series temporales (ARIMA y SARIMA), algoritmos de machine learning (XGBoost, LightGBM y CatBoost) y modelos de deep learning (LSTM y CNN). La comparación entre modelos se basa en su capacidad predictiva, evaluada mediante el **error RMSE** sobre un conjunto de test comprendido entre el 1 de julio de 2024 y el 30 de junio de 2025. Finalmente, para cada combinación MAGNITUD-ZONA se selecciona el modelo con mejor desempeño, cuyas proyecciones permiten analizar las tendencias futuras y estimar la concentración media anual prevista para 2030.

Por tanto, los cuatro subapartados siguientes se centran en cada una de las magnitudes mencionadas, aplicando en todos los casos el mismo proceso para obtener la predicción de la concentración media anual prevista para 2030 en las tres zonas de estudio (NORTE, CENTRO y SUR). Además, una vez generados y evaluados todos los modelos anteriormente mencionados, el procedimiento se complementa con una segunda fase de modelización basada exclusivamente en aquellos algoritmos que hayan mostrado el menor RMSE para cada MAGNITUD-ZONA. Estos modelos seleccionados se reentrenan ampliando el número de variables mediante la inclusión de la variable TRAFICO, altamente relevante en la dinámica de los contaminantes analizados, y restringiendo la base de datos a observaciones posteriores al 1 de enero de 2019 para asegurar coherencia temporal y homogeneidad metodológica. Adicionalmente, y con independencia de si fue inicialmente el modelo óptimo para una zona concreta, se incorpora en esta segunda fase el algoritmo CatBoost, dada su elevada eficiencia cuando el tamaño de entrenamiento es reducido y su capacidad para capturar interacciones no lineales sin requerir preprocesamiento específico de variables categóricas.

Todos estos modelos de predicción, se ejecutan dentro de los archivos *07.1\_Modelos de predicción\_MAG10.ipynb*, *07.2\_Modelos de predicción\_MAG14.ipynb*, *07.3\_Modelos de predicción\_MAG8.ipynb* y *07.4\_Modelos de predicción\_MAG9.ipynb*, incluidos íntegramente dentro del **Anexo A1** del presente documento.

Señalar además, que todos los gráficos y resultados generados en este apartado, se pueden localizar en la carpeta correspondiente al Anexo A3 Gráficos Modelos de Predicción.

A continuación, se muestra un diagrama de flujo con el proceso que se va a seguir en el presente apartado:

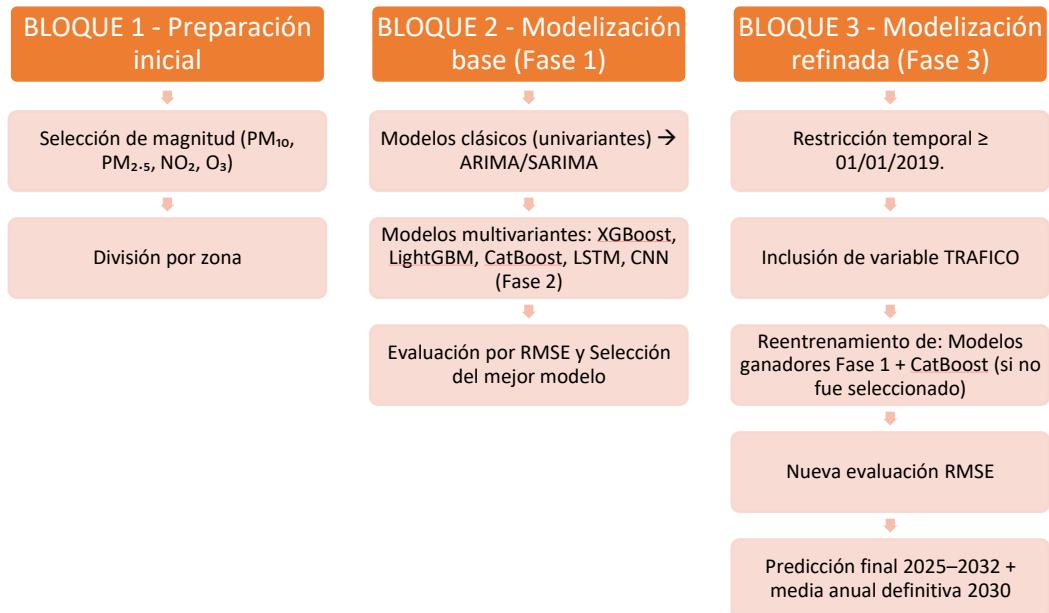


Figura 121 - Proceso Modelos de predicción

## 7.1. BLOQUE 1 – PREPARACIÓN INCIAL

Como se indicó al final del apartado anterior, el análisis se orienta a evaluar la evolución futura de las magnitudes críticas en las distintas zonas de la ciudad, en coherencia con los objetivos del PNIEC 2023–2030 y de la Agenda 2030, que priorizan la mejora de la calidad del aire y la protección de la salud pública en entornos urbanos. En consecuencia, este subapartado se dedica exclusivamente a la construcción de tantas bases de datos como combinaciones magnitud–zona existan, las cuales constituirán la estructura de entrada sobre la que posteriormente se ejecutarán los distintos modelos de predicción. Con el fin de garantizar estimaciones consistentes y adecuadas para los modelos clásicos de series temporales, **las bases de datos diarias se transforman en series mensuales mediante agregación**, obteniendo así la frecuencia temporal deseada para llevar a cabo las predicciones.

## 7.2. BLOQUE 2/FASE 1 – MODELOS CLÁSICOS (UNIVARIANTES)

En esta etapa se emplean modelos **ARIMA** y **SARIMA**, dos metodologías univariantes ampliamente consolidadas en el análisis de series temporales y especialmente adecuadas para capturar dependencias autorregresivas, efectos de memoria y componentes estacionales. Su utilización permite evaluar en qué medida la dinámica intrínseca de cada contaminante —sin incorporar variables exógenas— es capaz de explicar y predecir su evolución futura.

### i. MODELO ARIMA

El modelo ARIMA constituye una de las metodologías fundamentales en el análisis univariante de series temporales y se emplea para predecir la evolución futura de una variable en función **exclusivamente de su propio comportamiento histórico**. Su formulación combina componentes autorregresivos, de diferenciación y de media móvil, lo que permite capturar dependencias temporales, corregir tendencias y modelizar la influencia de errores pasados. Gracias a esta estructura flexible, ARIMA resulta

especialmente adecuado para series sin estacionalidad marcada y sirve como referencia metodológica sólida frente a enfoques más complejos.

### SERIES MENSUALES --> DETECTAMOS TENDENCIA/ESTACIONALIDAD

A partir de las series mensuales, sacamos conclusiones sobre la tendencia y la estacionalidad:

ZONA	MAGNITUD	SERIE MENSUAL	TENDENCIA	ESTACIONALIDAD
NORTE	PM <sub>10</sub>	Figura 64	Descenso hasta 2014/Estabilización posterior	Débil
CENTRO	PM <sub>10</sub>	Figura 68	Descenso hasta 2017/Estabilización posterior	Débil
SUR	PM <sub>10</sub>	Figura 72	Descenso hasta 2014/Estabilización posterior	Débil
NORTE	O <sub>3</sub>	Figura 100	Ligero ascenso sostenido	Muy marcada
CENTRO	O <sub>3</sub>	Figura 104	Ligero ascenso sostenido	Muy marcada
SUR	O <sub>3</sub>	Figura 108	Ligero ascenso sostenido	Muy marcada
NORTE	NO <sub>2</sub>	Figura 88	Descenso sostenido	Anual muy marcada
CENTRO	NO <sub>2</sub>	Figura 92	Descenso sostenido	Anual muy marcada
SUR	NO <sub>2</sub>	Figura 96	Descenso sostenido	Anual muy marcada
NORTE	PM <sub>2.5</sub>	Figura 76	Clara tendencia descendente	Débil
CENTRO	PM <sub>2.5</sub>	Figura 80	Clara tendencia descendente	Débil
SUR	PM <sub>2.5</sub>	Figura 84	Clara tendencia descendente	Débil

Tabla 20 - Series mensuales - Tendencia & Estacionalidad

### DIFERENCIAR (d) HASTA SERIE ESTACIONARIA

Se evalúa la estacionariedad de las series mediante el test de Dickey–Fuller, comenzando con una primera iteración sin diferenciar ( $d = 0$ ). Si el valor  $p$  supera el umbral de significación ( $p > 0,05$ ), se aplica una primera diferenciación (2<sup>a</sup> iteración) y se repite la prueba. En caso de persistir la no estacionariedad, se realiza una segunda diferenciación, evitando exceder este límite para no introducir ruido ni deteriorar la capacidad predictiva del modelo. Con este procedimiento se determina el valor óptimo de  $d$ . Obtenemos los siguientes resultados del parámetro  $d$  óptimo:

	PM <sub>10</sub>			PM <sub>2.5</sub>			NO <sub>2</sub>			O <sub>3</sub>		
	Norte	Centro	Sur	Norte	Centro	Sur	Norte	Centro	Sur	Norte	Centro	Sur
d óptimo	1	1	0	1	1	1	1	1	1	1	1	1

Tabla 21 - Parámetro  $d$  óptimo

## GRÁFICOS ACF y PACF → ELEGIR PARÁMETROS p y q

Se identifican los modelos ARIMA candidatos mediante el análisis de los gráficos ACF y PACF, que permiten proponer combinaciones iniciales de los parámetros p y q. Se muestran **algunos** de los gráficos ACF y PACF:

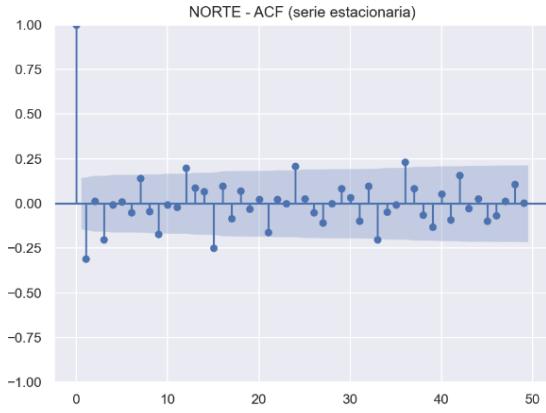


Figura 122 - PM<sub>10</sub> - ACF NORTE

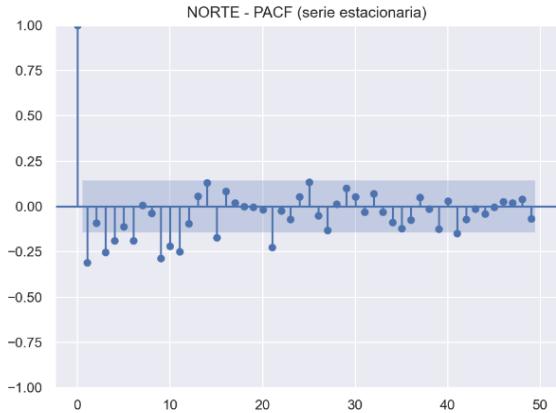


Figura 123 - PM<sub>10</sub> - PACF NORTE

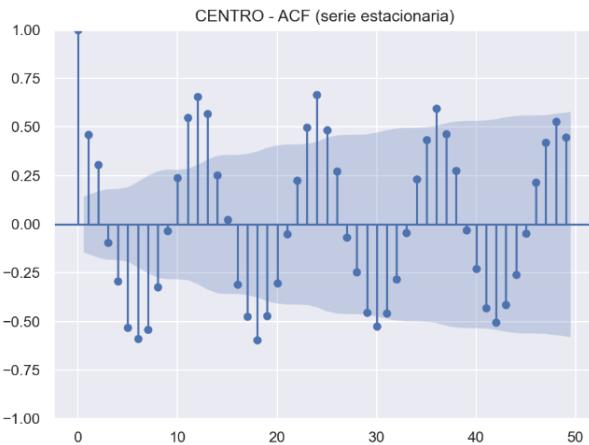


Figura 124 - O<sub>3</sub> - ACF CENTRO

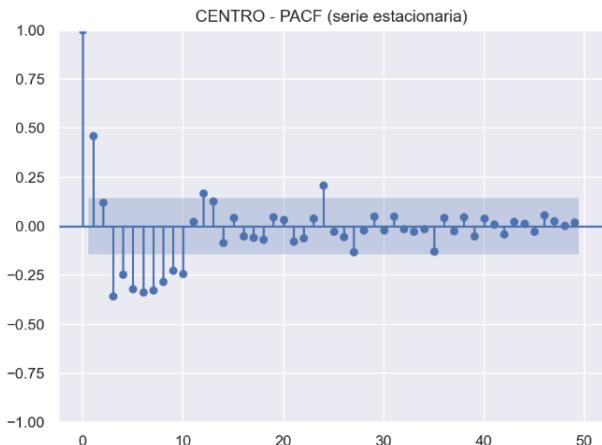


Figura 125 - O<sub>3</sub> - PACF CENTRO

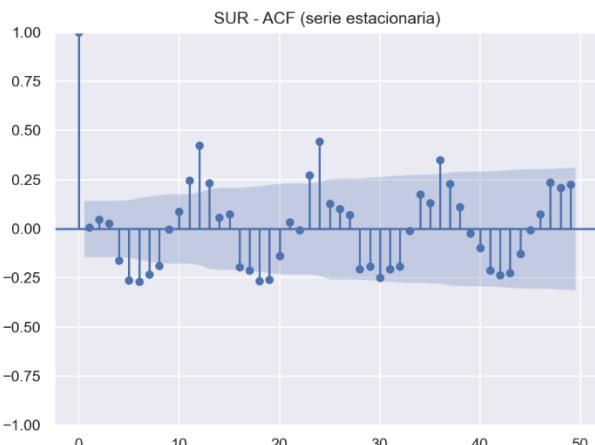


Figura 126 - NO<sub>2</sub> - ACF SUR

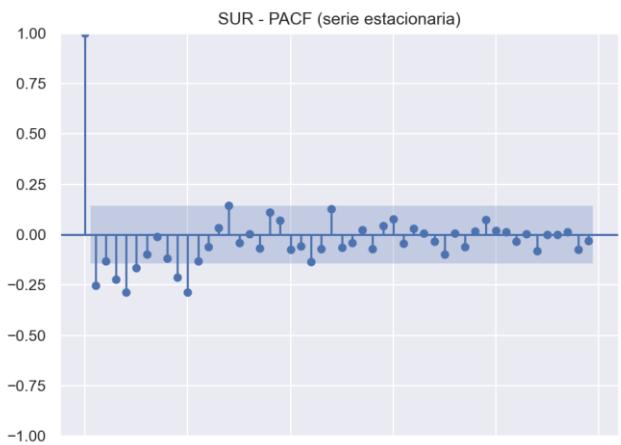


Figura 127 - PM<sub>2.5</sub> - PACF SUR

#### 4) COMPARAR ARIMAS CON AIC

Basándome en los gráficos ACF y PACF de arriba, puedo obtener algunos ARIMA candidatos, seleccionando los parámetros p y q, los cuales se muestran en la siguiente tabla:

