

Viabilidad e implicaciones de la Inteligencia Artificial para la estimación de la calidad del aire en España

Viability and implications of Artificial Intelligence for air quality estimation in Spain

Carlos Morillas¹, Rocco Giosa², Sergio Álvarez¹, Sara Martínez¹

REVISTA **MAPPING**
Vol. 35, 221, 86-92
2026
ISSN: 1131-9100

Resumen

Este estudio analiza el potencial de la inteligencia artificial (IA) para mejorar la estimación de la calidad del aire en España, con énfasis en el dióxido de nitrógeno (NO₂), uno de los contaminantes más vinculados a la movilidad urbana y a efectos adversos sobre la salud. La investigación se centra en la Comunidad de Madrid, que dispone de la red de vigilancia más completa del país y constituye un entorno idóneo para evaluar el rendimiento de modelos basados en aprendizaje automático. Se integran observaciones del sensor TROPOMI (Sentinel-5P) con variables meteorológicas procedentes tanto de estaciones de superficie como del reanálisis ERA5-Land, empleando diferentes configuraciones espaciales y dos familias de modelos: algoritmos basados en árboles de decisión (Random Forest y XGBoost) y redes neuronales artificiales.

Los mejores resultados se obtuvieron mediante redes neuronales combinadas con clústeres espaciales coherentes, alcanzando precisiones superiores a estudios previos y demostrando la capacidad de la IA para reproducir la variabilidad espacio-temporal del NO₂ incluso en entornos heterogéneos. El trabajo evidencia que, bajo un diseño metodológico adecuado, la IA puede complementar las redes de vigilancia tradicionales, ofrecer estimaciones fiables en áreas con escasa infraestructura y constituir una herramienta estratégica para afrontar las nuevas exigencias normativas y la transición hacia sistemas de monitorización más flexibles y escalables.

Palabras clave: Inteligencia artificial, NO₂, TROPOMI, aprendizaje automático, calidad del aire.

Abstract

This study examines the potential of artificial intelligence (AI) to enhance air-quality estimation in Spain, focusing on nitrogen dioxide (NO₂), a pollutant closely linked to urban mobility and adverse health impacts. The research is conducted in the Community of Madrid, which hosts the country's most comprehensive air-quality monitoring network and provides an ideal setting to assess machine-learning-based modeling. TROPOMI (Sentinel-5P) satellite observations are combined with meteorological variables derived from surface stations and ERA5-Land reanalysis. Several spatial configurations and two model families are evaluated: tree-based algorithms (Random Forest and XGBoost) and artificial neural networks. The best performance is achieved by neural networks trained on environmentally coherent spatial clusters, yielding accuracy metrics that surpass most previous studies and demonstrating AI's ability to capture the complex spatiotemporal behavior of NO₂, even in heterogeneous environments. Results show that, when supported by a rigorous methodological design, AI can effectively complement traditional monitoring networks, provide reliable estimates in areas with limited infrastructure, and support new regulatory requirements aimed at reducing pollutant exposure. The study highlights AI as a strategic tool for improving air-quality assessment in Spain and for facilitating the transition toward more flexible, scalable, and data-integrated monitoring systems.

Keywords: Artificial intelligence, NO₂, TROPOMI, machine learning, air quality.

¹Departamento de Ingeniería y Morfología del Terreno, Universidad Politécnica de Madrid
c.morillas@upm.es

²Dipartimento di Ingegneria, Università degli Studi della Basilicata

Recepción 15/01/2026
Aprobación 20/01/2026

1. INTRODUCCIÓN

La calidad del aire ha supuesto tradicionalmente uno de los principales retos ambientales y sanitarios en España. A pesar de los avances logrados durante la última década, la Organización Mundial de la Salud (OMS) señala que la exposición a contaminantes atmosféricos continúa generando impactos significativos sobre la salud pública, incluso en niveles inferiores a los umbrales regulados. Entre ellos, el dióxido de nitrógeno (NO₂) destaca por su estrecha vinculación con la movilidad urbana y por sus efectos contrastados sobre la morbilidad y la mortalidad. Ciudades como Madrid (en primer lugar), Barcelona (6º) y Valencia (68º) se encuentran entre las que registran mayores tasas de mortalidad evitable asociada a la exposición prolongada al NO₂ en Europa (Khomenko et al., 2021). Estos resultados evidencian no solo la magnitud del desafío, sino también la urgencia de desplegar herramientas capaces de evaluar la contaminación con mayor resolución espacial y temporal, especialmente en contextos urbanos densos donde los gradientes locales pueden cambiar abruptamente en distancias cortas.

