

# IMPLEMENTACIÓN Y EVALUACIÓN DE UN MODELO DE PRONÓSTICO ESTADÍSTICO DE PM2.5 PARA EL VALLE DE ABURRÁ

Mauricio Ramírez<sup>1</sup>, **K. Santiago Hernández<sup>1</sup>**, Jhayron S. Pérez-Carrasquilla<sup>2</sup>,  
Paola Montoya<sup>1</sup>, Juan M. Sánchez<sup>1</sup>, Manuel D. Zuluaga<sup>1</sup>

<sup>1</sup> Sistema de Alerta Temprana de Medellín y el Valle de Aburrá - SIATA

<sup>2</sup> Department of Atmospheric and Oceanic Science, University of Maryland, College Park, USA

**CASAP IX, 22-24 de marzo del 2023**

**Santa Marta - Colombia**

Presentador: K. Santiago Hernández

# Introducción

- Existe una creciente necesidad de gestionar la calidad del aire en las ciudades.
- Entender y pronosticar el estado de la calidad del aire es fundamental para hacer una gestión adecuada de la misma.



## Design of an early alert system for PM<sub>2.5</sub> through a stochastic method and machine learning models

Nathalia Celis<sup>a</sup>, Alejandro Casallas<sup>b</sup>, Ellie Anne López-Barrera<sup>d</sup>, Hermes Martínez<sup>a</sup>, Carlos A. Peña Rincón<sup>a</sup>, Ricardo Arenas<sup>c</sup>, Camilo Ferro<sup>d</sup>

Original Article | Published: 13 September 2021

## Long short-term memory artificial neural network approach to forecast meteorology and PM<sub>2.5</sub> local variables in Bogotá, Colombia

Alejandro Casallas<sup>a</sup>, Camilo Ferro, Nathalia Celis, Marco Andrés Guevara-Luna, Caroline Mogollón-Sotelo, Fredy Alejandro Guevara-Luna & Martha Merchán

Modeling Earth Systems and Environment 8, 2951-2964 (2022) | Cite this article  
410 Accesses | 10 Citations | Metrics



## Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in Aburrá Valley, Colombia

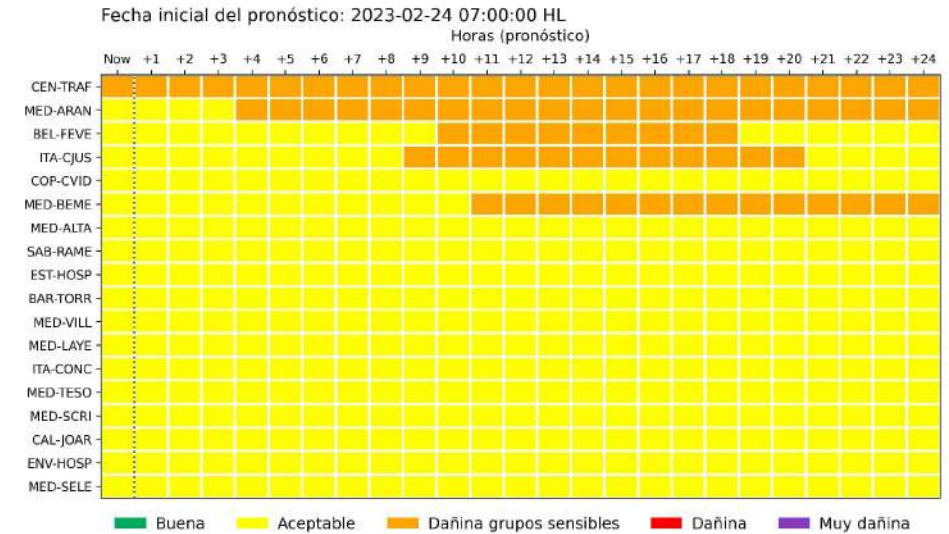
J. Murillo-Escobar<sup>a</sup>, J.P. Sepulveda-Suescun<sup>a</sup>, M.A. Correa<sup>b</sup>, D. Orrego-Metaute<sup>a</sup>



Original Article

## Discovering relationships and forecasting PM<sub>10</sub> and PM<sub>2.5</sub> concentrations in Bogotá, Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering

Fabiana Franceschi, Martha Cobo, Manuel Figueredo



- Las técnicas de Machine Learning (ML) han adquirido gran importancia en el estudio y pronóstico de la calidad del aire.
- Diferentes estudios en el país han mostrado la utilidad de este tipo de técnicas.



