



Análisis y predicción de series temporales de los niveles de concentración de NO<sub>2</sub> y PM<sub>10</sub> en el Eixample, Barcelona

Adrián Hinojosa Calleja

**IT ACADEMY**

**atmosphere**

**Article**

**Time Series Forecasting of Air Quality: A Case Study of Sofia City**

Evgeniy Marinov <sup>1,\*</sup>, Dessislava Petrova-Antonova <sup>1,2,\*</sup> and Simeon Malinov <sup>1</sup>

<sup>1</sup> GATE Institute, Sofia University "St. Kliment Ohridski", 1504 Sofia, Bulgaria; simeon.malinov@gate-ai.eu  
<sup>2</sup> Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", 1164 Sofia, Bulgaria  
\* Correspondence: evgeniy.marinov@gate-ai.eu (E.M.); dessislava.petrova@gate-ai.eu (D.P.-A.)

**Abstract:** Air pollution has a significant impact on human health and the environment, causing cardiovascular disease, respiratory infections, lung cancer and other diseases. Understanding the behavior of air pollutants is essential for adequate decisions that can lead to a better quality of life for citizens. Air quality forecasting is a reliable method for taking preventive and regulatory actions. Time series analysis produces forecasting models, which study the characteristics of the data points over time to extrapolate them in the future. This study explores the trends of air pollution at five air quality stations in Sofia, Bulgaria. The data collected between 2015 and 2019 is analyzed applying time series forecasting. Since the time series analysis works on complete data, imputation techniques are used to deal with missing values of pollutants. The data is aggregated by granularity periods of 3 h, 6 h, 12 h, 24 h (1 day). The AutoRegressive Integrated Moving Average (ARIMA) method is employed to create statistical analysis models for the prediction of pollutants' levels at each air quality station and for each granularity, including carbon dioxide (CO), nitrogen dioxide ( $\text{NO}_2$ ), ozone ( $\text{O}_3$ ) and fine particles (PM2.5). In addition, the method allows us to find out whether the pollutants' levels exceed the limits prescribed by the World Health Organization (WHO), as well as to investigate the correlation between levels of a given pollutant measured in different air quality stations.

**Keywords:** air quality; time series analysis; forecasting models; prediction of pollutants' levels

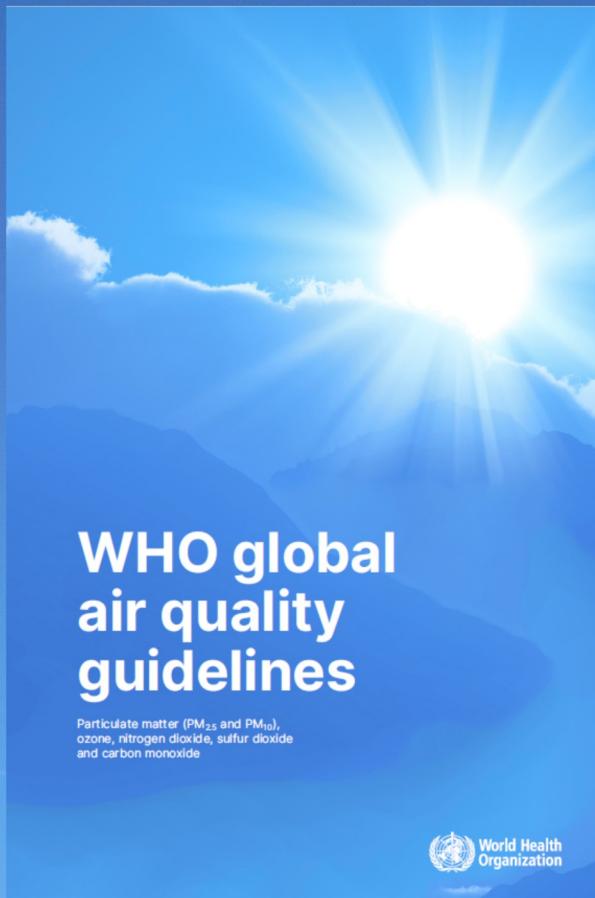
**check for updates**

**Citation:** Marinov, E.; Petrova-Antonova, D.; Malinov, S.

# Ingredient

1

# Ingrediente 2



**Calidad del Aire**

CALIDAD DEL AIRE ▾ LA SALUD ▾ LA CONTAMINACIÓN ▾

Calidad del aire/ Qué pasa con el aire que respiramos

## Qué pasa con el aire que respiramos

La intensa actividad industrial, el elevado volumen de tráfico en Barcelona y en su entorno metropolitano, el clima mediterráneo de escasa lluvia media a lo largo del año y la llegada de nubes de polvo sahariano favorecen concentraciones de elementos contaminantes en el aire que respiramos. Este hecho afecta negativamente a nuestra salud.