ZONA	MAGNITUD	ARIMA 1	ARIMA 2	ARIMA 3	ARIMA 4
NORTE	PM <sub>10</sub>	(0,1,1)	(1,1,1)		
CENTRO	PM <sub>10</sub>	(0,1,1)	(1,1,0)		
SUR	PM <sub>10</sub>	(1,0,0)	(1,0,1)	(0,0,1)	
NORTE	O <sub>3</sub>	(1,1,1)	(1,1,2)	(1,0,1)	(1,0,2)
CENTRO	O <sub>3</sub>	(1,1,1)	(1,1,2)	(1,0,1)	(1,0,2)
SUR	O <sub>3</sub>	(1,1,1)	(1,1,2)	(1,0,1)	(1,0,2) <sup>3</sup>
NORTE	NO <sub>2</sub>	(0,1,1)	(1,1,1)		
CENTRO	NO <sub>2</sub>	(0,1,1)	(1,1,0)		
SUR	NO <sub>2</sub>	(1,1,0)	(1,1,1)	(0,1,1)	
NORTE	PM <sub>2.5</sub>	(0,1,1)	(1,1,1)		
CENTRO	PM <sub>2.5</sub>	(0,1,1)	(1,1,0)	(0,1,0)	
SUR	PM <sub>2.5</sub>	(1,1,0)	(1,1,1)	(0,1,1)	

Tabla 22 - ARIMA candidatos

Por cada combinación MAGNITUD-ZONA se lleva a cabo la selección del modelo ARIMA óptimo a partir del conjunto de configuraciones candidatas predefinidas arriba. Para cada orden (p,d,q) propuesto, se ajusta el modelo sobre el conjunto de entrenamiento y se calcula su AIC, que actúa como criterio principal de comparación. La configuración con el AIC más bajo se considera la más adecuada para dicha combinación y se utiliza posteriormente para generar predicciones sobre el conjunto de prueba (o test), evaluando su precisión mediante el RMSE. De modo que, se obtienen los siguientes ARIMA óptimos por cada combinación MAGNITUD-ZONA:

ZONA	MAGNITUD	ARIMA ÓPT	AIC	RMSE
NORTE	PM <sub>10</sub>	(1, 1, 1)	1065.49	6.049
CENTRO	PM <sub>10</sub>	(0, 1, 1)	1104.08	6.525
SUR	PM <sub>10</sub>	(1, 0, 0)	1069.22	7.006
NORTE	O <sub>3</sub>	(1, 0, 2)	1283.93	19.028
CENTRO	O <sub>3</sub>	(1, 0, 2)	1267.54	20.669
SUR	O <sub>3</sub>	(1, 0, 2)	1280.43	18.894
NORTE	NO <sub>2</sub>	(0, 1, 1)	1205.93	11.929
CENTRO	NO <sub>2</sub>	(0, 1, 1)	1187.18	11.596
SUR	NO <sub>2</sub>	(1, 1, 0)	1243.45	12.907
NORTE	PM <sub>2.5</sub>	(1, 1, 1)	792.22	2.760

<sup>3</sup> Dada la alta estacionalidad anual que presentan los gráficos ACF, se prueba también con d=0.

CENTRO	$PM_{2.5}$	(0, 1, 1)	805.95	2.595
SUR	$PM_{2.5}$	(1, 1, 1)	845.21	2.666

Tabla 23 - ARIMAs Óptimos

Al ejecutar estos modelos se obtienen pronósticos a largo plazo (a partir de 2025) prácticamente planos y con **una estacionalidad apenas apreciable**. Este comportamiento es indicativo de que los modelos ARIMA seleccionados están sobreajustando el ruido a corto plazo o de que el proceso de diferenciación ha eliminado en exceso la estructura temporal. Por ello, se decide avanzar hacia la modelización mediante SARIMA, que incorpora de forma explícita la componente estacional y permite capturar oscilaciones más realistas en el horizonte de largo plazo.

## ii. MODELO SARIMA

El modelo SARIMA (Seasonal ARIMA) amplía la estructura clásica de ARIMA incorporando términos que representan explícitamente la estacionalidad de la serie. Mediante la inclusión de parámetros estacionales autorregresivos, de diferenciación y de medias móviles, SARIMA es capaz de reflejar patrones que se repiten de forma periódica y que no pueden capturarse adecuadamente con un ARIMA estándar. Esta formulación permite reproducir de manera más precisa la dinámica cíclica propia de las series mensuales o anuales, proporcionando proyecciones más coherentes y con mayor capacidad para mantener oscilaciones a medio y largo plazo.

Para cada combinación magnitud, se han elegido los siguientes parámetros de cada modelo SARIMA, teniendo en cuenta que  $s=12$ , por la periodicidad de la estacionalidad (12 para datos mensuales):

MAGNITUD	PARÁMETRO	VALOR PARÁMETRO	COMENTARIO
$PM_{10}$	P	1	Dependencia mismo mes año anterior
	D	1	Pasa a un patrón repetitivo anual
	Q	1	Corrige errores predicción estacionales
$O_3$	P	1	Dependencia mismo mes año anterior
	D	0	No necesita diferenciación. Patrón anual estable
	Q	1	Corrige errores predicción estacionales
$NO_2$	P	0	Los PACF estacionales (lags 12, 24...) no muestran cortes bruscos
	D	1	Pasa a un patrón repetitivo anual
	Q	1	Corrige errores de predicción estacionales
$PM_{2.5}$	P	0	Los PACF estacionales (lags 12, 24...) no muestran cortes bruscos
	D	1	Pasa a un patrón repetitivo anual
	Q	1	Corrige errores predicción estacionales

Tabla 24 - Parámetros SARIMA

Para cada combinación MAGNITUD-ZONA se ejecutan cada uno de sus SARIMA óptimo, con los siguientes resultados:

### PM<sub>10</sub>

NORTE	SARIMA (1,1,1)(1,1,1,12)
CENTRO	SARIMA (0,1,1)(1,1,1,12)
SUR	SARIMA (1,0,0)(1,1,1,12)

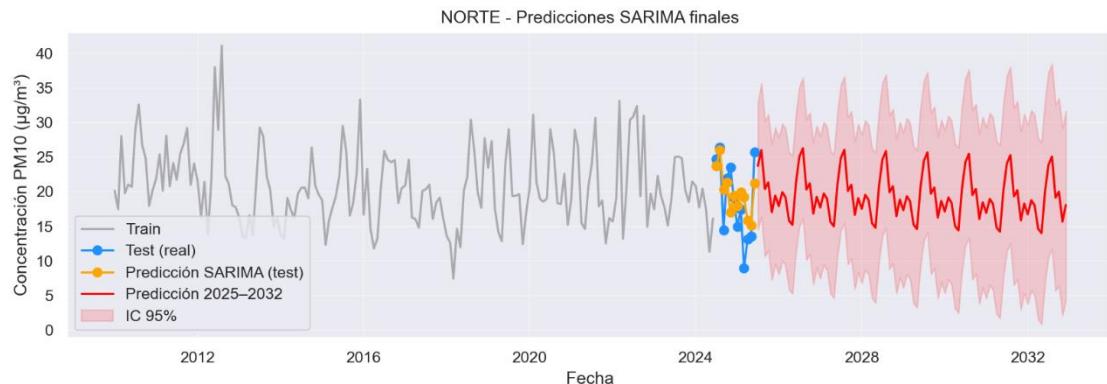


Figura 128 - PM<sub>10</sub> - SARIMA óptimo NORTE

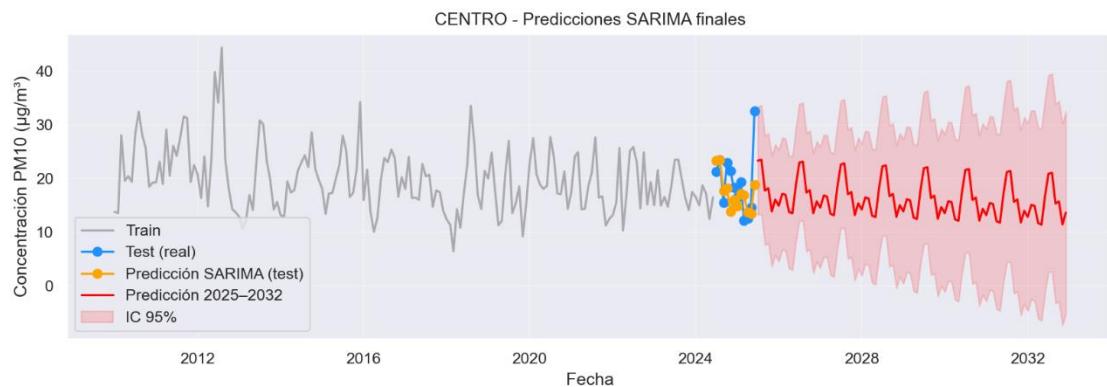


Figura 129 - PM<sub>10</sub> - SARIMA óptimo CENTRO

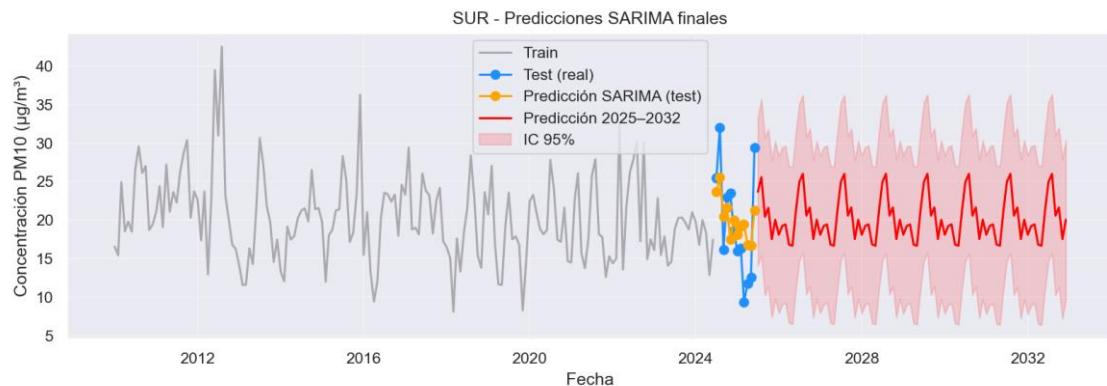


Figura 130 - PM<sub>10</sub> - SARIMA óptimo SUR

### $O_3$

NORTE	SARIMA (1,0,2)(1,0,1,12)
CENTRO	SARIMA (1,0,2)(1,0,1,12)
SUR	SARIMA (1,0,2)(1,0,1,12)

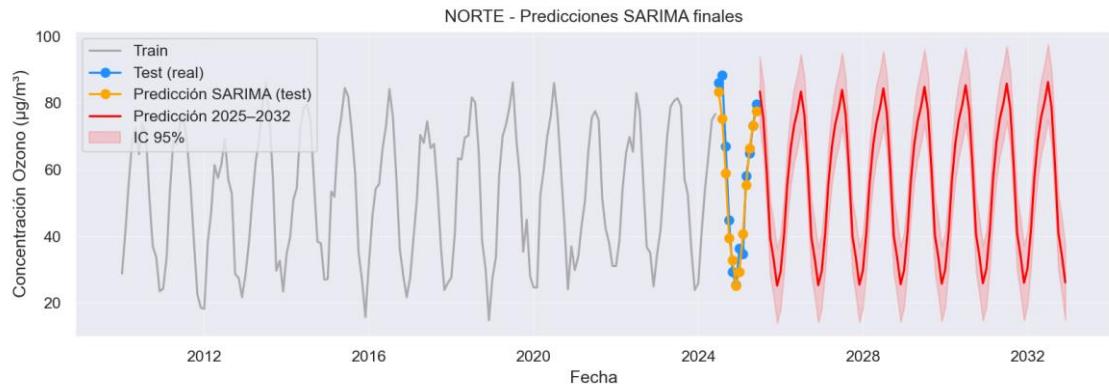


Figura 131 -  $O_3$  - SARIMA óptimo NORTE

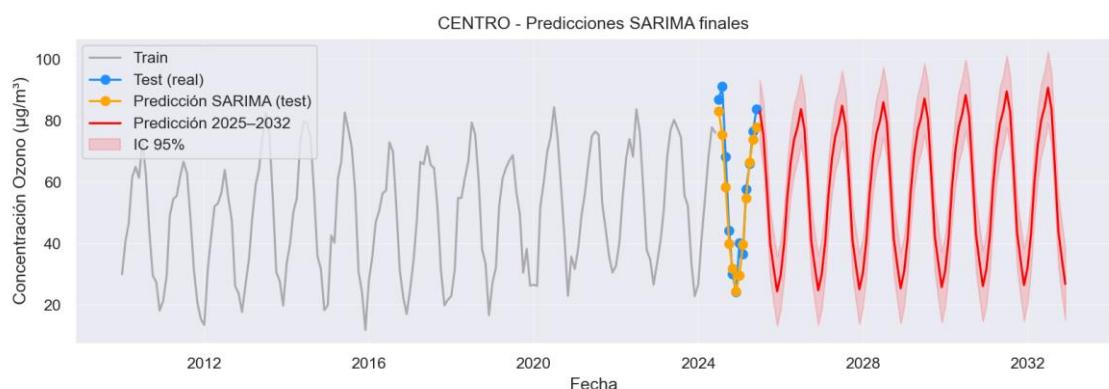


Figura 132 -  $O_3$  - SARIMA óptimo CENTRO

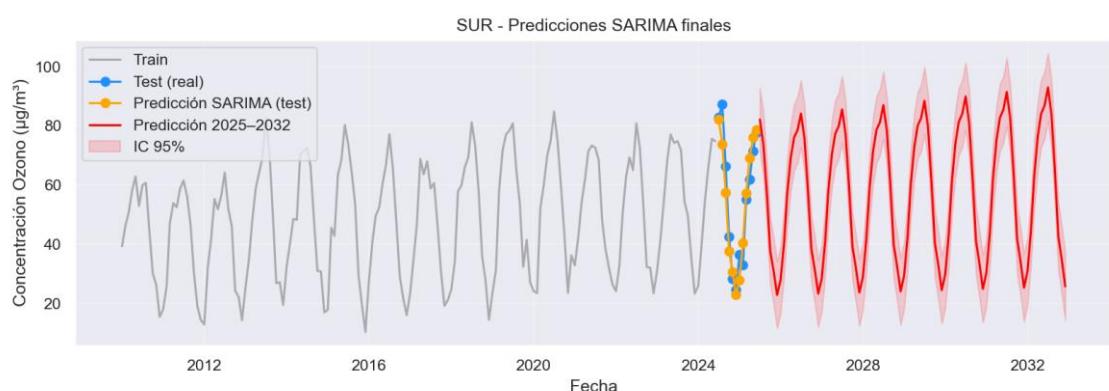


Figura 133 -  $O_3$  - SARIMA óptimo SUR

## NO<sub>2</sub>

NORTE	SARIMA (0,1,1)(0,1,1,12)
CENTRO	SARIMA (0,1,1)(0,1,1,12)
SUR	SARIMA (1,1,0)(0,1,1,12)