## ■ **Introducción**

---

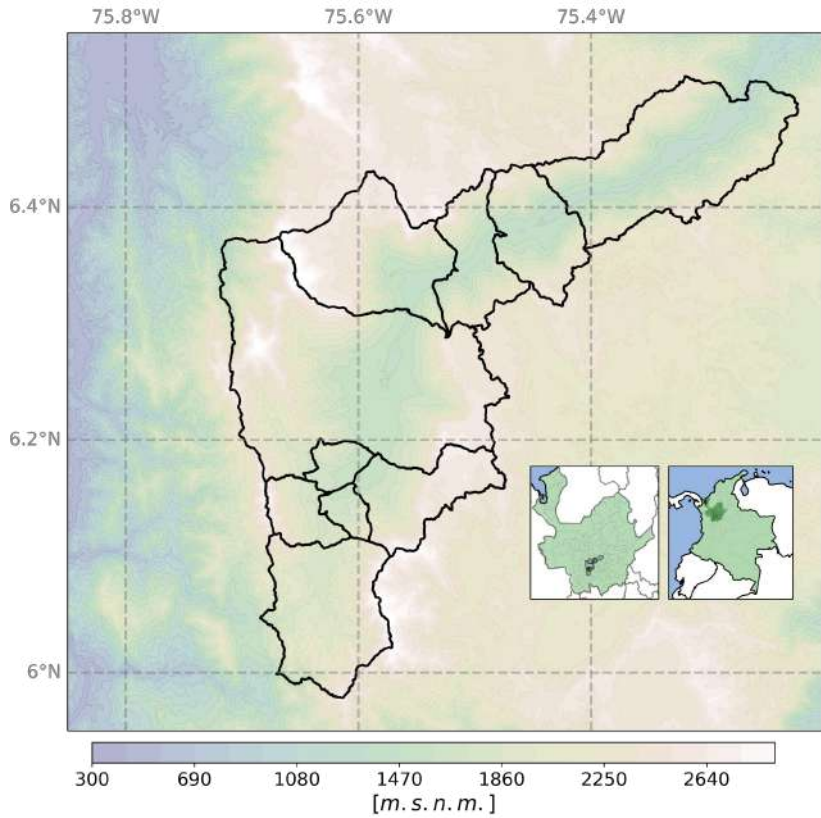
**Evaluar el desempeño de diferentes modelos estadísticos basados en ML e implementados por el Sistema de Alerta Temprana de Medellín y el Valle de Aburrá (SIATA) para el pronóstico del material particulado menor a 2.5 micras (PM2.5).**



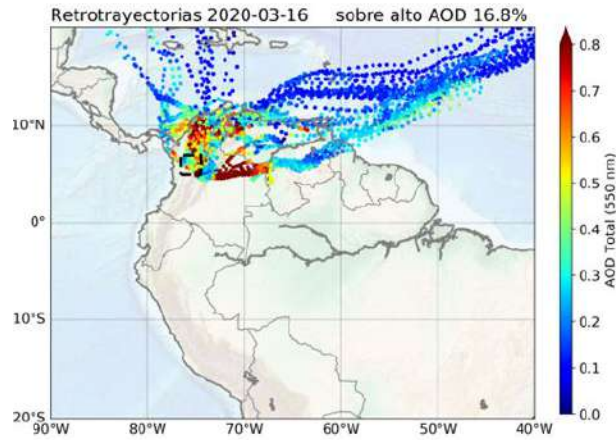


# Contexto

El Valle de Aburrá



Eventos de contaminación



SIATA



Datos pendientes de validación final.  
\*Cálculos a partir de registros de las 24 horas previas a este reporte



# Datos y métodos

- Periodo de entrenamiento: información histórica de cada estación hasta marzo 2020.
- 70% para entrenamiento - 30% para validación.
- Evaluación para el periodo comprendido entre el 14 de febrero y el 1 de abril del 2022.