Los dos elementos perjudiciales para la salud de las personas más abundantes en el aire de Barcelona son el **dióxido de nitrógeno (NO<sub>2</sub>)**, un gas irritante que ataca a los pulmones y está producido sobre todo por los motores de combustión interna, y **las partículas en suspensión de menos de 10 micras (PM<sub>10</sub>)**, que afectan especialmente a los sistemas respiratorio y cardiovascular y que tienen origen en los procesos de combustión, rodamientos y frenazos de vehículos, en las obras de construcción, o en el polvo que llega del Sáhara.

Aunque **son invisibles a nuestros ojos**, estos contaminantes son muy perjudiciales para la salud de todas las personas, pero especialmente para los colectivos más vulnerables, como niños y niñas, mujeres embarazadas, personas adultas con cardiopatías o afecciones respiratorias y personas mayores.

# Ingrediente

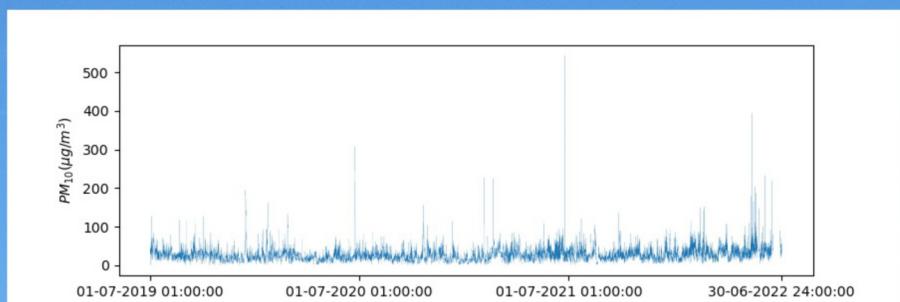
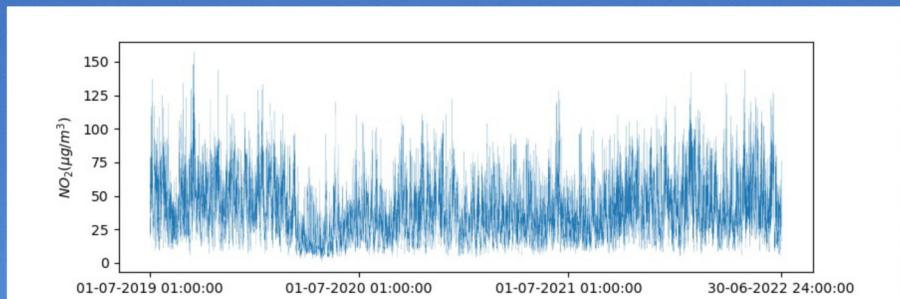
3



# Objetivos

- Evaluar el nivel de exposición a concentraciones diarias de  $\text{NO}_2$  y  $\text{PM}_{10}$  en los últimos tres años de acuerdo a la normativa de la OMS
- Estudiar un modelo de series temporales tipo ARIMA para predecir la concentración diaria de dichos contaminantes