Aunque las redes de vigilancia constituyen el pilar fundamental para la evaluación de la calidad del aire, su distribución sigue siendo desigual. Casi un centenar de municipios españoles con más de 50.000 habitantes dispone de una única estación de medición o carece de ella, lo que limita la representatividad espacial y compromete la capacidad de caracterizar adecuadamente la exposición real de la población. Este déficit estructural cobra especial relevancia ante el endurecimiento normativo: la reciente Directiva (UE) 2024/2881 fija un límite anual para NO₂ de 20 µg/m³ en 2030, la mitad del actualmente vigente, mientras que las guías de la OMS recomiendan niveles aún más estricto (OMS, 2021). Las administraciones deberán, por tanto, apoyarse en métodos más flexibles, escalables y compatibles con la diversidad territorial del país para cumplir estos objetivos.

En este escenario, la inteligencia artificial (IA) está transformando la manera de estimar y monitorizar la contaminación atmosférica. Los modelos de aprendizaje automático y profundo permiten integrar fuentes heterogéneas (datos de satélite, variables meteorológicas, indicadores de actividad humana o características del territorio) para inferir concentraciones de contaminantes allí donde las estaciones no existen o resultan insuficientes. La elección del NO₂ como contaminante de referencia no es casual: su detección desde satélite presenta un rendimiento particularmente elevado. El sensor TROPOMI a bordo del satélite Sentinel-5P ofrece observaciones diarias del NO₂ troposférico, que en regiones como la Comunidad de Madrid muestran una correlación especialmente alta con las mediciones en superficie, facilitando la construcción de modelos precisos y reduciendo la incertidumbre asociada al proceso de estimación. Esta combinación de alta sensibilidad

espectral, fuerte dependencia de las fuentes locales y vida media atmosférica relativamente corta convierte al NO₂ en uno de los contaminantes más adecuados para su trazabilidad mediante el uso conjunto de teledetección e IA.

España constituye además un laboratorio privilegiado para explorar el potencial de estas metodologías. Su marcada diversidad climática, la compleja distribución de emisiones, los contrastes topográficos y las diferencias entre áreas urbanas y rurales plantean desafíos que demandan enfoques de modelización capaces de aprender patrones locales sin perder capacidad de generalización. A su vez, la digitalización progresiva de las políticas ambientales y la creciente disponibilidad de datos abiertos facilitan la integración de nuevas soluciones tecnológicas en la gestión de la calidad del aire.

En este artículo se analizan las implicaciones científicas, tecnológicas y regulatorias que tiene el uso de la IA para la estimación de la calidad del aire en España, con especial atención al NO₂ como contaminante trazador. Se presenta el caso de la Comunidad de Madrid, cuyo interés radica en que alberga la red de vigilancia de calidad del aire más completa y consolidada del país. La densidad y continuidad de sus registros permiten emplear las mediciones de sus estaciones como referencia de calibración y validación para entrenar modelos basados en IA, proporcionando un escenario idóneo para evaluar el potencial de estas técnicas y explorar su futura extensión a otras regiones españolas.

2. MATERIALES Y MÉTODOS

2.1 Área de estudio y fuentes de datos

El análisis se centra en la Comunidad de Madrid, que dispone de la red de vigilancia de calidad del aire más completa de España, lo que permite contar con series temporales extensas de NO₂ superficial para la validación de los modelos. Además, la combinación de una topografía compleja, gradientes urbanos pronunciados y una marcada variabilidad meteorológica ofrece un escenario particularmente exigente para evaluar la capacidad de generalización de los algoritmos de machine learning.

Para obtener información de NO₂ continua a nivel espacial, se utilizaron observaciones del sensor TROPOMI a bordo del satélite Sentinel-5P, que proporciona la columna troposférica de NO₂ (moles/m²) diariamente en todo el planeta y con resolución espacial de 5,5 × 3,5 km² (Van Geffen et al., 2019; Veefkind et al., 2012). Para cada día del periodo de estudio se descargaron los productos Sentinel-5P TROPOMI NO₂ OFFL reprocesados en Google Earth Engine, proyectados en una rejilla regular de 0,01° × 0,01° (~1 × 1 km²), y filtrados por calidad (QA > 0,75) para eliminar observaciones afectadas por nubosidad o errores de recuperación (Earth Engine Data Catalog, 2024). La correlación relativamente elevada entre

la señal satelital y las observaciones en superficie (particularmente notable en Madrid, $r=0,78$, Morillas et al., 2024) supone la variable fundamental para estimar la concentración superficial ($\mu\text{g}/\text{m}^3$).