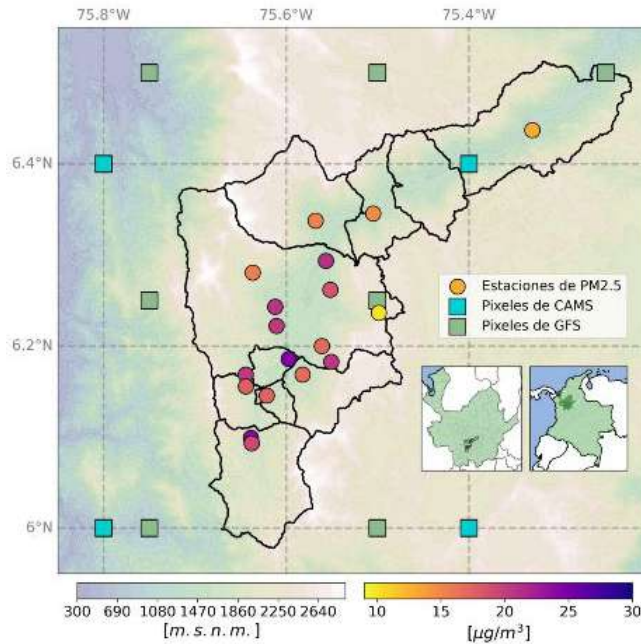
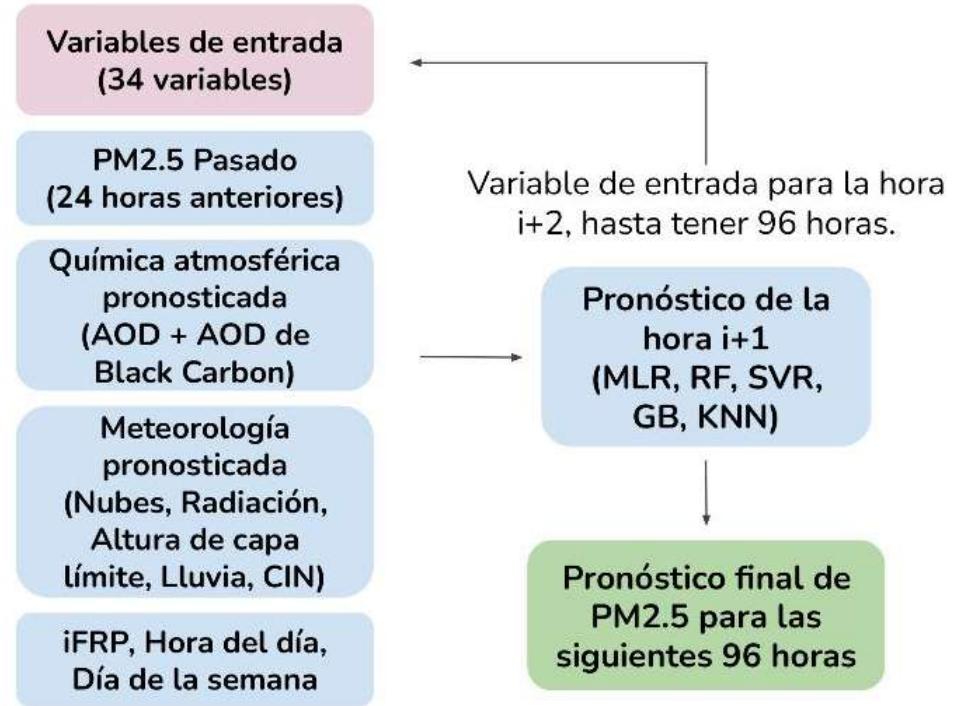


Figura 1. Ubicación espacial de las fuentes de información y estrategia de entrenamiento del modelo de pronóstico de PM2,5.



$$RMSE = \sqrt{\frac{\sum_{i=1}^N (F_i - O_i)^2}{N}}; \quad BIAS = \frac{1}{N} \sum_{i=1}^N (F_i - O_i)$$

$$CSI = \frac{Aciertos}{Aciertos - Falsas Alarmas - Desaciertos}$$

- Las métricas se calcularon en función del horizonte de pronóstico y la hora del día.



# Resultados

Tabla 1. Métricas de error calculadas durante todo el período de evaluación.