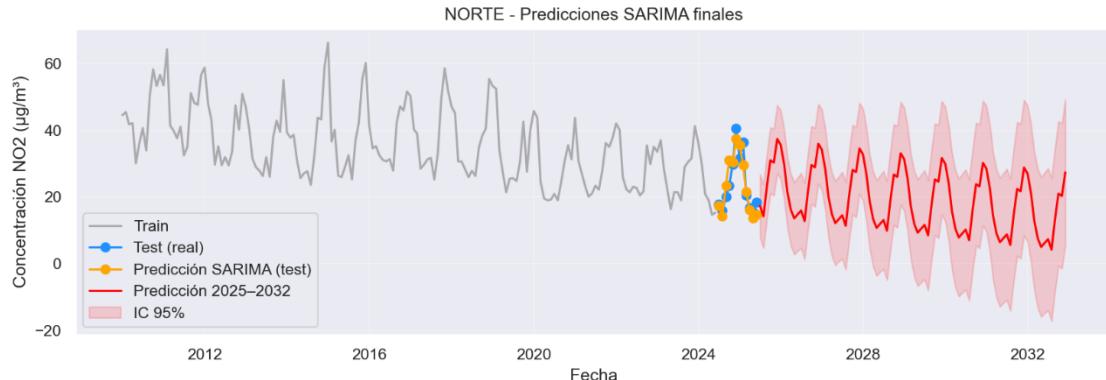


Figura 134 - NO<sub>2</sub> - SARIMA óptimo NORTE

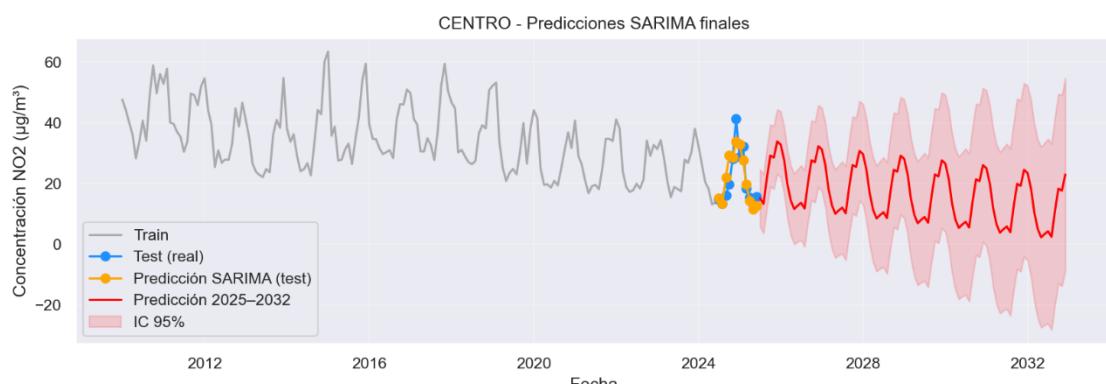


Figura 135 - NO<sub>2</sub> - SARIMA óptimo CENTRO

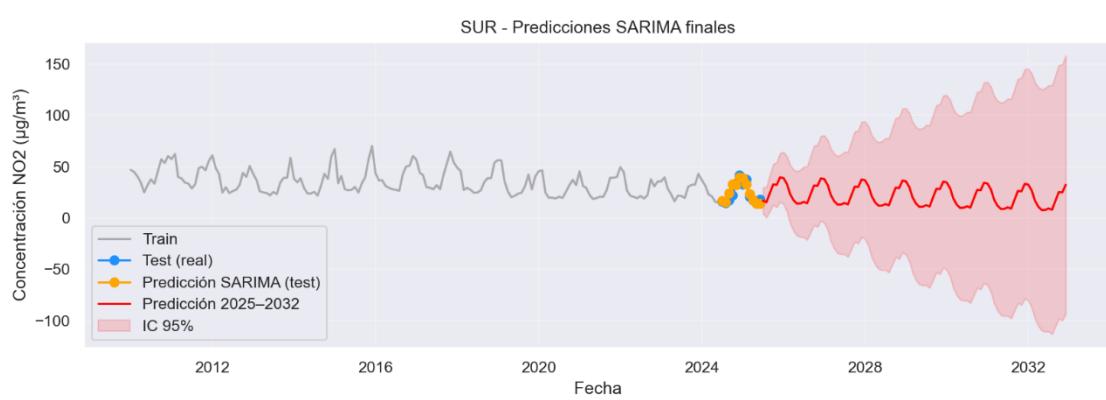


Figura 136 - NO<sub>2</sub> - SARIMA óptimo SUR

## PM<sub>2.5</sub>

NORTE	SARIMA (1,1,1)(0,1,1,12)
CENTRO	SARIMA (0,1,1)(0,1,1,12)
SUR	SARIMA (1,1,1)(0,1,1,12)

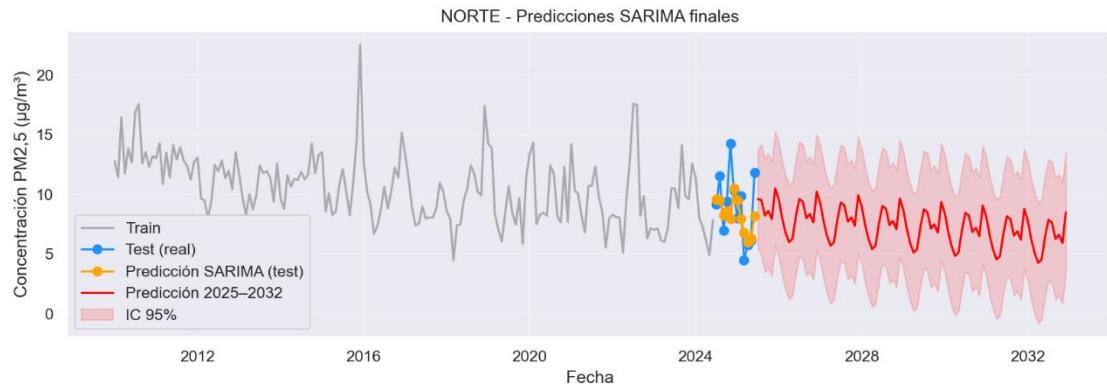


Figura 137 - PM<sub>2.5</sub> - SARIMA óptimo NORTE

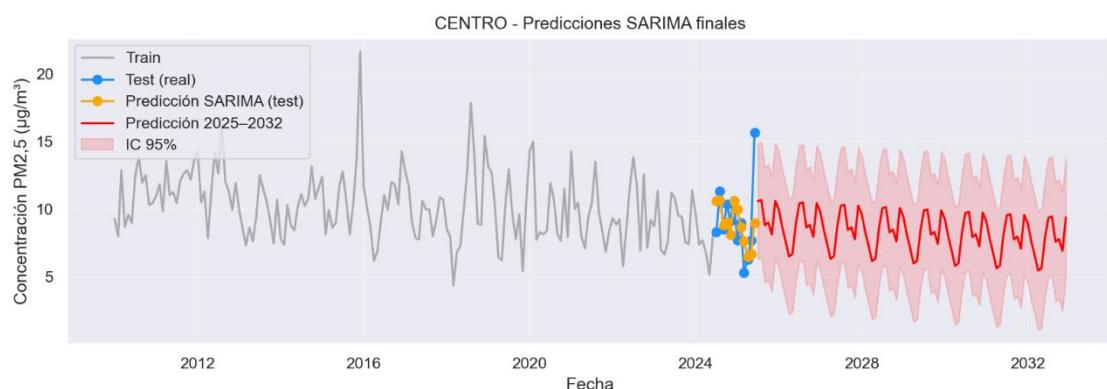


Figura 138 - PM<sub>2.5</sub> - SARIMA óptimo CENTRO

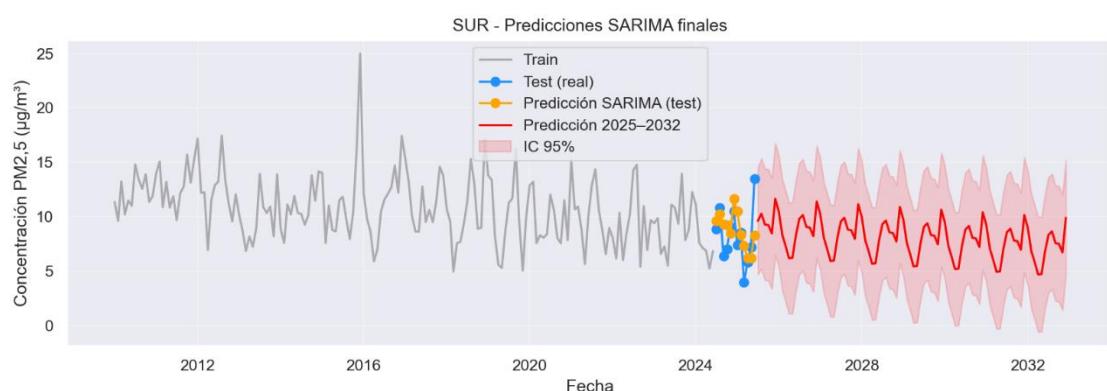


Figura 139 . PM<sub>2.5</sub> - SARIMA óptimo SUR

Se comparan las tres zonas en una única gráfica anual y se saca también la evolución mensual del año 2030 para ver en qué niveles nos encontramos con cada una de las magnitudes:

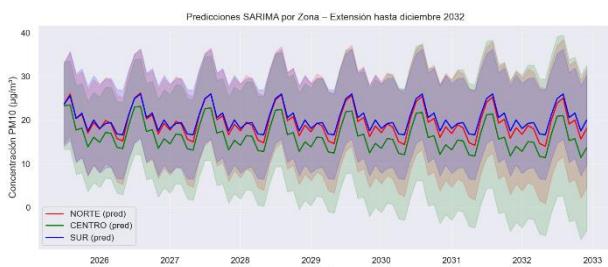


Figura 140 - PM<sub>10</sub> - Gráfico mensual - Pred 2032

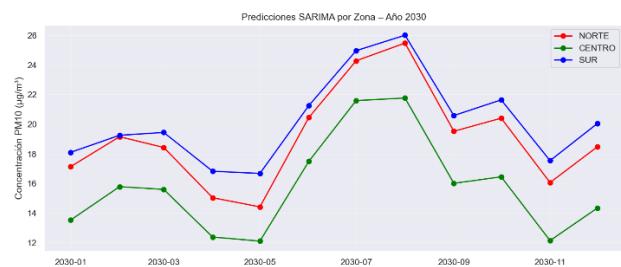


Figura 144 - PM<sub>10</sub> - Gráfico mensual - 2030

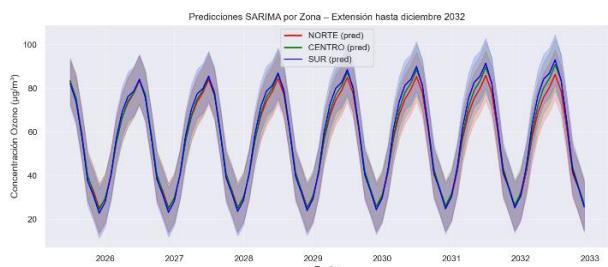


Figura 141 - O<sub>3</sub> - Gráfico mensual - Pred 2032

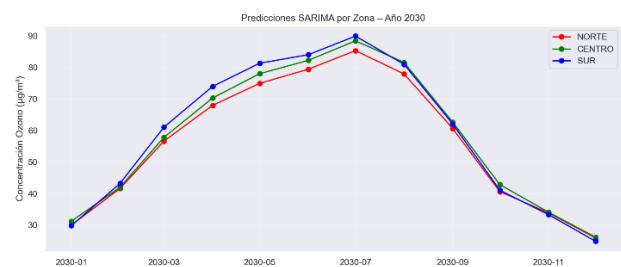


Figura 145 - O<sub>3</sub> - Gráfico mensual - 2030

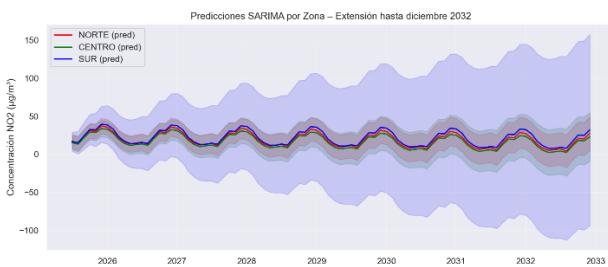


Figura 142 - NO<sub>2</sub> - Gráfico mensual - Pred 2032

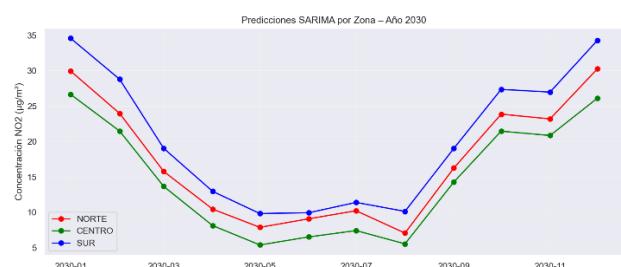


Figura 146 - NO<sub>2</sub> - Gráfico mensual - 2030

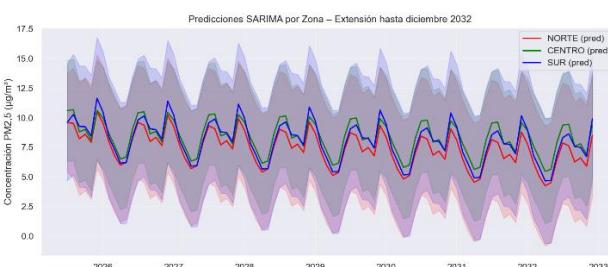


Figura 143 - PM<sub>2.5</sub> - Gráfico mensual - Pred 2032

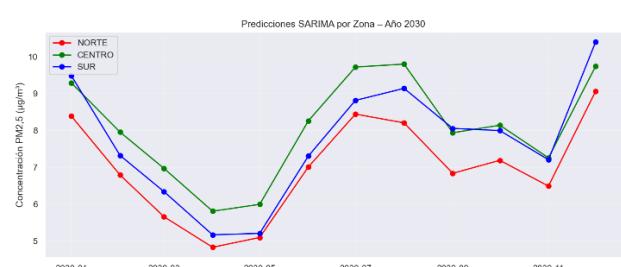


Figura 147 - PM<sub>2.5</sub> - Gráfico mensual - 2030

Con los resultados obtenidos mediante los modelos SARIMA (modelos UNIVARIANTES), se procede a evaluar si la evolución prevista de las concentraciones de cada combinación MAGNITUD-ZONA indica un avance adecuado hacia el cumplimiento de los límites que establecerán los organismos europeos, en consonancia con las recomendaciones de la OMS sobre calidad del aire (valores expresados en µg/m<sup>3</sup>).