Las medidas de referencia de NO_2 proceden de la red regional de estaciones, cuya selección se realizó siguiendo criterios de compleción temporal y continuidad operativa, tomando 48 estaciones diferentes entre 2020 y 2024 (Ayuntamiento de Madrid, 2025a; Comunidad de Madrid, 2025a). Estas observaciones se emplearon para el entrenamiento, validación y prueba de los modelos predictivos, bajo distintas configuraciones espaciales diseñadas para evaluar cómo la organización territorial de la información afecta al rendimiento de las técnicas de IA.

Dada la demostrada influencia de las condiciones atmosféricas en los niveles de NO_2 —particularmente en regiones como la Comunidad de Madrid, donde la dinámica del contaminante está fuertemente condicionada por la estabilidad atmosférica, la altura de la capa de mezcla, la radiación solar y los patrones de viento— se incorporaron a los modelos diversas variables meteorológicas como predictores fundamentales (Morillas et al., 2025). Estos datos se obtuvieron de la red conjunta de estaciones meteorológicas operadas por la Comunidad de Madrid (2025) y el Ayuntamiento de Madrid (2025), seleccionando únicamente aquellas que permanecieron plenamente operativas durante el periodo 2020–2024, haciendo un total de 50 puntos de muestreo (Figura 1).

Entre los predictores incluidos se encuentran temperatura, humedad relativa, presión atmosférica, precipitación, radiación solar y los componentes vectoriales del viento. Para su integración en los modelos, cada variable se promedió siguiendo dos ventanas temporales complementarias: una ventana corta, que abarca el intervalo de 11:00 a 14:00 UTC coincidiendo con el sobrevuelo de Sentinel-5P, y una ventana larga, que integra las condiciones atmosféricas de las 21 horas previas (de 14:00 del día anterior a 11:00 del día actual). Esta doble agregación permite capturar tanto el estado inmediato de la atmósfera como los procesos acumulativos que influyen

en la evolución del NO_2 en superficie, dotando a los modelos de IA de la información necesaria para reproducir los procesos físico-químicos que regulan su formación, dispersión y persistencia (Cedeno Jimenez et al., 2023).

Además, con el fin de explorar la viabilidad de enfoques basados exclusivamente en datos abiertos y homogéneos a escala nacional, se planteó un segundo conjunto de experimentos en el que las variables meteorológicas no se obtuvieron de estaciones de superficie, sino del reanálisis climático ERA5-Land (Centro Europeo de Predicción a Medio Plazo, ECMWF, Muñoz Sabater, 2019), con información diaria de resolución espacial $0,1^\circ \times 0,1^\circ$ ($\sim 9 \times 9 \text{ km}^2$). Esta ampliación metodológica resulta fundamental para determinar si la IA puede seguir proporcionando estimaciones fiables de la contaminación atmosférica en ausencia de infraestructura local de monitorización meteorológica.

2.2 Configuraciones espaciales y modelos de IA

El diseño experimental contempló diferentes estrategias de organización territorial y diversas arquitecturas de aprendizaje automático con el fin de evaluar cómo la estructura espacial de los datos y la complejidad del modelo influyen en la capacidad para estimar concentraciones superficiales de NO_2 . Para ello se implementaron cuatro configuraciones espaciales complementarias y dos familias principales de modelos de machine learning, aplicadas de manera sistemática tanto a los datos meteorológicos procedentes de estaciones como a los de ERA5.

2.2.1 Configuraciones espaciales

a) Configuración regional completa

En primer lugar, se consideró una configuración regional completa en la que todas las estaciones de medición se integraron en un único dominio homogéneo. Esta elección maximiza el volumen de datos disponible para el entrenamiento, lo cual puede ser beneficioso para modelos que requieren grandes cantidades de información. Sin embargo, también introduce una elevada diversidad espacial, puesto que mezcla estaciones situadas en entornos urbanos densos, áreas suburbanas, regiones industriales y zonas rurales o de montaña. La heterogeneidad resultante obliga al modelo a aprender patrones muy distintos, e incluso potencialmente contradictorios, lo que puede dificultar su capacidad de generalización.

b) Malla regular de $5 \times 5 \text{ km}^2$

Para capturar de manera más explícita la variabilidad espacial, se adoptó también una configuración basada en una rejilla regular de $5 \times 5 \text{ km}$. En este caso, el área de estudio se subdividió en celdas cuadradas y únicamente se consideraron aquellas donde coincidían una estación de NO_2 y una estación meteorológica. Las observaciones dentro de cada celda se promediaron, generando valores representativos del entorno inmediato.