Modelo	MLR	RF	SVR	GB	KNN	MEAN	CAMS
<b>CORR</b>	0,41	0,51	0,45	0,20	0,44	0,51	0,17
<b>RMSE (<math>\mu\text{g}/\text{m}^3</math>)</b>	10,66	9,60	9,85	11,62	9,60	9,23	12,62
<b>BIAS (<math>\mu\text{g}/\text{m}^3</math>)</b>	-5,83	-4,41	-3,31	-1,39	-3,19	-4,43	-6,87

- Random Forest (RF) y Máquinas de Soporte (SVR) obtuvieron un mejor desempeño.
- Cerca a la hora de inicialización, el promedio de los modelos y RF presentan una mejor capacidad.
- SVR es capaz de mantener correlaciones relativamente altas (por encima de 0.5) aún para horizontes de pronóstico mayores.

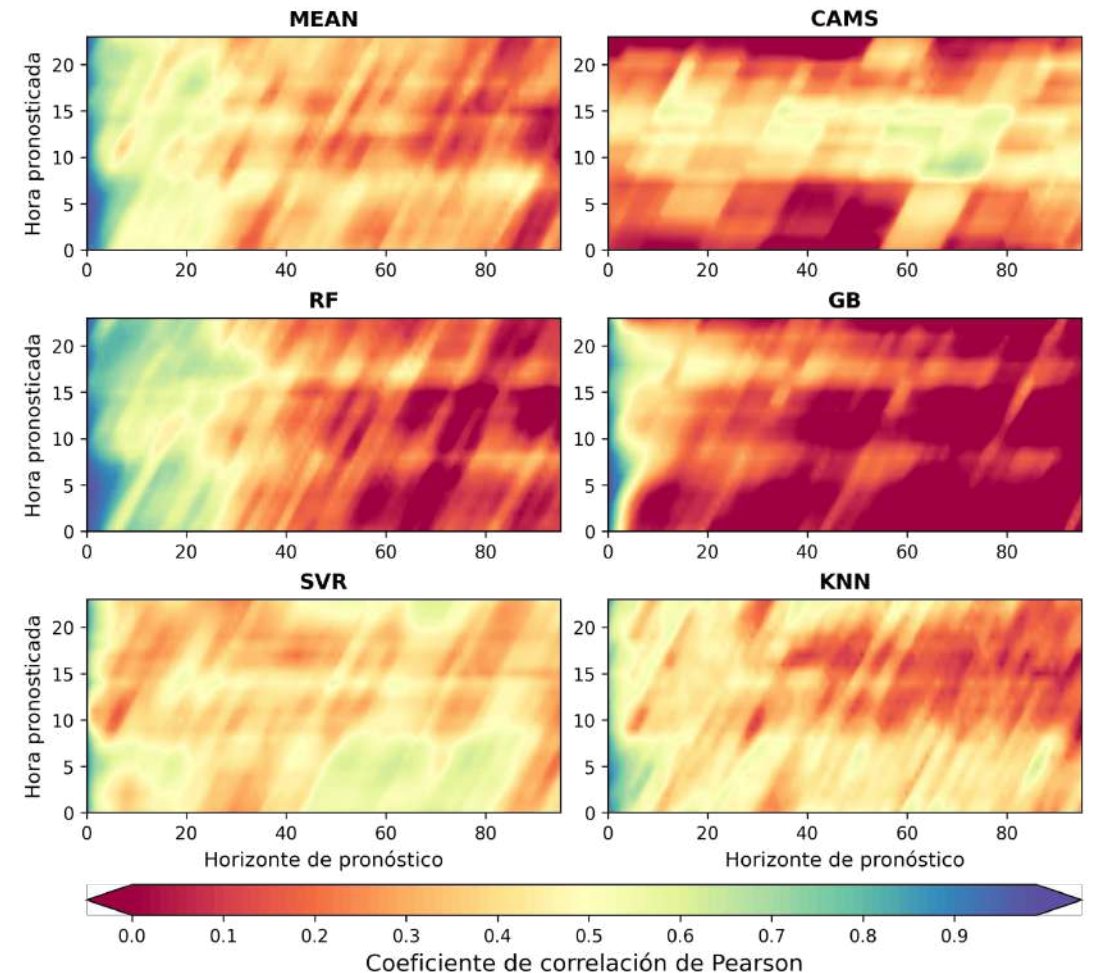


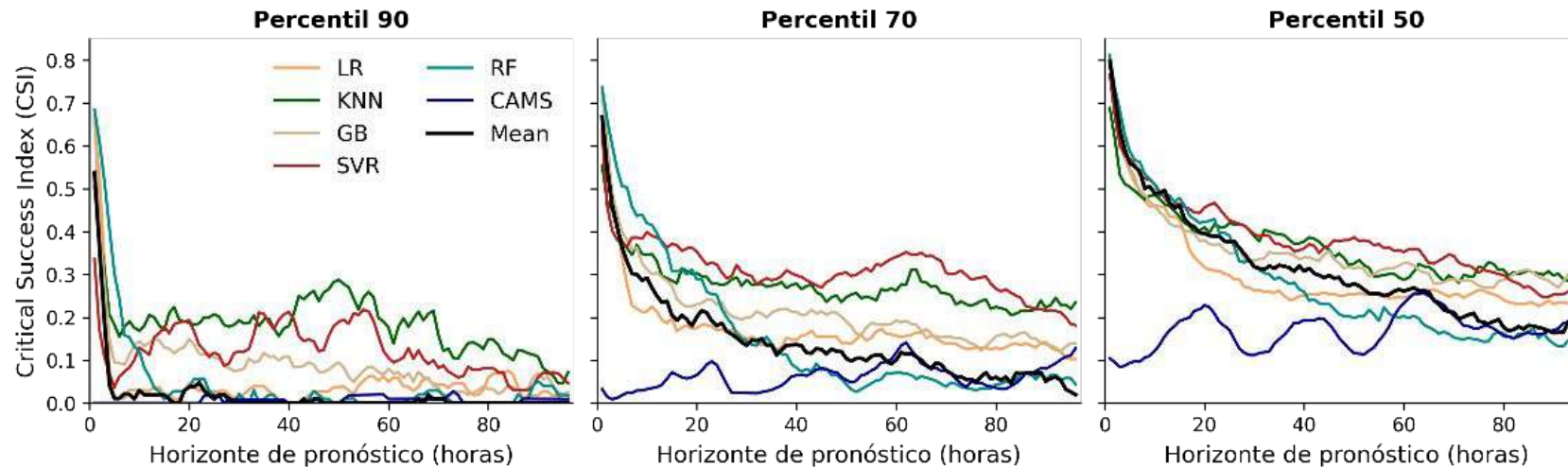
Figura 2. Coeficiente de correlación de Pearson en función de la hora del día y el horizonte de pronóstico. Cada uno de los paneles muestra un modelo distinto.