ZONA	MAGNITUD	Recomendación OMS	Predicción 2030 SARIMA	CUMPLIMIENTO
NORTE	PM <sub>10</sub>	15	19,073	NO
CENTRO	PM <sub>10</sub>	15	15,768	NO

SUR	PM <sub>10</sub>	15	20,197	<b>NO</b>
NORTE	O <sub>3</sub>	60	56,263	<b>SÍ</b>
CENTRO	O <sub>3</sub>	60	58,135	<b>SÍ</b>
SUR	O <sub>3</sub>	60	58,825	<b>SÍ</b>
NORTE	NO <sub>2</sub>	10	17,325	<b>NO</b>
CENTRO	NO <sub>2</sub>	10	14,787	<b>NO</b>
SUR	NO <sub>2</sub>	10	20,357	<b>NO</b>
NORTE	PM <sub>2.5</sub>	5	7,004	<b>NO</b>
CENTRO	PM <sub>2.5</sub>	5	8,076	<b>NO</b>
SUR	PM <sub>2.5</sub>	5	7,707	<b>NO</b>

Tabla 25 - Cumplimiento recomendación OMS año 2030

Los resultados obtenidos tras la aplicación de modelos univariantes (SARIMA) evidencian un patrón claro y consistente respecto al grado de cumplimiento de las recomendaciones de la OMS para el año 2030. Únicamente **el ozono (O<sub>3</sub>) presenta valores previstos por debajo del umbral establecido en las tres zonas analizadas**, lo que refleja una evolución favorable, a pesar de que, como se observó en los datos históricos, su tendencia era ascendente. En contraste, **las proyecciones para PM<sub>10</sub>, PM<sub>2.5</sub> y NO<sub>2</sub> se sitúan sistemáticamente por encima de los valores recomendados**, lo que confirma que estos contaminantes mantienen niveles elevados. En conjunto, los resultados apuntan a que, salvo en el caso del ozono, el horizonte 2030 no se ajusta aún a los estándares de la OMS, lo que subraya la necesidad de reforzar las políticas de reducción de emisiones y actuar de manera más decidida sobre las principales fuentes emisoras.

Se observan ahora los resultados previstos, empleando modelos univariantes, para 2030 con los obtenidos para finales de 2025, a fin de analizar la evolución proyectada durante ese intervalo de cinco años.

ZONA	MAGNITUD	Valores finales 2025	Predicción 2030 SARIMA	% DIFF
NORTE	PM <sub>10</sub>	18,53	19,073	<b>2,93%</b>
CENTRO	PM <sub>10</sub>	18,08	15,768	<b>-12,79%</b>
SUR	PM <sub>10</sub>	18,18	20,197	<b>11,09%</b>
NORTE	O <sub>3</sub>	57,5	56,263	<b>-2,15%</b>
CENTRO	O <sub>3</sub>	57,96	58,135	<b>0,30%</b>
SUR	O <sub>3</sub>	55,61	58,825	<b>5,78%</b>
NORTE	NO <sub>2</sub>	24,75	17,325	<b>-30,00%</b>
CENTRO	NO <sub>2</sub>	22,57	14,787	<b>-34,48%</b>

SUR	$\text{NO}_2$	25,38	20,357	-19,79%
NORTE	$\text{PM}_{2.5}$	8,56	7,004	-18,18%
CENTRO	$\text{PM}_{2.5}$	8,74	8,076	-7,60%
SUR	$\text{PM}_{2.5}$	8,41	7,707	-8,36%

Tabla 26 - Resultados SARIMA - Comparación valores anuales 2025

Los resultados muestran una evolución heterogénea entre 2025 y las proyecciones para 2030 obtenidas mediante el modelo univariante SARIMA, con mejoras significativas en algunos contaminantes y estancamiento o retrocesos en otros. En el caso de  $\text{NO}_2$ , las **tres zonas presentan reducciones muy acusadas** (entre -19,8 % y -34,5 %), lo que indica una **tendencia descendente clara y coherente con la progresiva electrificación del transporte y las restricciones a los vehículos más emisores**. También  $\text{PM}_{2.5}$  muestra descensos en todas las zonas (-7,6 % a -18,2 %), una señal positiva dada su elevada peligrosidad sanitaria y su fuerte componente de emisiones difusas. Por el contrario,  $\text{PM}_{10}$  presenta un **comportamiento dispar**: mientras que la zona Centro proyecta una reducción notable (-12,8 %), Norte y Sur. **En cuanto al  $\text{O}_3$ , las variaciones son muy pequeñas y con signo mixto**, lo que confirma la complejidad de su formación fotoquímica y la dificultad de modificar su comportamiento mediante reducciones directas de precursores.

## CONCLUSIÓN

En conjunto, los resultados obtenidos mediante los modelos univariantes SARIMA ofrecen una visión clara del estado actual y de la evolución prevista de la calidad del aire en Madrid. Por un lado, el análisis frente a los valores recomendados por la OMS para 2030 muestra que únicamente el ozono ( $\text{O}_3$ ) se sitúa dentro de los límites en todas las zonas, mientras que  $\text{PM}_{10}$ ,  $\text{PM}_{2.5}$  y  $\text{NO}_2$  continúan claramente por encima de los niveles exigidos, lo que refleja que el cumplimiento de los objetivos aún está lejos de alcanzarse. Por otro lado, la comparación entre los valores de 2025 y las proyecciones para 2030 revela dinámicas diferenciadas por contaminante: fuertes descensos en  $\text{NO}_2$  y reducciones algo más moderadas en  $\text{PM}_{2.5}$ , incrementos o comportamientos heterogéneos en  $\text{PM}_{10}$  y variaciones mínimas en  $\text{O}_3$ . Esta doble perspectiva —comparación con los estándares y evolución temporal proyectada— evidencia que, aunque se observan avances relevantes en algunos contaminantes clave, **persisten desafíos significativos que requieren intervenciones más específicas y focalizadas**. En definitiva, los **modelos univariantes apuntan a mejoras parciales, pero también ponen de manifiesto la necesidad de reforzar las estrategias de mitigación para progresar de forma sostenida hacia los objetivos de calidad del aire marcados para 2030**.

Pasamos ahora a analizar los modelos multivariantes, esto, teniendo en cuenta variables exógenas.

## 7.3. BLOQUE 2/FASE 2 – MODELOS MULTIVARIANTES

### i. MODELOS MACHINE LEARNIG (ML)

Para cada combinación MAGNITUD-ZONA, se entrenaron **tres modelos de machine learning: CatBoost, LightGBM y XGBoost**, configurados con un número elevado de iteraciones y una tasa de aprendizaje moderada con el fin de optimizar su

capacidad predictiva sin incurrir en sobreajuste. El proceso de entrenamiento se realizó sobre las particiones temporales previamente definidas (train-test), manteniendo la estructura cronológica inherente a la serie. Al igual que para los modelos univariantes, la muestra test corresponde al rango temporal entre el 01 de julio de 2024 al 30 de junio de 2025. Se muestran a continuación las principales diferencias entre los tres modelos:

CARACTERÍSTICA	CatBoost	LightGBM	XGBoost
<b>Manejo de categóricas</b>	Sin codificación previa	Requiere codificación (e.g. one-hot)	Requiere codificación
<b>Velocidad de entrenamiento</b>	Media	Muy rápida	Rápida
<b>Precisión</b>	Alta, especialmente en datos mixtos	Muy alta en grandes datasets	Alta y robusta
<b>Overfitting</b>	Menor riesgo por regularización	Puede sobreajustar si no se controla	Controlable con parámetros
<b>Rendimiento con pocos datos históricos</b>	Muy buena estabilidad	Puede perder estabilidad con datasets pequeños	Desempeño sólido, aunque puede requerir más datos para expresar todo su potencial predictivo

Tabla 27 - Comparativa modelos Machine Learning

En lo relativo a la inclusión de variables exógenas, se observa que dos de los tres modelos de ML presentan limitaciones en el tratamiento de variables categóricas. Por este motivo, para cada magnitud **se opta por emplear exclusivamente las variables exógenas de tipo numérico**, seleccionadas entre aquellas que mostraron relevancia según los análisis de correlación e importancia de variables realizados previamente (véase Tabla 19). Por tanto, para cada magnitud, se usarán las siguientes variables como variables exógenas:

MAGNITUD	Variables exógenas para ML
PM <sub>10</sub>	<ul style="list-style-type: none"> <li>• Temperatura: máxima, mínima y media (°C).</li> <li>• Presión: presión media (hPa).</li> <li>• Viento a 100 m: velocidad máxima, mínima y media (km/h) y dirección (°).</li> <li>• Viento a 10 m: velocidad máxima, mínima y media (km/h) y dirección (°).</li> <li>• Nieve: espesor de nieve acumulada (cm).</li> </ul>
O <sub>3</sub>	<ul style="list-style-type: none"> <li>• Temperatura: máxima, mínima y media (°C).</li> <li>• Presión: presión media (hPa).</li> <li>• Viento: velocidad máxima a 100 m (km/h) y dirección a 10 m (°).</li> <li>• Radiación solar: radiación global (W/m<sup>2</sup>).</li> <li>• Humedad relativa: valores máximo, mínimo y medio (%).</li> </ul>
NO <sub>2</sub>	<ul style="list-style-type: none"> <li>• Viento (10 m): velocidad máxima, mínima y media (km/h).</li> <li>• Viento (100 m): velocidad máxima, mínima y media (km/h).</li> <li>• Dirección del viento: dirección a 10 m (°).</li> <li>• Presión atmosférica: valores máximo, mínimo y medio (hPa).</li> <li>• Temperatura: temperatura media (°C).</li> </ul>
PM <sub>2.5</sub>	<ul style="list-style-type: none"> <li>• Viento (10 m): velocidad máxima, mínima y media (km/h).</li> <li>• Viento (100 m): velocidad máxima, mínima y media (km/h).</li> <li>• Presión atmosférica: valores máximo, mínimo y medio (hPa).</li> <li>• Superficie incendiada: hectáreas de zonas cercanas afectadas por incendios (ha).</li> </ul>

Tabla 28 - Variables exógenas para ML

Se muestran a continuación los resultados RMSE obtenidos tras aplicar los modelos de ML sobre cada una de las combinaciones MAGNITUD-ZONA:

**PM<sub>10</sub>**

ZONA	MODELO	RMSE
NORTE	CatBoost	<b>4,292</b>
CENTRO	CatBoost	<b>3,755</b>
SUR	CatBoost	4,562
NORTE	LightGBM	4,453
CENTRO	LightGBM	3,881
SUR	LightGBM	4,885
NORTE	XGBoost	4,638
CENTRO	XGBoost	3,996
<b>SUR</b>	<b>XGBoost</b>	<b>4,336</b>

Tabla 29 - PM<sub>10</sub> - Resultados RMSE ML

**O<sub>3</sub>**

ZONA	MODELO	RMSE
NORTE	CatBoost	5,731
CENTRO	CatBoost	<b>7,417</b>
SUR	CatBoost	6,230
NORTE	LightGBM	<b>4,629</b>
CENTRO	LightGBM	7,577
<b>SUR</b>	<b>LightGBM</b>	<b>5,576</b>
NORTE	XGBoost	5,694
CENTRO	XGBoost	8,459
SUR	XGBoost	6,718

Tabla 30 - O<sub>3</sub> - Resultados RMSE ML

**NO<sub>2</sub>**

ZONA	MODELO	RMSE
NORTE	CatBoost	2,837
CENTRO	CatBoost	3,370
SUR	CatBoost	3,489
NORTE	LightGBM	3,877
CENTRO	LightGBM	3,666
SUR	LightGBM	4,145
<b>NORTE</b>	<b>XGBoost</b>	<b>2,237</b>
<b>CENTRO</b>	<b>XGBoost</b>	<b>2,783</b>
<b>SUR</b>	<b>XGBoost</b>	<b>2,992</b>

Tabla 31 - NO<sub>2</sub> - Resultados RMSE ML

**PM<sub>2.5</sub>**

ZONA	MODELO	RMSE
NORTE	CatBoost	<b>1.895</b>
CENTRO	CatBoost	1.696
SUR	CatBoost	2.188
NORTE	LightGBM	2.100
CENTRO	LightGBM	<b>1.587</b>
SUR	LightGBM	<b>2.072</b>
NORTE	XGBoost	2.199
CENTRO	XGBoost	1.724
SUR	XGBoost	2.608

Tabla 32 - PM<sub>2.5</sub> - Resultados RMSE ML

Por tanto, las predicciones de los modelos con menor RMSE por cada combinación MAGNITUD-ZONA se grafican como sigue:

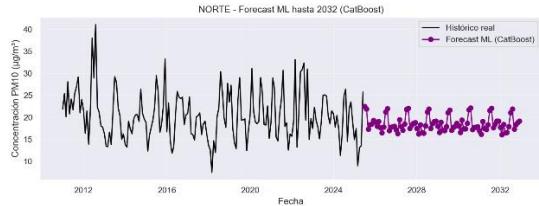


Figura 148 - PM<sub>10</sub> - Pred 2032 - CatBoost - NORTE

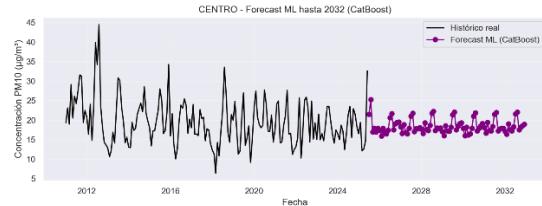


Figura 149 - PM<sub>10</sub> - Pred 2032 - CatBoost - CENTRO

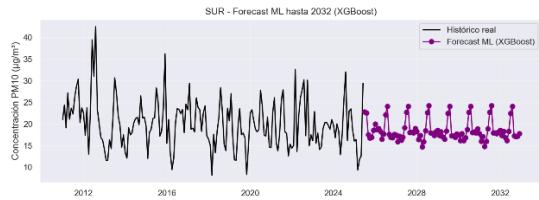


Figura 150 - PM<sub>10</sub> - Pred 2032 - XGBoost - SUR

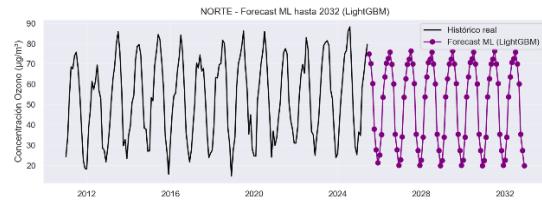


Figura 151 - O<sub>3</sub> - Pred 2032 - LightGBM - NORTE

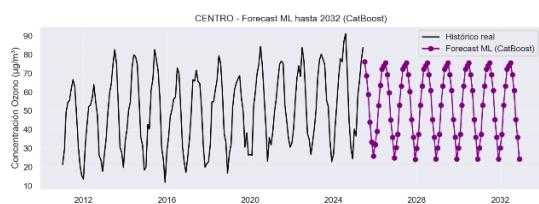


Figura 152 - O<sub>3</sub> - Pred 2032 - CatBoost - CENTRO

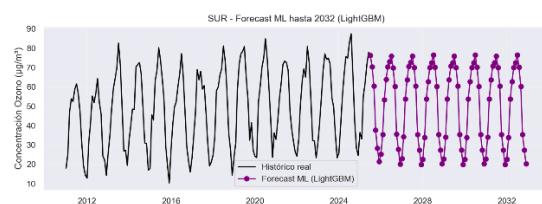


Figura 153 - O<sub>3</sub> - Pred 2032 - LightGBM - SUR

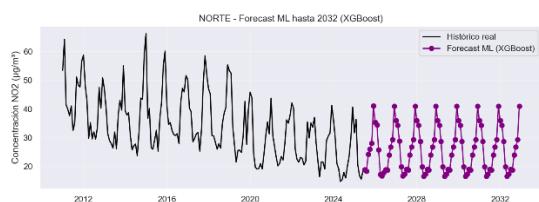


Figura 154 - NO<sub>2</sub> - Pred 2032 - XGBoost - NORTE

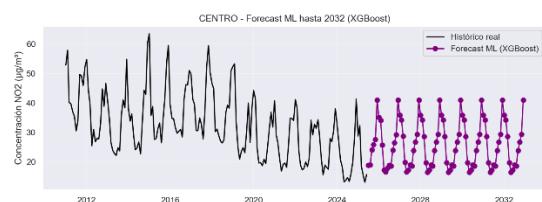


Figura 155 - NO<sub>2</sub> - Pred 2032 - XGBoost - CENTRO

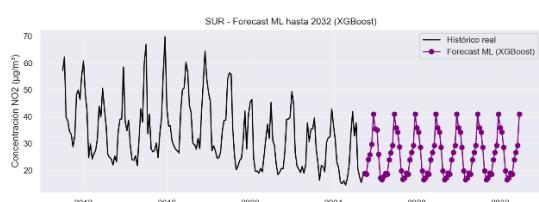


Figura 156 - NO<sub>2</sub> - Pred 2032 - XGBoost - SUR

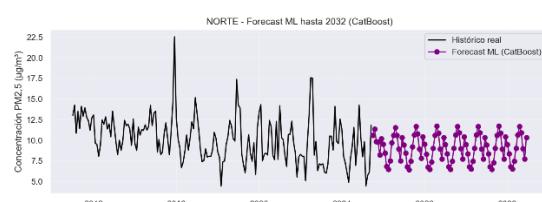


Figura 157 - PM<sub>2.5</sub> - Pred 2032 - CatBoost - NORTE

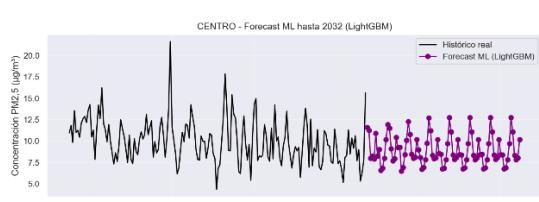


Figura 158 - PM<sub>2.5</sub> - Pred 2032 - LightGBM - CENTRO

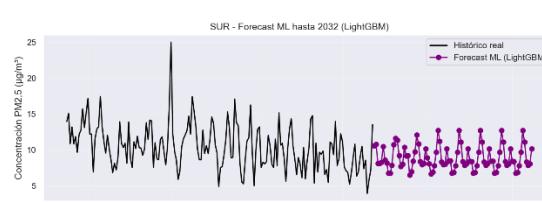


Figura 159 - PM<sub>2.5</sub> - Pred 2032 - LightGBM - SUR

## ii. MODELOS DEEP LEARNIG (DL)

Para cada combinación MAGNITUD–ZONA se entrenaron **dos modelos de DL diseñados específicamente para el tratamiento de series temporales: una red LSTM (Long Short-Term Memory) y una CNN1D (Convolutional Neural Network unidimensional)**. Ambas arquitecturas trabajan sobre secuencias multivariantes construidas mediante una ventana deslizante de 12 meses, permitiendo capturar tanto dependencias de largo alcance como patrones locales en la dinámica temporal del contaminante. El entrenamiento se llevó a cabo sobre las particiones temporales previamente definidas (train–test), preservando la estructura cronológica de los datos y utilizando early stopping para evitar fenómenos de sobreajuste. Al igual que en los modelos univariantes y de ML, el conjunto de prueba corresponde al periodo comprendido entre el 1 de julio de 2024 y el 30 de junio de 2025, lo que garantiza la comparabilidad entre enfoques. A continuación, se muestran las principales diferencias entre ambas arquitecturas:

CARACTERÍSTICA	LSTM	CNN1D
Manejo de categóricas	No admite categóricas directamente; requiere convertirlas a valores numéricos o embeddings <sup>4</sup>	No admite categóricas directamente; requiere codificación previa o embeddings
Velocidad de entrenamiento	Más lenta, debido al procesamiento secuencial y celdas de memoria	Más rápida, gracias a operaciones convolucionales paralelizables
Precisión	Alta cuando existen dependencias temporales de largo alcance	Muy buena para capturar patrones locales y estructuras repetitivas
Overfitting	Propensa al sobreajuste si no se regula; requiere early stopping y regularización	Más robusta, pero también puede sobreajustar con arquitecturas demasiado profundas
Rendimiento con pocos datos históricos	Mejor en series largas; puede perder rendimiento con muestras reducidas	Comportamiento estable incluso con menos datos; detecta patrones locales de forma eficiente

Tabla 33 - Comparativa modelos Deep Learning

En lo relativo a la inclusión de variables exógenas, se observa que ambos modelos no admiten la incorporación directa de variables categóricas. Por ello, se descarta la variable “dow”, dado que carece de sentido en una estructura mensual, y la variable “mes” se transforma en un vector numérico. Además, se integran las variables numéricas utilizadas en los modelos de ML y se añaden las variables derivadas “target”, “t\_index”, “roll3\_mean”, “roll6\_mean” y “roll12\_mean”. El significado de estas últimas se resume en la siguiente tabla.

<sup>4</sup> Es una forma de convertir categorías o palabras en vectores numéricos continuos que capturan relaciones y similitudes entre ellas.

VARIABLE	SIGNIFICADO
target	Valor del contaminante en el instante anterior. Actúa como componente autoregresiva dentro de la secuencia temporal
t_index	Índice temporal creciente (1, 2, 3, ...), útil para modelizar tendencias de largo plazo en la serie.
roll3_mean	Media móvil de 3 períodos. Suaviza el comportamiento de corto plazo y reduce el ruido puntual de la serie
Roll6_mean	Media móvil de 6 períodos. Captura tendencias intermedias entre corto y medio plazo
Roll12_mean	Media móvil de 12 períodos. Representa la estacionalidad y la tendencia estructural anual

Tabla 34 - Variables derivadas modelos DL

Tras la ejecución de estos modelos, los resultados obtenidos no muestran un comportamiento plenamente coherente, probablemente debido a fenómenos de sobreajuste. La excepción la constituye la magnitud  $O_3$ , especialmente en las zonas **NORTE** y **CENTRO**, donde los modelos ofrecen un rendimiento más consistente. A continuación, se presentan los resultados correspondientes a dichas zonas:

ZONA	MODELO	RMSE
<b>NORTE</b>	<b>LSTM</b>	<b>11,25</b>
CENTRO	LSTM	10.365
NORTE	CNN1D	19,591
<b>CENTRO</b>	<b>CNN1D</b>	<b>6.548</b>

Tabla 35 -  $O_3$  - Resultados RMSE DL

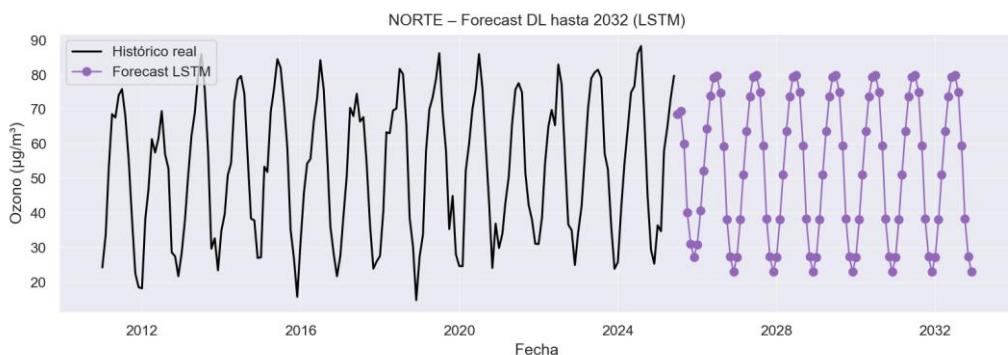


Figura 160 -  $O_3$  - Pred 2032 - LSTM - NORTE

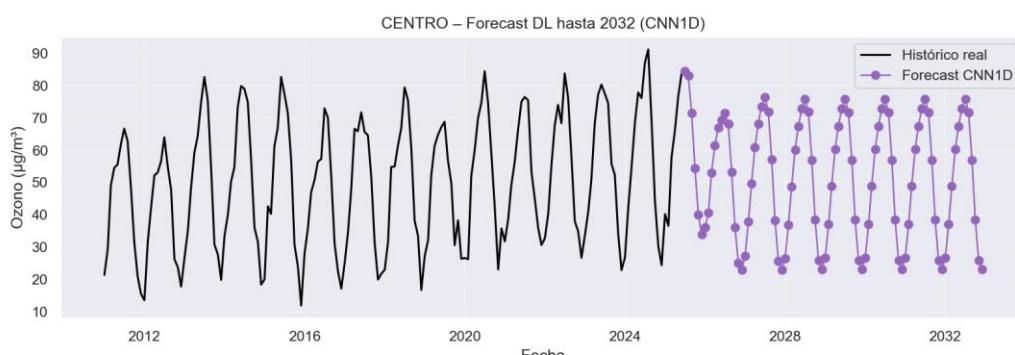


Figura 161 -  $O_3$  - Pred 2032 - CNN1D - CENTRO

## 7.4. BLOQUE 2/FASE 3 Evaluación por RMSE y Selección del mejor modelo

Una vez ejecutados todos los modelos, se evalúan sus valores de RMSE y se establece un ranking de desempeño para cada combinación MAGNITUD-ZONA, seleccionando como mejor modelo aquel que presenta el menor RMSE.