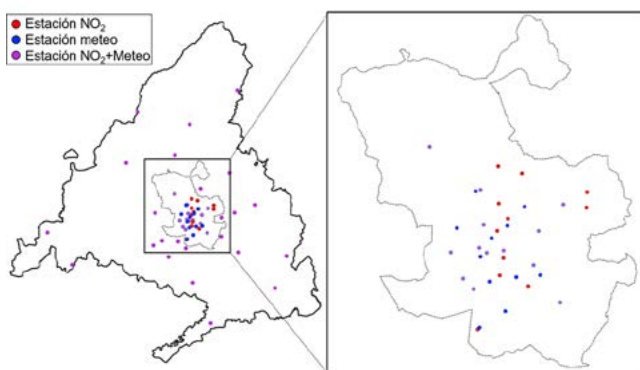


Figura 1. Estaciones de monitorización de NO_2 y de meteorología de la Comunidad de Madrid y, en detalle, del Ayuntamiento de Madrid.

c) Sectores ambientales

Con el fin de lograr un equilibrio entre resolución espacial y coherencia ambiental, se empleó una tercera aproximación basada en grandes sectores ambientales. Estos sectores se definieron atendiendo a características topográficas, urbanísticas y de intensidad de emisiones, agrupando estaciones situadas en entornos con dinámicas atmosféricas similares. De este modo, se evita la fragmentación excesiva propia de la rejilla, a la vez que se minimiza la heterogeneidad interna propia de la agregación regional. La delimitación sectorial permite capturar procesos atmosféricos característicos de cada tipo de entorno —por ejemplo, regiones montañosas del norte, áreas industriales del sur o zonas periurbanas— y proporciona un marco espacial adecuado para analizar cómo varían las relaciones entre las observaciones satelitales y las concentraciones superficiales.

d) Clústeres coherentes simplificados

Finalmente, se desarrolló una configuración basada en clústeres coherentes simplificados, concebida para maximizar la similitud ambiental dentro de cada agrupación. Este enfoque excluye estaciones periféricas o situadas en entornos muy singulares que podrían introducir una variabilidad difícil de integrar mediante un modelo único. El resultado es un conjunto de agrupaciones compactas, definidas no solo por la proximidad geográfica, sino también por la similitud en las dinámicas del NO₂, en los patrones meteorológicos predominantes y en las fuentes locales de emisión. Esto preserva la coherencia ambiental sin reducir en exceso el tamaño de las muestras disponibles.

2.2.2 Modelos de aprendizaje automático empleados

a) Modelos basados en árboles de decisión

Sobre cada una de las configuraciones espaciales descritas se entrenaron dos familias principales de modelos de aprendizaje automático, seleccionadas tanto por su capacidad para manejar relaciones no lineales como por su complementariedad metodológica. La primera familia estuvo formada por modelos basados en árboles de decisión. En una primera etapa se empleó un Random Forest (RF, Breiman, 2001) como herramienta de selección de variables, permitiendo identificar los predictores más relevantes entre el conjunto de variables meteorológicas, temporales y satelitales. Este paso resultó útil para reducir la dimensionalidad del problema, eliminar redundancias y centrar el entrenamiento en aquellos factores con mayor capacidad explicativa. A continuación, las variables seleccionadas alimentaron un modelo XGBoost (Chen and Guestrin, 2016; Friedman, 2001), optimizado mediante técnicas de búsqueda bayesiana de hiperparámetros. Este algoritmo, basado en el prin-

cipio del gradiente descendente aplicado a árboles de decisión secuenciales, ofrece un equilibrio notable entre capacidad predictiva, control del sobreajuste y eficiencia computacional, lo que lo convierte en una herramienta especialmente sólida para problemas con interacciones complejas entre predictores, como los relacionados con la calidad del aire.

b) Redes neuronales artificiales

La segunda familia metodológica estuvo compuesta por redes neuronales artificiales. A diferencia de los modelos basados en árboles, estas redes no requieren selección previa de variables, sino que aprenden internamente las ponderaciones que determinan su relevancia. Para diseñar su arquitectura se empleó AutoKeras, una herramienta de búsqueda automática que permite evaluar múltiples configuraciones sin imponer a priori un diseño concreto. Tras seleccionar la arquitectura óptima en función del rendimiento sobre un conjunto de validación, se aplicaron etapas adicionales de entrenamiento progresivo para refinar los pesos y mejorar la estabilidad del modelo, manteniendo bajo control el riesgo de sobreajuste. Las redes neuronales resultan especialmente adecuadas para entornos complejos y dinámicos, como el de la Comunidad de Madrid, donde los niveles de NO₂ dependen simultáneamente de factores meteorológicos, características del territorio y variaciones temporales que interactúan de forma no lineal.