# Resultados

$$CSI = \frac{Aciertos}{Aciertos - Falsas\ Alarmas - Desaciertos}$$

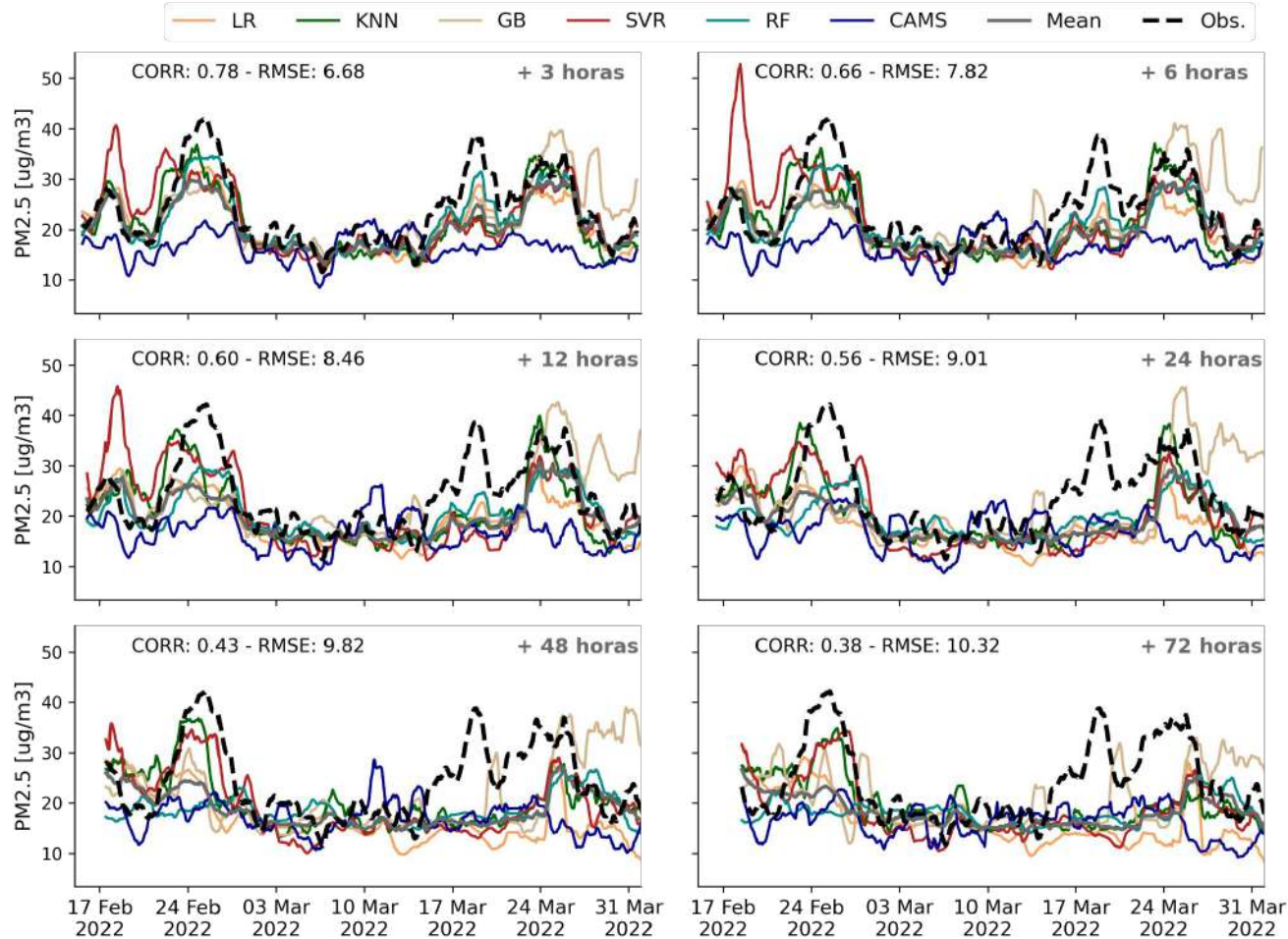


**Figura 3.** Índice de éxito crítico calculado en función del horizonte de pronóstico. Las líneas de colores muestran los resultados para los diferentes modelos.

- Los modelos presentan una menor capacidad para captar concentraciones con umbrales más altos. CAMS no captura eventos con concentraciones superiores al percentil 90 en ningún horizonte de pronóstico.
- RF tiene mejores resultados cerca de la hora de inicialización de los pronósticos.
- Los modelos KNN y SVR presentan una mejor capacidad para horizontes de tiempo mayores en comparación con los otros modelos.



# Resultados



- CAMS no capta los picos, mientras que los modelos de ML pronostican las variaciones y magnitudes de las concentraciones.
- SVR y KNN pronostican adecuadamente el pico de contaminación ocurrido en la semana del 27 de febrero del 2022 con 3 días de anticipación.
- Los aumentos de las concentraciones después del 15 de marzo no fueron pronosticados por ninguno de los modelos 48 horas antes de su ocurrencia.

