PREDICTING AIR POLLUTION WITH OPEN DATA

PiperLab

AIR POLLUTION OPEN DATA
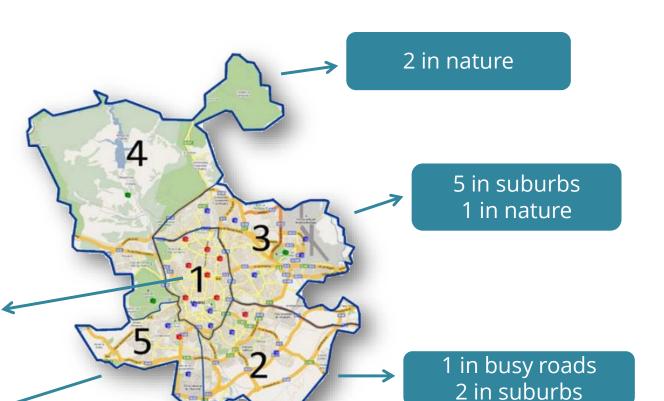
Madrid has **24 measuring stations** that collect **hourly data** about different air pollutants:

- NO_2
- O_3
- PM2.5
- PM10
- SO_2
- CO
- …

The city is divided into **five zones**. Zone 1 (central) is more populated and, therefore, where most of the stations lie. Stations are classified depending on the location: in busy roads, suburbs or nature.

2 in nature

5 in suburbs
1 in nature

7 in busy roads
2 in nature

1 in busy roads
2 in suburbs

1 in busy roads
2 in suburbs

Each hour, a **raw file** with the measures is made publicly available in the City of Madrid's Data Portal:

datos.madrid.es

¡MADRID! datos abiertos

WHY IS IT SO USEFUL?
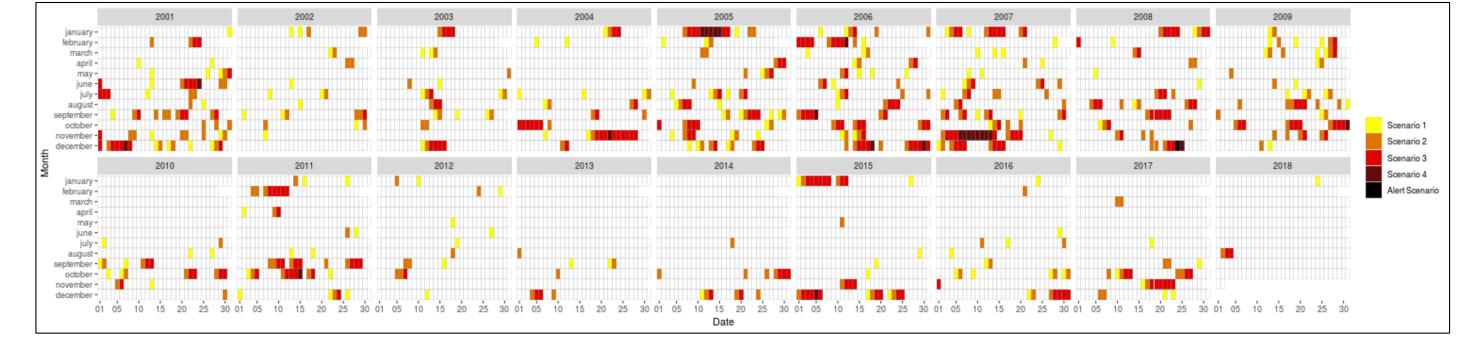
AIR POLLUTION → PROTOCOL ACTIVATION → SPEED, PARKING AND TRAFFIC RESTRICTIONS

Madrid has a **Protocol for High Pollution Episodes** that might be activated when **NO_2** levels exceed certain limits. Different scenarios of traffic restrictions apply depending on the severity.

Conditions for activation are a bit complex: when daily measures in a zone reach a certain level, an **action level** is set for that day. If this happens for one or more consecutive days, an **scenario** is activated. Restrictions bound to that scenario are applied the following day.

Action levels

Forewarning *(preaviso)*	Any 2 stations in a zone reach **180 μg/m³** for 2 consecutive hours, or any 3 stations reach 180 μg/m³ for 3 consecutive hours.
Warning *(aviso)*	Any 2 stations in a zone reach **200 μg/m³** for 2 consecutive hours, or any 3 stations reach 200 μg/m³ for 3 consecutive hours.
Alert *(alerta)*	Any 3 stations in a zone reach **400 μg/m³** for 3 consecutive hours.