# Procesamiento de datos

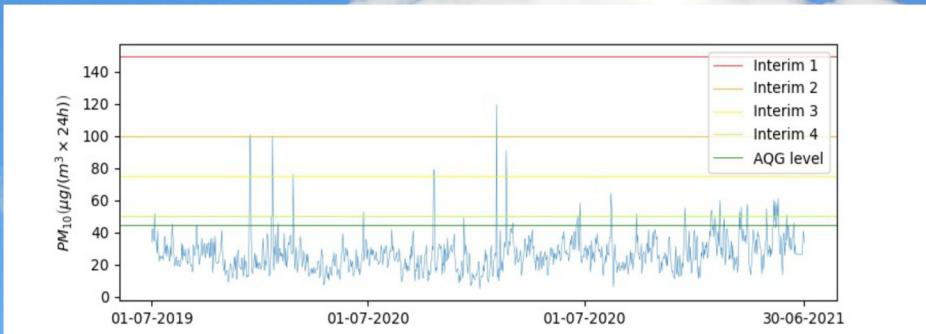
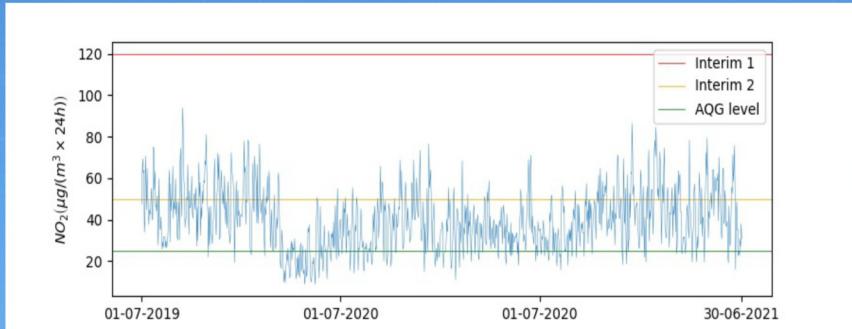


- Se sustituyeron los valores faltantes con la media de la serie

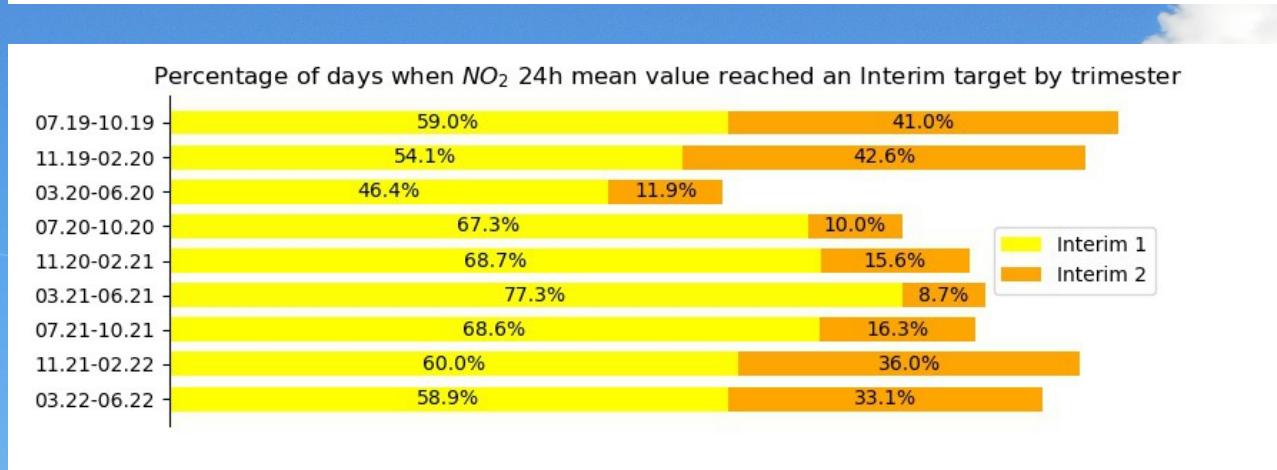
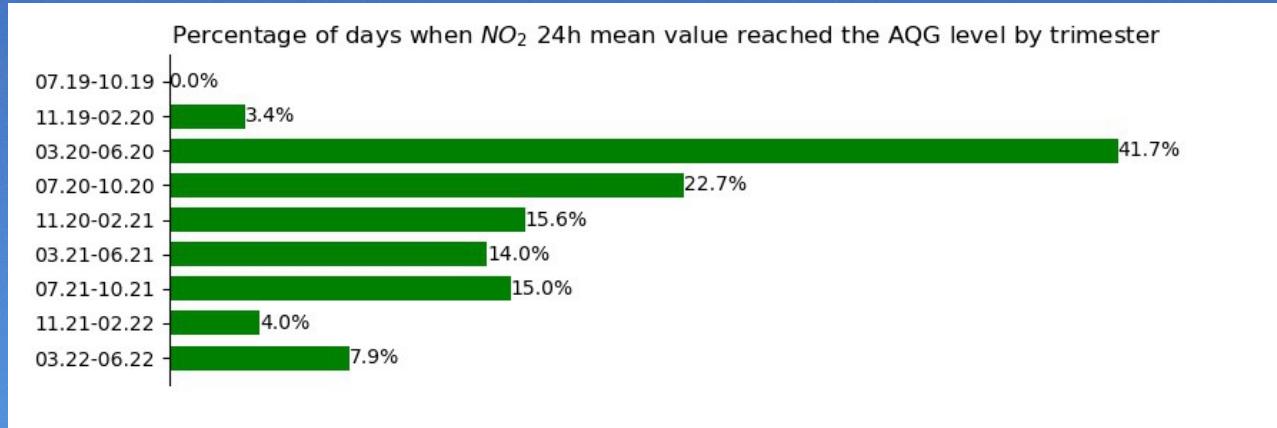
Polluter	Mean	Std	Min	Max	Missing
N02	39.76	21.67	1	157	3.67%
PM10	26.59	16.89	1	543	6.10%