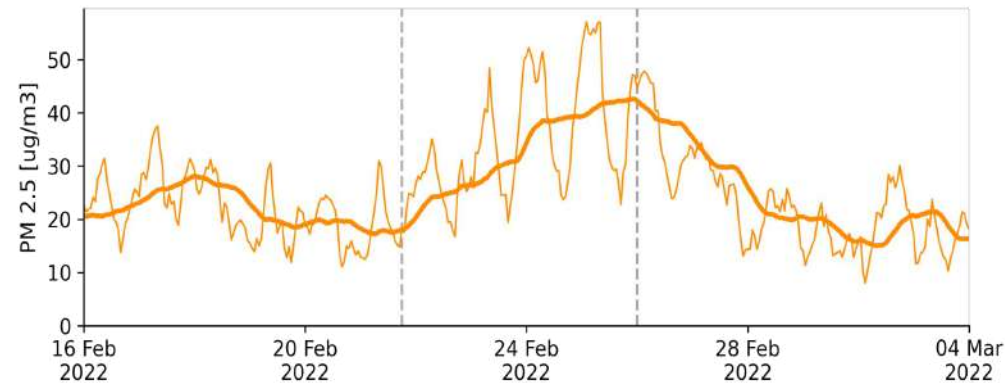
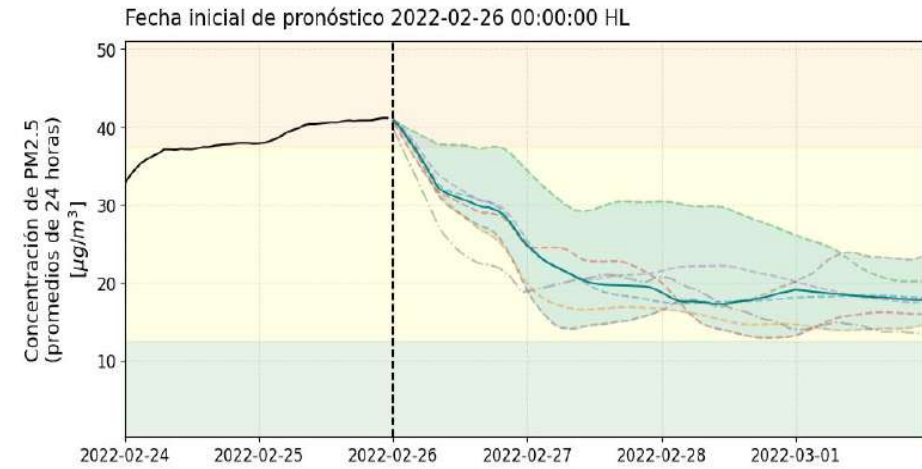
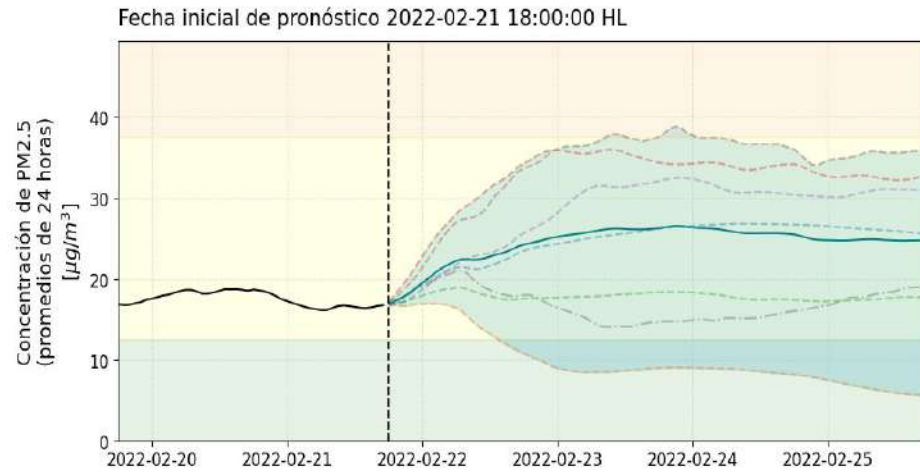
**Figura 4.** Series de tiempo de la concentración de PM2.5 promedio de 24 horas observada (línea negra discontinua) y pronosticada por los distintos modelos (líneas de colores) en diferentes ventanas de pronóstico (paneles). En cada panel se presentan métricas de correlación (CORR) y RMSE para el promedio de todos los modelos (MEAN).





# Resultados

Algunos ejemplos de la utilidad de los modelos



# ■ Conclusiones y trabajo a futuro

---

De forma general, los resultados presentados resaltan diferencias en la capacidad de los distintos modelos para captar variaciones en las concentraciones de material particulado en el Valle de Aburrá. Los análisis mostraron que **los modelos de ML implementados en la región tienen un mejor desempeño que el pronóstico global de CAMS**, y en particular para eventos con altas concentraciones, lo cual es relevante en términos de gestión de la calidad del aire. Finalmente, se espera que estos resultados contribuyan a la mejora de los modelos de pronóstico de la calidad del aire en el Valle de Aburrá.