Scenarios

Scenario 1	1 forewarning ⇒ **70 km/h speed** in M-30 and accesses.
Scenario 2	2 forewarnings or 1 warning ⇒ previous + **traffic and parking restrictions inside M-30.**
Scenario 3	3 forewarnings or 2 warnings ⇒ previous + **traffic restrictions all across the municipality.**
Scenario 4	4 warnings ⇒ previous + **more severe traffic restrictions.**
Alert Scenario	1 alert ⇒ previous + **more severe traffic and parking restrictions.**

The first protocol was designed in March 2015. The last version, explained above, came into force in October 2018.
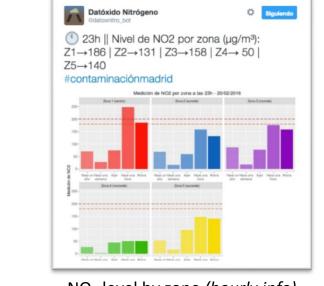
The image on the left shows when scenarios would have been activated in case the current protocol had been in force since 2001.

⚠ **Problem**: highest levels of NO2 usually occur **at night.** It is necessary to use predictive models to be able to notify restrictions sooner, when people are still awake.

OUR FIRST TOOL: TWITTER BOT

🐦 **@datoxnitro_bot**

🎯 **GOAL: MAKE INFORMATION -AND NOT JUST DATA- ACCESIBLE TO EVERYONE IN REAL TIME**

@datoxnitro_bot downloads and processes the air pollution open data and posts tweets explaining the **current situation**. If pollution rises and an scenario is activated, it **immediately warns all its followers**.
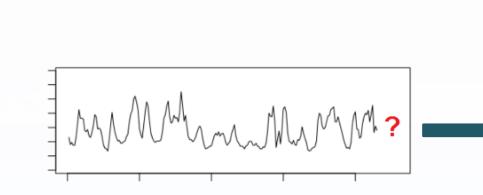
NO_2 level by zone *(hourly info)*

Pollution evolution in the last 4 days *(twice a day)*
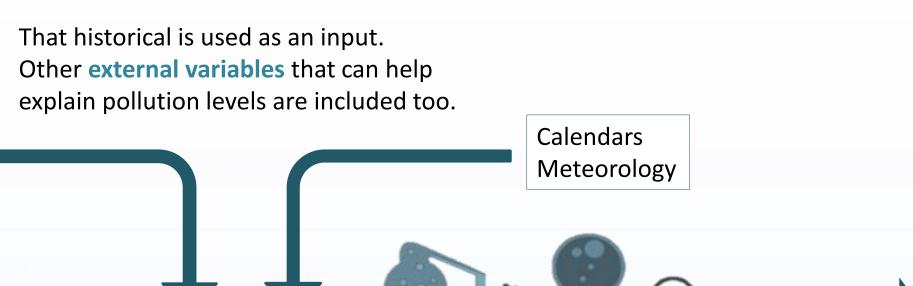
Scenario activation warning and restrictions info

PROOF OF CONCEPT: PREDICTING AIR POLLUTION

🎯 **GOAL: ANTICIPATE SCENARIOS SO CITIZENS, COMPANIES AND ADMINISTRATIONS CAN BE PREPARED FOR RESTRICTIONS**
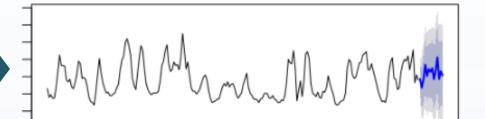
Statistical and computational methods can be used to get a **pollution forecast** for several days ahead. PiperLab, together with CITET, CEL and UNO has developed a project to prove that this can be achieved with high accuracy.

We have **historical data** on NO_2 pollution and want to know how levels will vary on a short future.

That historical is used as an input. Other **external variables** that can help explain pollution levels are included too.

Calendars
Meteorology

As a result we get a **forecast for the following days**. Accuracy decreases for farther predictions.

Horizon	Peak detection	Accuracy
24 hours	**90.24%**	87.98%
48 hours	73.17%	87.36%
7 days	51.22%	77.97%

An **algorithm** is used to detect patterns in data and project them in the future. We have focused on autoregressive families: ARIMA, Prophet.

NEXT STEPS: PREDICTION AS A SERVICE

Citizens, companies and administrations would benefit by having more time to react to the scenarios. This would help reduce incidents and encourage the use of public transport.

During 2019, PiperLab, CITET, CEL and UNO will work on developing

CLEAN AIR APP

a solution that will generate air pollution predictions in real time.

The applications will focus on the last mile delivery ecosystem, which is greatly impacted by traffic restrictions triggered by high air pollution levels. Predictions will be served through CITET to all members of the innovation cluster.

The service will be available in five cities in Spain:

- Madrid
- Barcelona
- Valencia
- Sevilla
- Zaragoza

www.piperlab.es