**PM<sub>10</sub>**

modelo	zona	RMSE_test
CatBoost	CENTRO	3.755
LightGBM	CENTRO	3.881
XGBoost	CENTRO	3.996
SARIMA	CENTRO	5.137
LSTM	CENTRO	5.845
ARIMA	CENTRO	6.525
CNN1D	CENTRO	22.226
CatBoost	NORTE	4.292
SARIMA	NORTE	4.360
LightGBM	NORTE	4.453
XGBoost	NORTE	4.638
LSTM	NORTE	5.598
CNN1D	NORTE	5.736
ARIMA	NORTE	6.049
XGBoost	SUR	4.336
CatBoost	SUR	4.562
LightGBM	SUR	4.885
SARIMA	SUR	5.207
ARIMA	SUR	7.006
CNN1D	SUR	7.493
LSTM	SUR	18.584

Figura 162 - PM<sub>10</sub> - Ranking mejores modelos por zona

**O<sub>3</sub>**

modelo	zona	RMSE_test
CNN1D	CENTRO	6.548
SARIMA	CENTRO	6.784
CatBoost	CENTRO	7.417
LightGBM	CENTRO	7.577
XGBoost	CENTRO	8.459
LSTM	CENTRO	10.365
ARIMA	CENTRO	20.669
LightGBM	NORTE	4.629
SARIMA	NORTE	5.641
XGBoost	NORTE	5.694
CatBoost	NORTE	5.731
LSTM	NORTE	11.250
ARIMA	NORTE	19.028
CNN1D	NORTE	19.591
LightGBM	SUR	5.576
CatBoost	SUR	6.230
SARIMA	SUR	6.454
XGBoost	SUR	6.718
ARIMA	SUR	18.894
LSTM	SUR	19.091
CNN1D	SUR	20.094

Figura 163 - O<sub>3</sub> - Ranking mejores modelos por zona

**NO<sub>2</sub>**

modelo	zona	RMSE_test
XGBoost	CENTRO	2.783
CatBoost	CENTRO	3.370
LightGBM	CENTRO	3.666
SARIMA	CENTRO	4.405
ARIMA	CENTRO	11.596
CNN1D	CENTRO	13.604
LSTM	CENTRO	22.338
XGBoost	NORTE	2.237
CatBoost	NORTE	2.837
SARIMA	NORTE	3.674
CNN1D	NORTE	3.790
LightGBM	NORTE	3.877
LSTM	NORTE	8.581
ARIMA	NORTE	11.929
XGBoost	SUR	2.992
CatBoost	SUR	3.489
LightGBM	SUR	4.145
SARIMA	SUR	4.624
ARIMA	SUR	12.907
LSTM	SUR	24.617
CNN1D	SUR	33.655

Figura 164 - NO<sub>2</sub> - Ranking mejores modelos por zona

**PM<sub>2.5</sub>**

modelo	zona	RMSE_test
LightGBM	CENTRO	1.587
CatBoost	CENTRO	1.696
XGBoost	CENTRO	1.724
SARIMA	CENTRO	2.328
ARIMA	CENTRO	2.595
LSTM	CENTRO	2.670
CNN1D	CENTRO	9.413
CatBoost	NORTE	1.895
LightGBM	NORTE	2.100
XGBoost	NORTE	2.199
SARIMA	NORTE	2.446
CNN1D	NORTE	2.718
ARIMA	NORTE	2.760
LSTM	NORTE	2.776
LightGBM	SUR	2.072
CatBoost	SUR	2.188
SARIMA	SUR	2.335
XGBoost	SUR	2.608
ARIMA	SUR	2.666
CNN1D	SUR	6.118
LSTM	SUR	7.478

Figura 165 - PM<sub>2.5</sub> - Ranking mejores modelos por zona

Por tanto, tras terminar de ejecutar este segundo bloque del proceso definido en la Figura 121, los mejores modelos por cada combinación MAGNITUD-ZONA y sus predicciones promedio anuales para 2030 son los siguientes:

**PM<sub>10</sub>**

zona	modelo	promedio_anual_2030
CENTRO	CatBoost	18.219
NORTE	CatBoost	18.504
SUR	XGBoost	18.583

Figura 166 - PM<sub>10</sub> - Predicciones anuales 2030 mejores modelos

**O<sub>3</sub>**

zona	modelo	promedio_anual_2030
CENTRO	CNN1D	50.443
NORTE	LightGBM	50.547
SUR	LightGBM	50.422

Figura 167 - O<sub>3</sub> - Predicciones anuales 2030 mejores modelos

**NO<sub>2</sub>**

zona	modelo	promedio_anual_2030
CENTRO	XGBoost	25.918
NORTE	XGBoost	25.918
SUR	XGBoost	25.918

Figura 168 - PM<sub>10</sub> - Predicciones anuales 2030 mejores modelos

**PM<sub>2,5</sub>**

zona	modelo	promedio_anual_2030
CENTRO	LightGBM	8.861
NORTE	CatBoost	8.994
SUR	LightGBM	8.861

Figura 169 - PM<sub>2,5</sub> - Predicciones anuales 2030 mejores modelos

## 7.5. BLOQUE 3/FASE 1 - MODELIZACIÓN REFINADA

Se observa que las predicciones anuales para 2030 correspondientes a las tres zonas de la magnitud NO<sub>2</sub> coinciden exactamente. Esto se explica porque, en este tramo avanzado del horizonte de forecast, los modelos dependen casi por completo de las variables exógenas, mientras que la información específica de cada serie histórica pierde peso. Dado que dichas exógenas (solo meteorológicas) son idénticas para NORTE, CENTRO y SUR, los modelos multivariantes generan trayectorias futuras muy similares. Además, al tratarse de un procedimiento iterativo, los lags autoregresivos, que en las primeras iteraciones aún incorporan información real de cada zona, pasan a estar formados exclusivamente por predicciones previas del propio modelo. Esto provoca que la señal diferenciada de cada serie se diluya progresivamente, haciendo que la dinámica del forecast converja hacia un patrón “climatológico” común determinado por esas exógenas compartidas. Como consecuencia, los valores previstos para 2030 resultan finalmente idénticos entre las tres zonas.

Dado lo expuesto anteriormente y considerando la fuerte influencia del tráfico en los niveles de NO<sub>2</sub>, y, de forma indirecta, en el comportamiento del O<sub>3</sub>, **se procede a reentrenar los modelos incorporando la variable TRAFICO y restringiendo el análisis a datos posteriores al 1 de enero de 2019**, con el fin de garantizar coherencia temporal y homogeneidad metodológica. Asimismo, se incluye CatBoost en esta segunda fase para todas las zonas, independientemente de su desempeño previo, debido a su eficacia con muestras pequeñas y su capacidad para modelizar relaciones no lineales sin necesidad de preprocesar variables categóricas.

Tras ejecutar todos los modelos con los datos históricos filtrados e incluyendo la variable tráfico entre las variables exógenas, obtengo los siguientes gráficos que pueden resultar de interés:

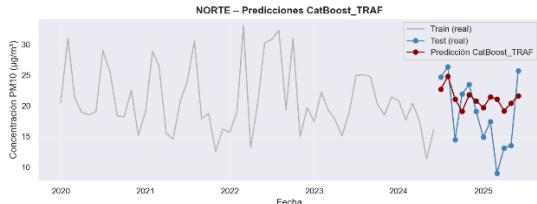


Figura 170 - PM<sub>10</sub> - Error Test (+Tráfico) - CatBoost - NORTE

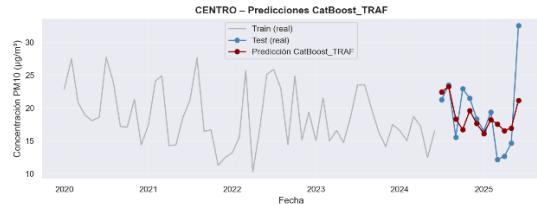


Figura 171 - PM<sub>10</sub> - Error Test (+Tráfico) - CatBoost - CENTRO



Figura 172 - PM<sub>10</sub> - Error Test (+Tráfico) - CatBoost - SUR

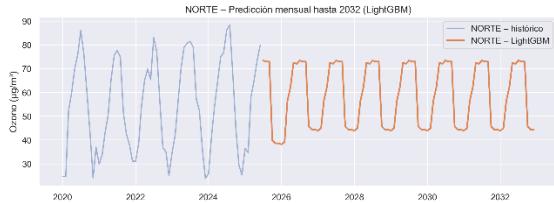


Figura 173 - O<sub>3</sub> - pred 2032 (+Tráfico) - LightGBM - NORTE

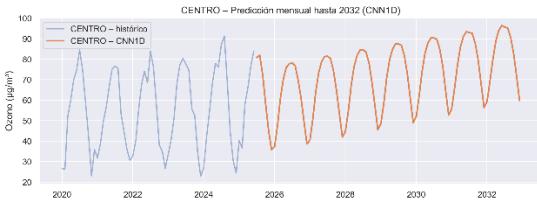


Figura 174 - O<sub>3</sub> - pred 2032 (+Tráfico) - CNN1D - CENTRO



Figura 175 - O<sub>3</sub> - pred 2032 (+Tráfico) - LightGBM - SUR



Figura 176 - NO<sub>2</sub> - Error Test (+Tráfico) - XGBoost - NORTE



Figura 177 - NO<sub>2</sub> - Error Test (+Tráfico) - XGBoost - CENTRO



Figura 178 - NO<sub>2</sub> - Error Test (+Tráfico) - XGBoost - SUR

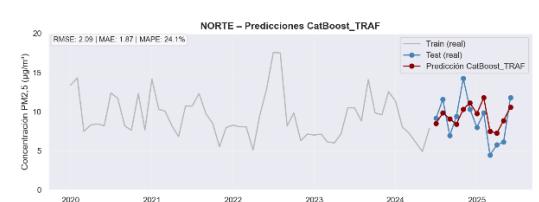


Figura 179 - PM<sub>2.5</sub> - Error Test (+Tráfico) - CatBoost - NORTE



Figura 180 - PM<sub>2.5</sub> - Error Test (+Tráfico) - CatBoost - CENTRO



Figura 181 - PM<sub>2.5</sub> - Error Test (+Tráfico) - CatBoost - SUR

## 7.6. BLOQUE 3/FASE 2 - NUEVA EVALUACIÓN RMSE

Una vez, se ha incorporado la variable exógena tráfico y se han ejecutados los modelos susceptibles a mejorar a los modelos del bloque dos, se vuelve a mostrar el ranking de los modelos según su error RMSE:

**PM<sub>10</sub>**

modelo	zona	RMSE_test
CatBoost	CENTRO	3.755
LightGBM	CENTRO	3.881
XGBoost	CENTRO	3.996
CatBoost_TRAF	CENTRO	4.410
SARIMA	CENTRO	5.137
LSTM	CENTRO	5.845
ARIMA	CENTRO	6.525
CNN1D	CENTRO	22.226
CatBoost	NORTE	4.292
SARIMA	NORTE	4.360
LightGBM	NORTE	4.453
XGBoost	NORTE	4.638
CatBoost_TRAF	NORTE	5.412
LSTM	NORTE	5.598
CNN1D	NORTE	5.736
ARIMA	NORTE	6.049
XGBoost	SUR	4.336
CatBoost	SUR	4.562
LightGBM	SUR	4.885
SARIMA	SUR	5.207
CatBoost_TRAF	SUR	5.621
XGBoost_TRAF	SUR	5.973
ARIMA	SUR	7.006
CNN1D	SUR	7.493
LSTM	SUR	18.584

Figura 182 - PM<sub>10</sub> - Ranking mejores modelos por zona (+Tráfico)

**O<sub>3</sub>**

modelo	zona	RMSE_test
CNN1D_TRAF	CENTRO	5.174
CNN1D	CENTRO	6.548
SARIMA	CENTRO	6.784
CatBoost	CENTRO	7.417
LightGBM	CENTRO	7.577
XGBoost	CENTRO	8.459
CatBoost_TRAF	CENTRO	8.910
LSTM	CENTRO	10.365
ARIMA	CENTRO	20.669
LightGBM	NORTE	4.629
SARIMA	NORTE	5.641
XGBoost	NORTE	5.694
CatBoost	NORTE	5.731
CatBoost_TRAF	NORTE	7.240
LightGBM_TRAF	NORTE	8.568
LSTM	NORTE	11.250
ARIMA	NORTE	19.028
CNN1D	NORTE	19.591
LightGBM	SUR	5.576
CatBoost	SUR	6.230
SARIMA	SUR	6.454
XGBoost	SUR	6.718
LightGBM_TRAF	SUR	7.239
CatBoost_TRAF	SUR	7.375
ARIMA	SUR	18.894
LSTM	SUR	19.091
CNN1D	SUR	20.094

Figura 183 - O<sub>3</sub> - Ranking mejores modelos por zona (+Tráfico)

**NO<sub>2</sub>**

modelo	zona	RMSE_test
XGBoost	CENTRO	2.783
XGBoost_TRAF	CENTRO	2.836
CatBoost	CENTRO	3.370
LightGBM	CENTRO	3.666
CatBoost_TRAF	CENTRO	4.118
SARIMA	CENTRO	4.405
ARIMA	CENTRO	11.596
CNN1D	CENTRO	13.604
LSTM	CENTRO	22.338
XGBoost_TRAF	NORTE	2.200
XGBoost	NORTE	2.237
CatBoost	NORTE	2.837
CatBoost_TRAF	NORTE	3.189
SARIMA	NORTE	3.674
CNN1D	NORTE	3.790
LightGBM	NORTE	3.877
LSTM	NORTE	8.581
ARIMA	NORTE	11.929
XGBoost_TRAF	SUR	2.443
XGBoost	SUR	2.992
CatBoost	SUR	3.489
CatBoost_TRAF	SUR	4.065
LightGBM	SUR	4.145
SARIMA	SUR	4.624
ARIMA	SUR	12.907
LSTM	SUR	24.617
CNN1D	SUR	33.655

Figura 184 - NO<sub>2</sub> - Ranking mejores modelos por zona (+Tráfico)

**PM<sub>2.5</sub>**

modelo	zona	RMSE_test
LightGBM	CENTRO	1.587
CatBoost	CENTRO	1.696
XGBoost	CENTRO	1.724
CatBoost_TRAF	CENTRO	2.008
SARIMA	CENTRO	2.328
LightGBM_TRAF	CENTRO	2.352
ARIMA	CENTRO	2.595
CNN1D	CENTRO	9.413
CatBoost	NORTE	1.895
CatBoost_TRAF	NORTE	2.089
LightGBM	NORTE	2.100
XGBoost	NORTE	2.199
SARIMA	NORTE	2.446
CNN1D	NORTE	2.718
ARIMA	NORTE	2.760
LSTM	NORTE	2.776
CatBoost_TRAF	SUR	1.974
LightGBM	SUR	2.072
CatBoost	SUR	2.188
SARIMA	SUR	2.335
XGBoost	SUR	2.608
ARIMA	SUR	2.666
LightGBM_TRAF	SUR	2.735
CNN1D	SUR	6.118
LSTM	SUR	7.478

Figura 185 - PM<sub>2.5</sub> - Ranking mejores modelos por zona (+Tráfico)

Por tanto, tras terminar de ejecutar este tercer bloque del proceso definido en la Figura 121, los mejores modelos por cada combinación MAGNITUD-ZONA y sus predicciones promedio anuales para 2030 son los siguientes:

### **PM<sub>10</sub>**

zona	modelo	promedio_anual_2030
CENTRO	CatBoost	18.219
NORTE	CatBoost	18.504
SUR	XGBoost	18.583

Figura 186 - PM<sub>10</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico)

### **O<sub>3</sub>**

zona	modelo	promedio_anual_2030
CENTRO	CNN1D_TRAF	75.610
NORTE	LightGBM	50.547
SUR	LightGBM	50.422

Figura 187 - O<sub>3</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico)

### **NO<sub>2</sub>**

zona	modelo	promedio_anual_2030
CENTRO	XGBoost	25.918
NORTE	XGBoost_TRAF	25.920
SUR	XGBoost_TRAF	26.416