# Análisis de la media a 24h

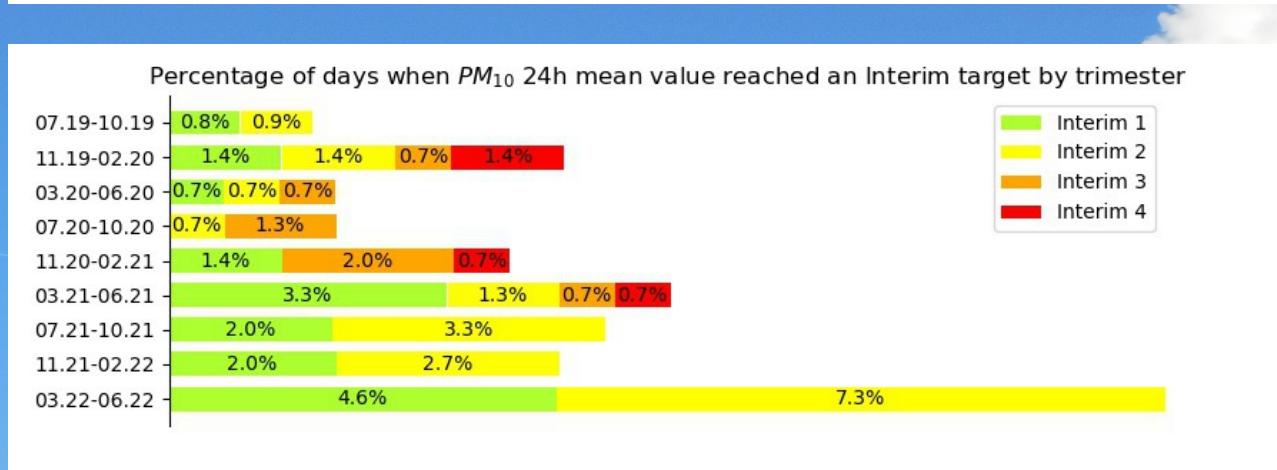
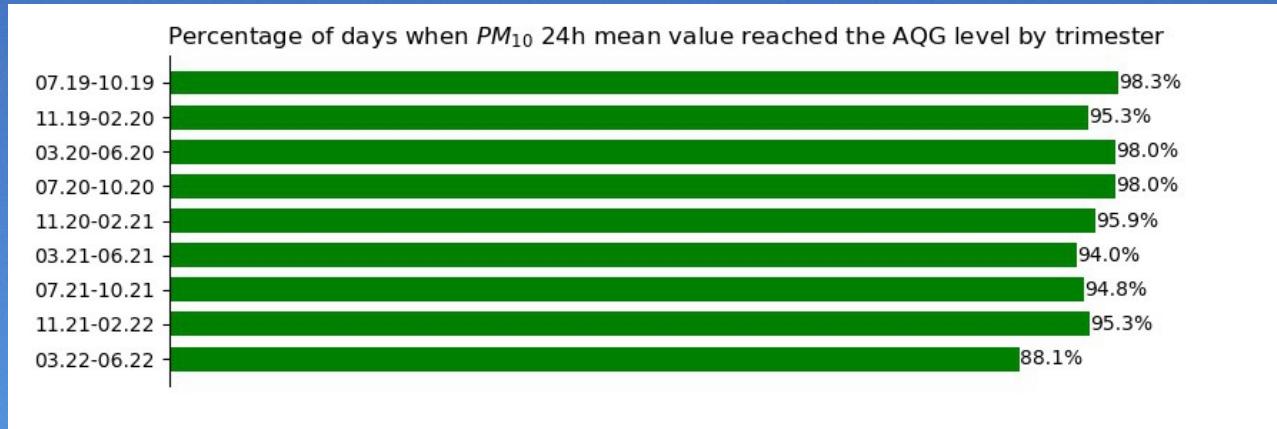
Pollutant	Interim target				AQG level
	1	2	3	4	
$NO_2 (\mu g/(m^3 \times 24h))$	120	50	-	-	25
$PM_{10} (\mu g/(m^3 \times 24h))$	150	100	75	50	45