3. RESULTADOS Y DISCUSIÓN

Entre todos los experimentos realizados, la combinación formada por redes neuronales artificiales y la configuración de clústeres coherentes simplificados fue la que proporcionó los mejores resultados, alcanzando métricas de RMSE (Raíz del Error Cuadrático Medio) = 2.49 µg/m³, R² (Coeficiente de determinación) = 0.93 y MAE (Error Absoluto Medio) = 1.63 µg/m³ (Figura 2). Este nivel de precisión supera de manera consistente al obtenido con el marco híbrido RF/XGBoost en las mismas condiciones, y confirma la capacidad de las ANN para capturar relaciones altamente no lineales entre la señal satelital, los predictores meteorológicos y las características espaciales del territorio.

Este desempeño sitúa el presente estudio entre los más precisos publicados hasta la fecha que integran datos Sentinel-5P para estimar concentraciones superficiales de NO₂. Trabajos previos realizados en Europa, como los de Chan et al. (2021), Balamurugan et al., (2023) o Das et al., (2021), presentan valores de RMSE notablemente superiores (entre 4,7 y 7,2 µg/m³ en la mayoría de los casos) y coeficientes de determinación más bajos (habitualmente entre 0,60 y 0,88). Estudios recientes en China y Estados Unidos, donde se han aplicado arquitecturas

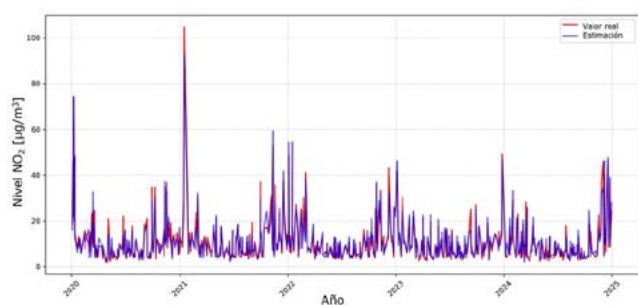


Figura 2. Comparación de la tendencia del NO₂ observado y estimado en la Comunidad de Madrid (2020-2024).

más complejas como redes profundas o fusiones multifuente, han logrado valores próximos a los nuestros (por ejemplo, Cao (2023), con $RMSE = 2,26 \mu g/m^3$ o Xing et al. (2024), aunque suelen basarse en dominios nacionales mucho más extensos y en conjuntos de variables considerablemente mayores. En este contexto, el rendimiento alcanzado en la Comunidad de Madrid es especialmente notable si se considera que el área combina una topografía compleja, fuertes gradientes de emisiones y episodios frecuentes de estabilidad atmosférica.

Un resultado importante es que la estructura espacial utilizada para organizar la información determina en gran medida el rendimiento final de los modelos. La configuración regional completa, que asume homogeneidad en todo el territorio, produjo errores mayores y un comportamiento más irregular, lo que evidencia que la mezcla indiscriminada de entornos muy distintos dificulta la capacidad del modelo para aprender patrones coherentes. La malla regular de $5 \times 5 \text{ km}^2$, aunque mejora la representación local, introduce una fragmentación excesiva y depende fuertemente de la presencia conjunta de estaciones meteorológicas y de NO₂, limitando su robustez. Por el contrario, los clústeres coherentes simplificados permiten agrupar únicamente estaciones con dinámicas atmosféricas similares, evitando la presencia de outliers territoriales y mejorando la internalidad ambiental de cada grupo. La superioridad de esta aproximación sugiere que la coherencia geográfica y meteorológica es un elemento crítico en el diseño de modelos basados en IA para estimación de contaminantes atmosféricos.

El análisis se replicó utilizando datos del reanálisis ERA5-Land en lugar de observaciones meteorológicas de superficie. Aunque las métricas mostraron un deterioro moderado ($RMSE = 3.19 \mu g/m^3$, $R^2 = 0.88$ y $MAE = 2.05$), la calidad de las predicciones sigue siendo elevada. Si bien los modelos entrenados con ERA5-Land no alcanzan la precisión de los basados en estaciones meteorológicas (algo esperable dado que el reanálisis presenta menor resolución espacial y puede no capturar adecuadamente micro escenarios urbanos), el rendimiento logrado es lo suficientemente alto como para considerarlos como una alternativa viable allí donde no exista infraestructura meteorológica local.

4. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

La IA, aplicada de manera rigurosa y bajo un diseño metodológico adecuado, constituye una herramienta sólida para complementar los sistemas tradicionales de monitorización de la calidad del aire. No se trata de utilizarla como sustituto ni como guía autónoma, sino como acompañante metodológico, capaz de aportar continuidad espacial, mejorar la resolución en áreas sin infraestructura suficiente y ofrecer una perspectiva más completa del comportamiento del NO₂ en superficie. El éxito de esta aproximación depende, sin embargo, de decisiones clave en el diseño del estudio: la selección de variables, el tratamiento temporal de la meteorología, la coherencia ambiental de las configuraciones espaciales y la elección del modelo de aprendizaje automático. Todo ello condiciona de manera decisiva la estabilidad y la precisión del sistema.

Uno de los elementos más relevantes derivados de esta investigación es su potencial aplicabilidad a ámbitos territoriales donde los recursos para la monitorización atmosférica son limitados. El precio de una estación convencional de referencia —con costes que superan fácilmente los 50.000 €, y que pueden duplicarse cuando incluyen analizadores de varios contaminantes— supone una barrera significativa para numerosos municipios españoles, especialmente aquellos de tamaño medio y pequeño que deberán cumplir las nuevas obligaciones legales en materia de cambio climático. En este contexto, disponer de metodologías capaces de generar estimaciones fiables de NO₂ a partir de observaciones satelitales, y con el apoyo de información meteorológica procedente de reanálisis como ERA5, representa una oportunidad real para acortar la brecha tecnológica entre territorios y facilitar la toma de decisiones en aquellos lugares donde aún no existe una red de vigilancia consolidada.

Esto tiene implicaciones directas para la próxima década: muchos municipios podrán cumplir y anticipar las obligaciones normativas mediante herramientas híbridas que combinen IA, satélites y datos abiertos, reduciendo su dependencia de infraestructuras costosas. En este sentido, la IA se perfila como un catalizador para democratizar el acceso a la información ambiental en España.

Asimismo, el horizonte tecnológico ofrece perspectivas muy prometedoras. La reciente puesta en órbita de Sentinel-4, y la futura operación sistemática de Sentinel-5, garantizarán una mejora sustancial tanto en frecuencia como en resolución espectral y espacial respecto a Sentinel-5P/TROPOMI. Sentinel-4 proporcionará observaciones geoestacionarias con una alta frecuencia temporal (una medición por hora sobre Europa), lo que permitirá caracterizar episodios de contaminación con gran detalle y capturar dinámicas intradiarias

imposibles de observar con los sensores actuales. Por su parte, Sentinel-5 ampliará la continuidad de las series troposféricas a escala global, manteniendo la calidad radiométrica y espectral que ha caracterizado a TROPOMI. Ambas misiones abrirán la puerta a modelos de IA más ricos, capaces de integrar información temporal casi continua y generar estimaciones de NO₂ con una precisión aún mayor.

Finalmente, merece especial atención el papel que pueden desempeñar los sensores de bajo coste en este ecosistema de herramientas. Aunque no alcanzan la precisión de las estaciones de referencia, estos dispositivos ofrecen una densidad de datos muy superior y una capacidad de despliegue rápido en entornos urbanos y periurbanos. Su integración como fuentes secundarias de validación, o como elementos auxiliares para corregir estimaciones satelitales, podría mejorar notablemente la robustez de los modelos. Explorar este potencial, así como desarrollar protocolos para la fusión adecuada de datos procedentes de sensores de bajo coste, satélites y estaciones convencionales, constituye una línea de investigación clave para los próximos años.