## TRABAJO A FUTURO

- Exploración de nuevas variables
- Implementación de pronósticos WRF - SIATA
- Aumentar el periodo de entrenamiento
- Nueva estrategia de pronóstico y nuevos modelos



# ■ Invitación a ponencias

---

## **Ciudadanos Científicos: una estrategia para la apropiación social del conocimiento y la gestión del recurso aire en el Valle de Aburrá**

Mauricio Ramírez Arias<sup>1</sup>, Maura Melissa Herrera<sup>2</sup>, Natalia Bernal<sup>3</sup>, Alejandra Montejo Barato<sup>4</sup>, Juan David Berrio<sup>5</sup>,  
Manuel D. Zuluaga<sup>6</sup>, Ana Z. Orrego<sup>7</sup>, Laura C. Gil<sup>8</sup>

<sup>1</sup> Área Metropolitana del Valle de Aburrá, Sistema de Alerta Temprana de Medellín y el Valle de Aburrá,

---

## **Procesos de dispersión de PM2.5 en un evento de contaminación por transporte regional de aerosoles: El rol de la turbulencia atmosférica en un valle urbanizado y topográficamente complejo.**

Guiller Alexis Ayala Parra<sup>1</sup>, K. Santiago Hernández<sup>2</sup>, Camilo Chalán Fernandez<sup>3</sup>, Mauricio Ramírez Arias<sup>4</sup>, Manuel D. Zuluaga<sup>5</sup>

<sup>1</sup> Área Metropolitana del Valle de Aburrá, Sistema de Alerta Temprana de Medellín y el Valle de Aburrá,





# Agradecimientos

Esta investigación fue liderada por **Jhayron S. Pérez-Carrasquilla**, y se llevó a cabo por el Sistema de Alerta Temprana del Valle de Aburrá, **SIATA**, proyecto del **Área Metropolitana del Valle de Aburrá** durante la vigencia del contrato de Ciencia y Tecnología CCT 799 de 2021. Los autores agradecen a NASA, ECMWF y NCEP por proveer datos de incendios, del modelo CAMS y de GFS, respectivamente.



# ■ Referencias

---

- Casallas, A., Ferro, C., Celis, N., Guevara-Luna, M. A., Mogollón-Sotelo, C., Guevara-Luna, F. A., & Merchán, M. (2021). Long short-term memory artificial neural network approach to forecast meteorology and PM 2.5 local variables in Bogotá, Colombia. *Modeling Earth Systems and Environment*, 1-14.
- Celis, N., Casallas, A., López-Barrera, E. A., Martínez, H., Rincón, C. A. P., Arenas, R., & Ferro, C. (2022). Design of an Early Alert System for PM2. 5 through a stochastic method and machine learning models. *Environmental Science & Policy*, 127, 241-252.
- Franceschi, F., Cobo, M., & Figueredo, M. (2018). Discovering relationships and forecasting PM10 and PM2. 5 concentrations in Bogotá, Colombia, using artificial neural networks, principal component analysis, and k-means clustering. *Atmospheric Pollution Research*, 9(5), 912-922.
- Henao, J. J., Rendón, A. M., Hernández, K. S., Giraldo-Ramirez, P. A., Robledo, V., Posada-Marín, J. A., ... & Mejía, J. F. (2021). Differential Effects of the COVID-19 Lockdown and Regional Fire on the Air Quality of Medellín, Colombia. *Atmosphere*, 12(9), 1137.
- Murillo-Escobar, J., Sepulveda-Suescun, J. P., Correa, M. A., & Orrego-Metaute, D. (2019). Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in Aburrá Valley, Colombia. *Urban climate*, 29, 100473.





# ¡MUCHAS GRACIAS!

K. Santiago Hernández

[ksantiago.hernandez@udea.edu.co](mailto:ksantiago.hernandez@udea.edu.co)

[ksantiago.hernandezu@gmail.com](mailto:ksantiago.hernandezu@gmail.com)

SIATA

[contacto@siata.gov.co](mailto:contacto@siata.gov.co)

<https://siata.gov.co/>

Más información



<https://casap.science/>



[casap@casap.science](mailto:casap@casap.science)