Figura 188 - NO<sub>2</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico)

### **PM<sub>2.5</sub>**

zona	modelo	promedio_anual_2030
CENTRO	LightGBM	8.861
NORTE	CatBoost	8.994
SUR	CatBoost_TRAF	9.254

Figura 189 - PM<sub>2.5</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico)

Dados los resultados anteriores y antes de proceder con las predicciones de los valores de las magnitudes, observamos que el modelo de red convolucional CNN1D\_TRAF presenta un valor promedio anual para el año 2030 muy por encima de los otros modelos (véase Figura 190), por lo que se decide, hacer una comparativa con los otros dos modelos que le persiguen en el ranking, estos son: CNN1D y SARIMA, de modo que:

modelo	zona	promedio_anual_2030
ARIMA	CENTRO	50.202
CNN1D	CENTRO	50.443
CatBoost	CENTRO	53.493
SARIMA	CENTRO	58.135
CNN1D_TRAF	CENTRO	75.610

Figura 190 - O<sub>3</sub> - Predicciones anuales 2030 (+Tráfico) - CENTRO

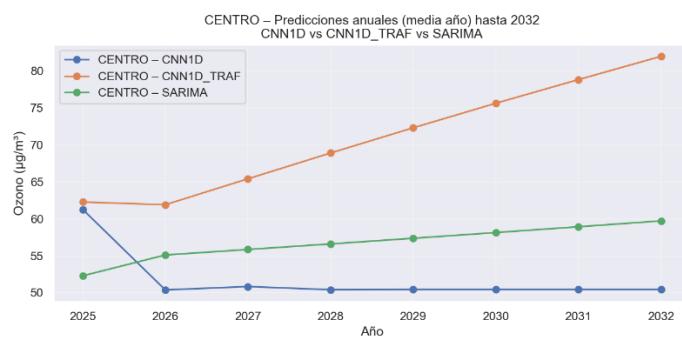


Figura 191 - O<sub>3</sub> - Predicciones anuales - CENTRO - CNN1D\_TRAF vs CNN1D vs SARIMA

Tras revisar la Figura 191 y compararla con los valores históricos mostrados en la Figura 106, se concluye que:

1. El modelo CNN1D queda descartado, dado que no presenta ningún tipo de tendencia.
2. Entre los modelos restantes, el que mejor captura la tendencia ascendente observada en los datos históricos es el modelo **SARIMA**. Vease tablas siguientes:

HISTÓRICO	2018	50
	2024	57,5
	TENDENCIA ANUAL <sup>5</sup>	<b>2,50%</b>

Tabla 36 - % Tendencia anual - O3 - CENTRO - HISTÓRICO

CNN1D_TRAF	2026	62,5
	2032	84
	TENDENCIA ANUAL	<b>5,73%</b>

Tabla 37 - % Tendencia anual - O3 - CENTRO - CNN1D\_TRAF

SARIMA	2026	55
	2032	60
	TENDENCIA ANUAL	<b>1,52%</b>

Tabla 38 - % Tendencia anual - O3 - CENTRO - SARIMA

Tomando como definitivo el modelo SARIMA, obtengo que los valores medios de la magnitud O<sub>3</sub> para 2030 en las 3 zonas son los siguientes:

modelo	zona	promedio_anual_2030
SARIMA	CENTRO	58.135
LightGBM	NORTE	50.547
LightGBM	SUR	50.422

Figura 192 - O<sub>3</sub> - Predicciones anuales 2030 mejores modelos (+Tráfico) - Definitivo

## 7.7. BLOQUE 3/FASE 3 – PREDICCIONES FINALES

A continuación, se listan todos los modelos seleccionados para cada combinación magnitud-zona y la diferencia de sus predicciones para 2030 con las obtenidas de los históricos en 2025:

ZONA	MAGNITUD	MODELO SELECCIONADO	Valores finales 2025	Predicción 2030	% DIFF
NORTE	PM <sub>10</sub>	CatBoost	18,53	18,504	<b>-0,14%</b>
CENTRO	PM <sub>10</sub>	CatBoost	18,08	18,219	<b>0,77%</b>
SUR	PM <sub>10</sub>	XGBoost	18,18	18,583	<b>2,22%</b>
NORTE	O <sub>3</sub>	LightGBM	57,5	50,547	<b>-12,09%</b>
CENTRO	O <sub>3</sub>	SARIMA	57,96	58,135	<b>0,30%</b>
SUR	O <sub>3</sub>	LightGBM	55,61	50,422	<b>-9,33%</b>
NORTE	NO <sub>2</sub>	XGBoost_TRAF	24,75	25,920	<b>4,73%</b>
CENTRO	NO <sub>2</sub>	XGBoost	22,57	25,918	<b>14,83%</b>
SUR	NO <sub>2</sub>	XGBoost_TRAF	25,38	26,416	<b>4,08%</b>
NORTE	PM <sub>2,5</sub>	CatBoost	8,56	8,994	<b>5,07%</b>
CENTRO	PM <sub>2,5</sub>	LightGBM	8,74	8,861	<b>1,38%</b>
SUR	PM <sub>2,5</sub>	CatBoost_TRAF	8,41	9,266	<b>10,18%</b>

Tabla 39 - Resumen modelos menor RMSE

<sup>5</sup> Se calcula como el % de incremento de la medición entre el valor de 2024 y el de 2018, dividido entre los años de diferencia entre 2024 y 2018 (6)

Se observa que, aunque los modelos seleccionados presentan el menor RMSE dentro de los ensayos realizados para O<sub>3</sub>, NO<sub>2</sub> y PM<sub>2.5</sub>, las trayectorias generadas no reproducen adecuadamente la tendencia observada en los datos históricos. En consecuencia, se decide sustituir estos modelos mediante la adopción de enfoques SARIMA, más adecuados para capturar la estructura temporal de las series. Por tanto, se presentan los modelos finales

ZONA	MAGNITUD	MODELO FINAL	TENDENCIA DATOS HISTÓRICOS	Valores finales 2025	Predicción 2030 FINAL	% DIFF
NORTE	PM <sub>10</sub>	CatBoost	Descenso hasta 2014/Estabilización posterior	18,53	18,504	-0,14%
CENTRO	PM <sub>10</sub>	CatBoost	Descenso hasta 2017/Estabilización posterior	18,08	18,219	0,77%
SUR	PM <sub>10</sub>	XGBoost	Descenso hasta 2014/Estabilización posterior	18,18	18,583	2,22%
NORTE	O <sub>3</sub>	SARIMA	Ligero ascenso sostenido	57,5	56,263	-2,15%
CENTRO	O <sub>3</sub>	SARIMA	Ligero ascenso sostenido	57,96	58,135	0,30%
SUR	O <sub>3</sub>	SARIMA	Ligero ascenso sostenido	55,61	58,825	5,78%
NORTE	NO <sub>2</sub>	SARIMA	Descenso sostenido	24,75	17,325	-30,00%
CENTRO	NO <sub>2</sub>	SARIMA	Descenso sostenido	22,57	14,787	-34,48%
SUR	NO <sub>2</sub>	SARIMA	Descenso sostenido	25,38	20,357	-19,79%
NORTE	PM <sub>2.5</sub>	SARIMA	Clara tendencia descendente	8,56	7,004	-18,18%
CENTRO	PM <sub>2.5</sub>	SARIMA	Clara tendencia descendente	8,74	8,076	-7,60%
SUR	PM <sub>2.5</sub>	SARIMA	Clara tendencia descendente	8,41	7,707	-8,36%

Tabla 40 - Resumen Mejores modelos

Por tanto, resultando los modelos los indicados en la Tabla 40, se muestran las predicciones hasta 2032:

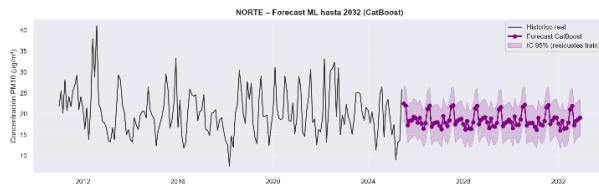


Figura 193 -  $\text{PM}_{10}$  - pred 2032 FINAL - CatBoost - NORTE

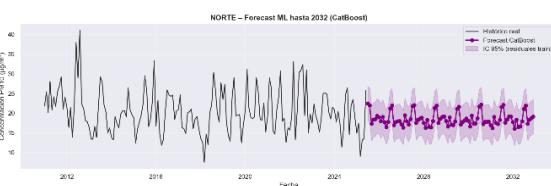


Figura 194 -  $\text{PM}_{10}$  - pred 2032 FINAL - CatBoost - CENTRO

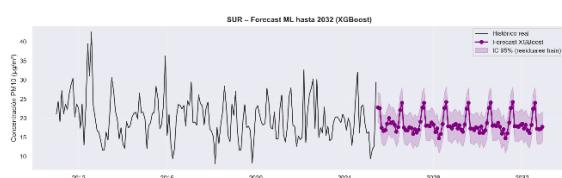


Figura 195 -  $\text{PM}_{10}$  - pred 2032 FINAL - XGBoost - SUR

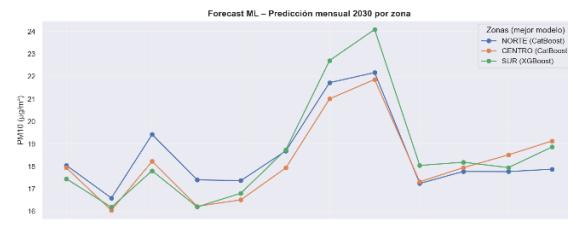


Figura 196 -  $\text{PM}_{10}$  - Comparación zonas pred 2030 FINAL

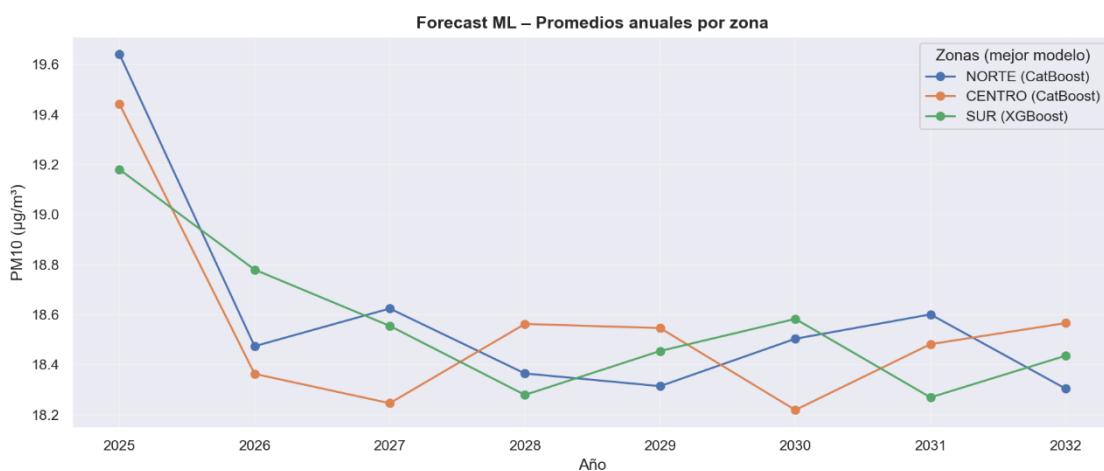


Figura 197 -  $\text{PM}_{10}$  - Comparación zonas pred 2032 FINAL

### $\text{O}_3$

- Pred 2032 FINAL - NORTE: véase Figura 131
- Pred 2032 FINAL - CENTRO: véase Figura 132
- Pred 2032 FINAL - SUR: véase Figura 133
- Comparación zonas pred 2030 FINAL: véase Figura 145
- Comparación zonas pred 2032 FINAL: véase Figura 141

### $\text{NO}_2$

- Pred 2032 FINAL - NORTE: véase Figura 134
- Pred 2032 FINAL - CENTRO: véase Figura 135
- Pred 2032 FINAL - SUR: véase Figura 136
- Comparación zonas pred 2030 FINAL: véase Figura 146
- Comparación zonas pred 2032 FINAL: véase Figura 142

### $\text{PM}_{2.5}$

- Pred 2032 FINAL - NORTE: véase Figura 137

- Pred 2032 FINAL - CENTRO: véase Figura 138
- Pred 2032 FINAL - SUR: véase Figura 139
- Comparación zonas pred 2030 FINAL: véase Figura 147
- Comparación zonas pred 2032 FINAL: véase Figura 143

## CONCLUSIÓN

Los resultados obtenidos permiten anticipar con un grado elevado de confianza que la evolución de la calidad del aire en Madrid mantiene la inercia observada en los últimos quince años. Para los contaminantes de origen primario, NO<sub>2</sub> y PM<sub>2.5</sub>, las series temporales apuntan a descensos adicionales de aquí a 2030, coherentes con la electrificación progresiva del parque móvil, la renovación tecnológica y la consolidación de políticas de baja emisión. Además, como se ha puesto de manifiesto en la Tabla 39, ambas magnitudes presentan una alta dependencia del tráfico rodado, lo que confirma que la combustión asociada a los vehículos constituye una de las fuentes principales de su variabilidad. Este vínculo directo explica que las medidas orientadas a la movilidad urbana tengan un impacto tan significativo en la reducción proyectada de estos contaminantes.

En contraste, otros contaminantes como, el PM<sub>10</sub> se mantiene en torno a valores estables, con variaciones menores y sin riesgos aparentes de repunte. Por otro lado, el O<sub>3</sub> confirma su perfil de aumento suave pero persistente, especialmente en la zona Sur, donde se anticipa un incremento cercano al 6 %. Esta dinámica refleja la complejidad fotoquímica del ozono y la necesidad de abordar estrategias más integrales que incluyan **no solo la reducción de precursores urbanos, sino también enfoques de ámbito regional**. En conjunto, las proyecciones evidencian avances significativos, pero también áreas donde los objetivos de sostenibilidad requerirán intervenciones más ambiciosas si se pretende cumplir plenamente con los estándares de la Agenda 2030.

## 8. ESTUDIOS ADICIONALES BASADOS EN LAS CONCLUSIONES

En este último punto, se van a estudiar las predicciones de las zonas más desfavorables de cada una de las magnitudes.

El código referente a este apartado se encuentra recogido dentro del archivo *08\_Estudios Adicionales.ipynb*, incluido íntegramente dentro del **Anexo A4** del presente documento.