5. REFERENCIAS

- Balamurugan, V., Chen, J., Wenzel, A., Keutsch, F.N., 2023. Spatiotemporal modeling of air pollutant concentrations in Germany using machine learning. *Atmos. Chem. Phys.* 23, 10267–10285. <https://doi.org/10.5194/acp-23-10267-2023>
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- Cao, E.L., 2023. National ground-level NO₂ predictions via satellite imagery driven convolutional neural networks. *Front. Environ. Sci.* 11, 1285471. <https://doi.org/10.3389/FENVS.2023.1285471/BIBTEX>
- Cedeno Jimenez, J.R., Pugliese Vioria, A. de J., Brovelli, M.A., 2023. Estimating Daily NO₂ Ground Level Concentrations Using Sentinel-5P and Ground Sensor Meteorological Measurements. *ISPRS Int. J. Geo-Information* 12, 107. <https://doi.org/10.3390/ijgi12030107>
- Chan, K.L., Khorsandi, E., Liu, S., Baier, F., Valks, P., 2021. Estimation of Surface NO₂ Concentrations over Germany from TROPOMI Satellite Observations Using a Machine Learning Method. *Remote Sens.* 2021, Vol. 13, Page 969 13, 969. <https://doi.org/10.3390/RS13050969>
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 13-17-August-2016, 785–794. <https://doi.org/10.1145/2939672.2939785;CSUBTYPE:STRING:-CONFERENCE>
- City of Madrid, 2025a. Hourly air quality data since 2001 [WWW Document]. URL [https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vg-](https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vg-nextoid=f3c0f7d512273410VgnVCM2000000c205a0aR-CRD&vgnextchannel=374512b9ace9f310VgnVCM-100000171f5a0aRCRD&vgnextfmt=default)
- nextoid=f3c0f7d512273410VgnVCM2000000c205a0aR-CRD&vgnextchannel=374512b9ace9f310VgnVCM-100000171f5a0aRCRD&vgnextfmt=default (accessed 11.20.25).
- City of Madrid, 2025b. Hourly meteorological data since 2019 [WWW Document]. URL <https://datos.madrid.es/sites/v/index.jsp?vgnextoid=fa8357cec5efa610VgnVCM-1000001d4a900aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD> (accessed 11.20.25).
- Community of Madrid, 2025a. Hourly air quality data since 2005 [WWW Document]. URL https://datos.comunidad.madrid/dataset/calidad_aire_datos_historico (accessed 11.20.25).
- Community of Madrid, 2025b. Hourly meteorological data since 2020 [WWW Document]. URL https://datos.comunidad.madrid/catalogos/#/dataset/calidad_aire_datos_meteo_historico/?view=info (accessed 11.20.25).
- Das, B.P., Pathan, M.S., Lee, Y.H., Dev, S., 2021. Estimating Ground-level Nitrogen Dioxide Concentration from Satellite Data. *Prog. Electromagn. Res. Symp.* 2021-November, 1176–1182. <https://doi.org/10.1109/PIERS53385.2021.9694752>
- Earth Engine Data Catalog, 2024. Sentinel-5P OFFL NO₂: Offline Nitrogen Dioxide [WWW Document]. URL https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_OFFL_L3_NO2#description (accessed 4.17.23).
- Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. <https://doi.org/10.1214/aos/1013203451> 29, 1189–1232. <https://doi.org/10.1214/AOS/1013203451>
- Khomenko, S., Cirach, M., Pereira-Barboza, E., Mueller, N., Barrera-Gómez, J., Rojas-Rueda, D., de Hoogh, K., Hoek, G., Nieuwenhuijsen, M., 2021. Premature mortality due to air pollution in European cities: a health impact assessment. *Lancet Planet. Heal.* 5, e121–e134. [https://doi.org/10.1016/S2542-5196\(20\)30272-2](https://doi.org/10.1016/S2542-5196(20)30272-2)
- Morillas, C., Álvarez, S., Pires, J.C.M., García, A.J., Martínez, S., 2025. Linking Satellite and Ground Observations of NO₂ in Spanish Cities: Influence of Meteorology and O₃. *Nitrogen (Switzerland)* 6, 32. <https://doi.org/10.3390/NITROGEN6020032/S1>
- Morillas, C., Alvarez, S., Serio, C., Masiello, G., Martinez, S., 2024. TROPOMI NO₂ Sentinel-5P data in the Community of Madrid: A detailed consistency analysis with in situ surface observations. *Remote Sens. Appl. Soc. Environ.* 33, 101083. <https://doi.org/10.1016/J.RSASE.2023.101083>
- Muñoz Sabater, J., 2019. ERA5-Land hourly data from 1950 to present. <https://doi.org/10.24381/cds.e2161bac>
- Van Geffen, J.H.G.M., Eskes, H.J., Boersma, K.F., Maasakkers, J.D., Veefkind, J.P., 2019. TROPOMI ATBD of the total and tropospheric NO₂ data products.
- Veefkind, J.P., Aben, I., McMullan, K., Förster, H., de Vries, J.,

Otter, G., Claas, J., Eskes, H.J., de Haan, J.F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., Levelt, P.F., 2012. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens. Environ.* 120, 70–83. <https://doi.org/10.1016/J.RSE.2011.09.027>

World Health Organization (WHO), 2021. WHO global air

quality guidelines, Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.