# Análisis de la media a 24 h



# Análisis de la media a 24 h



# ARIMA ( p, d, q)

Modelo autorregresivo integrado de media móvil

- Número de observaciones previas p
- Grado de incrementos d
- Tamaño de la media móvil q

# Métricas de evaluación

- Error Absoluto Medio (MAE)

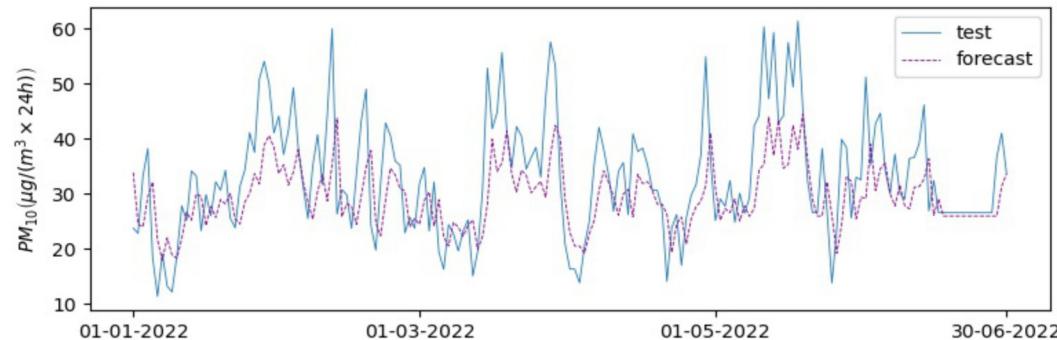
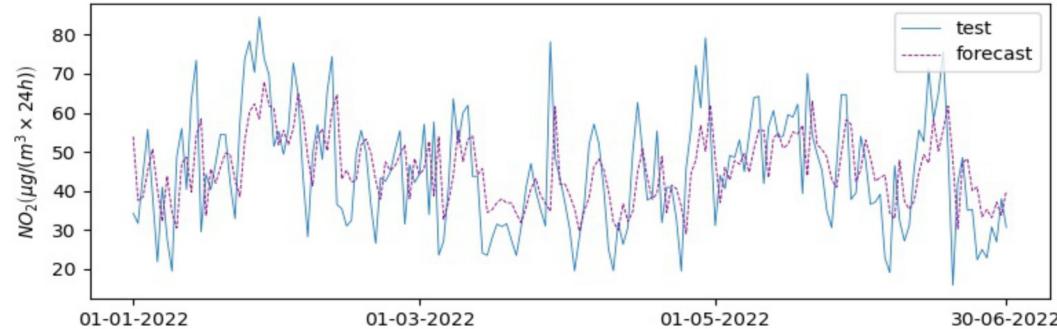
$$MAE = \frac{1}{n} \sum_{t=1}^n |x_t - y_t|.$$

- Porcentaje de Error Absoluto Medio (MAPE)

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|x_t - y_t|}{|x_t|}.$$

# Ajuste de parámetros y predicción

Pollutant	p	d	q	MAE train	MAPE train	MAE test	MAPE test
$NO_2$	2	1	1	8.32	24.96%	10.48	26.72%
$PM_{10}$	1	0	0	5.54	24.77%	6.86	22.69%



# Conclusiones...