### 8.1. PM<sub>10</sub> - SUR

Se realiza un análisis específico de la predicción de la magnitud PM<sub>10</sub> en la zona SUR con el objetivo de evaluar tres aspectos fundamentales de la evolución prevista hasta 2032:

1. Identificar si, dentro del horizonte de predicción, existen meses en los que la concentración prevista se sitúe por debajo del límite anual establecido para el año 2030 (15 µg/m<sup>3</sup>);
2. Determinar en qué trimestre se registran los valores más elevados de concentración del contaminante; y
3. Llevar a cabo un análisis de la variabilidad de la serie predicha para caracterizar su dispersión y estabilidad temporal.

En cuanto al punto primero, se obtiene el siguiente gráfico:

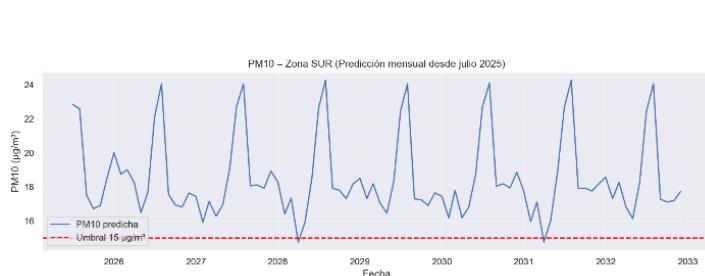


Figura 198 -  $PM_{10}$  - SUR - Valor límite OMS

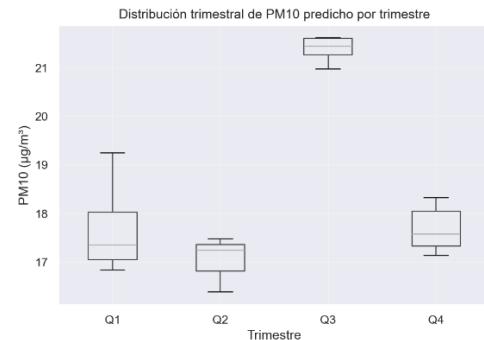


Figura 199 -  $PM_{10}$  - SUR - Boxplot trimestres

```
===== VARIABILIDAD DE LA SERIE (Mensual y Trimestral) =====
media_mensual: 18.526
std_mensual: 2.407
cv_mensual: 0.130
iqr_mensual: 1.632
media_trimestral: 18.526
std_trimestral: 1.853
cv_trimestral: 0.100
iqr_trimestral: 3.295
```

Tabla 41 -  $PM_{10}$  - SUR - Variabilidad

En conjunto, el análisis evidencia que solo en un número muy reducido de meses las concentraciones previstas de  $PM_{10}$  se sitúan por debajo del valor límite anual fijado para 2030 ( $15 \mu g/m^3$ ), lo que indica **una baja probabilidad de cumplimiento normativo bajo las tendencias actuales**. Asimismo, se observa que **los valores máximos de concentración se concentran sistemáticamente en el tercer trimestre del año**, periodo en el que las administraciones deberían intensificar las medidas de mitigación debido a la mayor probabilidad de superar niveles críticos. En términos de variabilidad, la serie predicha presenta una dispersión moderada a nivel mensual, con una desviación estándar de  $2.407 \mu g/m^3$  e índice de variación relativo ( $CV = 0.130$ ). A escala trimestral, la variabilidad disminuye aún más ( $std = 1.853; CV = 0.100$ ), lo que refleja una serie más suavizada cuando se agregan las observaciones, si bien el IQR trimestral (3.295) confirma la existencia de fluctuaciones estructurales relevantes entre estaciones. Estas evidencias subrayan la necesidad de **actuaciones sostenidas y focalizadas, especialmente en el tercer trimestre del año**, para aproximar la concentración de  $PM_{10}$  a los niveles exigidos, como por ejemplo, la limitación del tráfico más contaminante, control de emisiones urbanas e industriales, gestión del polvo en superficie, monitorización intrusión del polvo sahariano y acciones de sensibilización ciudadana, especialmente durante los meses del tercer trimestre del año.

## 8.2. $NO_2$ y $O_3$ - SUR

Se realiza un análisis específico de la predicción de la magnitud  $NO_2$  en la zona SUR con el objetivo de evaluar dos aspectos fundamentales de la evolución prevista hasta 2032:

1. Identificar si, dentro del horizonte de predicción, existen meses en los que la concentración prevista se sitúe por debajo del límite anual establecido para el año 2030 ( $10 \mu g/m^3$ );
2. Identificar la correlación indirecta que existe con el contaminante  $O_3$

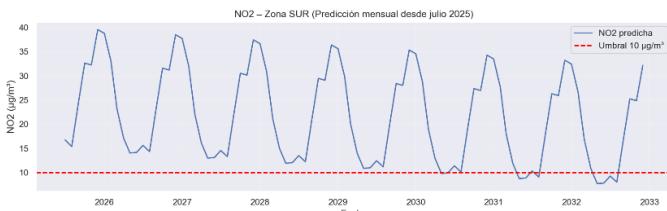


Figura 200 - NO<sub>2</sub> - SUR - Valor límite OMS

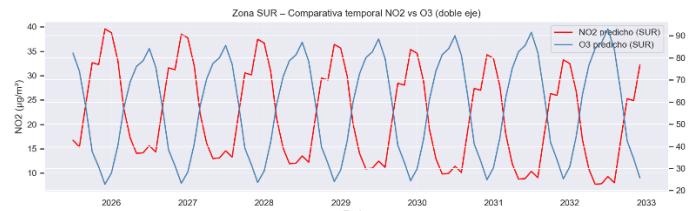


Figura 201 - Comparativa magnitudes NO<sub>2</sub> y O<sub>3</sub>

La serie de NO<sub>2</sub> permanece consistentemente por encima del umbral de 10 µg/m<sup>3</sup>, pese a las caídas estivales. Esto implica que la zona SUR no alcanzará niveles especialmente bajos de NO<sub>2</sub> en el corto y medio plazo, por lo que será necesario mantener medidas de control de emisiones.

En paralelo, las proyecciones para la zona SUR muestran una estacionalidad muy marcada tanto en NO<sub>2</sub> como en O<sub>3</sub>, con ciclos anuales que se mantienen estables durante todo el horizonte 2025–2033. **Ambas series evolucionan de forma inversa: el ozono alcanza sus máximos en los meses cálidos cuando los niveles de NO<sub>2</sub> desciden, reflejando el comportamiento fotoquímico típico en el que la reducción de NOx y la mayor radiación solar favorecen la formación de O<sub>3</sub>.**

La combinación de ambos resultados sugiere que la **gestión ambiental deberá abordar NO<sub>2</sub> y O<sub>3</sub> de forma integrada**: reducciones de NOx pueden ayudar a aliviar la carga urbana de NO<sub>2</sub>, pero deben planificarse cuidadosamente para no intensificar episodios de ozono durante el verano.

Por otro lado, se analiza la predicción a largo plazo de la magnitud O<sub>3</sub> y se comprueba si, en la zona más desfavorable, que nuevamente corresponde a la SUR, existen meses en los que se supera el umbral anual recomendado por la OMS (60 µg/m<sup>3</sup>).



Figura 202 - O<sub>3</sub> - SUR - Valor límite OMS

Como puede observarse, numerosos meses superan este límite, especialmente durante el verano, lo que indica que gran parte de las medidas destinadas a mitigar las concentraciones de O<sub>3</sub> deberían centrarse en el periodo estival.

Comentado lo de arriba, se proponen algunas medidas reducir el NO<sub>2</sub> sin que ello suponga un aumento considerable, sobre todo en los meses de verano:

- Reducir simultáneamente NOx y COV, especialmente en sectores tráfico-industria, para evitar incrementos indeseados de ozono.
- Renovar flotas y electrificar movilidad.
- Controlar emisiones de COV industriales y disolventes.
- Aplicar restricciones temporales en verano para actividades emisoras en días de riesgo alto de formación de ozono.

- Actuar sobre calefacción residencial, sustituyendo calderas contaminantes y mejorando la eficiencia energética.

### 8.3. PM<sub>2.5</sub> – CENTRO

Se realiza un análisis específico de la predicción de la magnitud PM<sub>2.5</sub> en la zona CENTRO con el objetivo de evaluar dos aspectos fundamentales de la evolución prevista hasta 2032:

1. Identificar si, dentro del horizonte de predicción, existen meses en los que la concentración prevista se sitúe por debajo del límite anual establecido para el año 2030 ( $5 \mu\text{g}/\text{m}^3$ );
2. Determinar en qué trimestre se registran los valores más elevados de concentración del contaminante; y



Figura 203 - PM<sub>2.5</sub> - SUR - Valor límite OMS

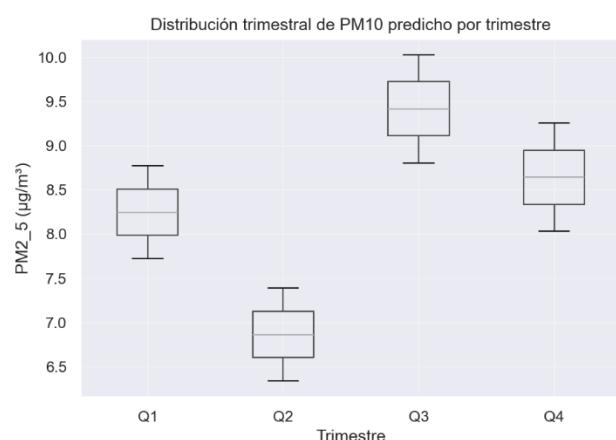


Figura 204 - PM<sub>2.5</sub> - SUR - Boxplot trimestres

Las predicciones de PM<sub>2.5</sub> muestran un patrón estacional muy definido, con máximos persistentes en los meses de invierno y valores más bajos en verano. Aun así, las concentraciones se mantienen siempre por encima del umbral recomendado por la OMS ( $5 \mu\text{g}/\text{m}^3$ ) durante todo el periodo analizado. El análisis trimestral confirma esta dinámica: Q1 y Q4 presentan los niveles más altos, mientras que Q3 registra los mínimos, reflejando la fuerte influencia de la meteorología fría y de las emisiones asociadas al tráfico y a la calefacción residencial.

Para mitigar la concentración de PM<sub>2.5</sub>, las actuaciones más eficaces deben centrarse en los meses fríos e incluir: reducción del tráfico rodado, sustitución de calderas contaminantes y mejora de la eficiencia energética en edificios, control de emisiones industriales y polvo urbano, y activación de protocolos de calidad del aire en episodios de estabilidad atmosférica.

## 9. BIBLIOGRAFÍA

ID	Título	Autor/Institución	Año	Tipo de fuente	Link
01	Transforming our world: the 2030 Agenda for Sustainable Development	Naciones Unidas	2015	Documento institucional	<a href="#">Link</a>
02	Directiva 2008/50/CE	Parlamento Europeo y Consejo de la Unión Europea	2008	Legislación europea	<a href="#">Link</a>
03	Ley 34/2007	Gobierno de España	2007	Legislación nacional	<a href="#">Link</a>
04	Real Decreto 102/2011	Gobierno de España	2011	Legislación nacional	<a href="#">Link</a>
05	Real Decreto 34/2023	Gobierno de España	2023	Legislación nacional	<a href="#">Link</a>
06	Compromiso de reducción del 90 % de las emisiones de gases de efecto invernadero para 2040 respecto a 1990	El País	2025	Noticia/Política ambiental	<a href="#">Link</a>
07	Guías de Calidad del Aire: Partículas en suspensión ( $PM_{2.5}$ y $PM_{10}$ ), Ozono ( $O_3$ ), Dióxido de nitrógeno ( $NO_2$ ), Dióxido de azufre ( $SO_2$ ) y Monóxido de carbono (CO)	Organización Mundial de la Salud (OMS)	2021	Documento técnico / Referencia científica internacional	<a href="#">Link</a>
08	PM2.5 and cardiovascular diseases: State-of-the-Art review	International Journal of Cardiology Cardiovascular Risk and Prevention	2023	Artículo científico	<a href="#">Link</a>
09	Efectos del material particulado (PM) sobre la salud y el medioambiente	Agencia de Protección Ambiental de Estados Unidos	2025	Informe técnico de divulgación científica.	<a href="#">Link</a>
10	Contaminación atmosférica efectos en la salud respiratoria en el niño	Revista Médica Clínica Las Condes	2017	Artículo científico	<a href="#">Link</a>
11	Efectos del ozono troposférico en la salud humana en el	Agencia Europea de Medio Ambiente (EEA) y la Comisión Europea	2023	Informe técnico de divulgación científica	<a href="#">Link</a>

ID	Título	Autor/Institución	Año	Tipo de fuente	Link
	contexto del cambio climático				
12	Dióxido de Azufre - Efectos en la Salud	Centro Canadiense de Seguridad y Salud Ocupacional (CCSSO)	1997	Informe técnico de divulgación científica	<a href="#">Link</a>
13	Planes nacionales de mejora de la calidad del aire	Gobierno de España	2023	Documento técnico de planificación ambiental	<a href="#">Link</a>
14	Plan de Calidad del Aire y Cambio Climático de la Comunidad de Madrid - Orden 2322/2023	Comunidad Autónoma de Madrid	2024	Legislación autonómica	<a href="#">Link</a>
15	Plan de Calidad del Aire y Cambio Climático "Madrid 360"	Ayuntamiento de Madrid	2019	Plan municipal de acción climática	<a href="#">Link</a>
16	Portal de Datos Abiertos del Ayuntamiento de Madrid	Ayuntamiento de Madrid	2025	Base de datos – Calidad del aire	<a href="#">Link</a>
17	Portal Meteoblue – Historical Weather Data	Meteoblue	2025	Base de datos – Meteorológicos	<a href="#">Link</a>
18	Portal de Datos Abiertos del Ayuntamiento de Madrid	Ayuntamiento de Madrid	2025	Base de datos – Tráfico	<a href="#">Link</a>
19	Portal de datos abiertos de Civio	Civio	2023	Base de datos – Incendios	<a href="#">Link</a>
20	Información sanitaria y ambiental sobre el Tolueno	MurciaSalud	(s.f.)	Divulgación técnica / Información ambiental	<a href="#">Link</a>
21	Ficha toxicológica del Etilbenceno (ATSDR)	ATSDR (Agency for Toxic Substances and Disease Registry)	2016	Documento técnico / Toxicología ambiental	<a href="#">Link</a>