Xing, J., Baek, B.H., Li, S., Wang, C.T., Song, G., Ma, S., Zheng, S., Liu, C., Tong, D., Woo, J.H., Liu, T.Y., Fu, J.S., 2024. A Physically Constrained Deep-Learning Fusion Method for Estimating Surface NO₂ Concentration from Satellite and Ground Monitors. *Environ. Sci. Technol.* 58, 21218–21228. <https://doi.org/10.1021/ACS.EST.4C07341>

Sobre los autores

Carlos Morillas

Profesor Ayudante Doctor en la Universidad Politécnica de Madrid (ETSI de Caminos, Canales y Puertos) y doctor en Ingeniería Geomática. Su investigación se centra en el análisis geoespacial aplicado a la calidad del aire, inventarios de gases de efecto invernadero y resiliencia climática, integrando teledetección satelital, datos in situ e inteligencia artificial. Ha participado en proyectos de investigación nacionales y europeos y es autor de diversas publicaciones en revistas indexadas en los ámbitos de teledetección, cambio climático y evaluación ambiental. Es coordinador del área de empresas del Observatorio de Acción Climática.

Rocco Giosa

Doctorando del programa Engineering for Innovation and Sustainable Development en la Universidad de Basilicata, donde también obtuvo el Máster en Ingeniería Informática y Ciencia de la Información. Forma parte del grupo de investigación AS ART, centrando su trabajo en monitorización ambiental y observación de la Tierra mediante técnicas avanzadas de inteligencia artificial. Está especializado en el procesamiento de datos satelitales y en el desarrollo de modelos de IA para la estimación de variables atmosféricas, utilizando herramientas como Python y MATLAB. Durante una estancia en la Universidad Politécnica de Madrid, investigó la contaminación por NO₂ en la Comunidad de Madrid, analizando la relación entre datos satelitales (TROPOMI – Sentinel-5P) y mediciones en superficie, y desarrollando modelos de aprendizaje automático (Random Forest y redes neuronales) para la estimación de concentraciones a nivel del suelo. Actualmente trabaja en el desarrollo de una Physics-Informed Neural Network (PINN) para identificar fuentes de emisión de metano y reconstruir perfiles verticales de concentración a partir de datos IASI simulados. Entre sus próximos objetivos se encuentra el diseño de una metodología de inversión multigás para estimar simultáneamente perfiles verticales de los principales gases de efecto invernadero.

Sergio Álvarez

Profesor Contratado Doctor en el Departamento de Ingeniería y Morfología del Terreno de la Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos de la Universidad Politécnica de Madrid. Es doctor Ingeniero de Montes. Su investigación se centra en la evaluación de la sostenibilidad mediante Análisis de Ciclo de Vida, huella de carbono y modelos input-output multirregionales aplicados a organizaciones, productos y sistemas territoriales. Es autor de numerosas publicaciones en revistas internacionales de alto impacto y de varios libros especializados sobre huella de carbono y análisis ambiental. Ha participado en proyectos competitivos nacionales e internacionales y ha dirigido tesis y trabajos de investigación en el ámbito de la sostenibilidad y la evaluación ambiental. Es socio fundador y presidente del Observatorio de Acción Climática.

Sara Martínez

Profesora Permanente Laboral en la Universidad Politécnica de Madrid (E.T.S.I. de Caminos, Canales y Puertos) y doctora en Ingeniería de Montes por la UPM. Desarrolla su actividad docente e investigadora en el ámbito de la ingeniería ambiental, con especialización en huella de carbono, huella ambiental y huella de nitrógeno, así como en el uso de datos satelitales (Sentinel-5P) para el estudio de la calidad del aire. Su investigación abarca desde la evaluación ambiental de sistemas agroalimentarios y proyectos de construcción hasta el análisis de emisiones atmosféricas y la integración de observaciones satelitales y datos in situ. Ha participado en proyectos competitivos de investigación y en contratos de I+D con entidades públicas y privadas, y es coautora de dos patentes en el ámbito de la innovación tecnológica aplicada al sector ambiental y agrícola. Complementa su actividad académica con más de ocho años de experiencia profesional en la empresa Técnicas Reunidas S.A., en el área de procesos de agua y tratamiento de aguas residuales. Es además tesorera y socia fundadora del Observatorio de Acción Climática, impulsando iniciativas orientadas a la sostenibilidad ambiental.