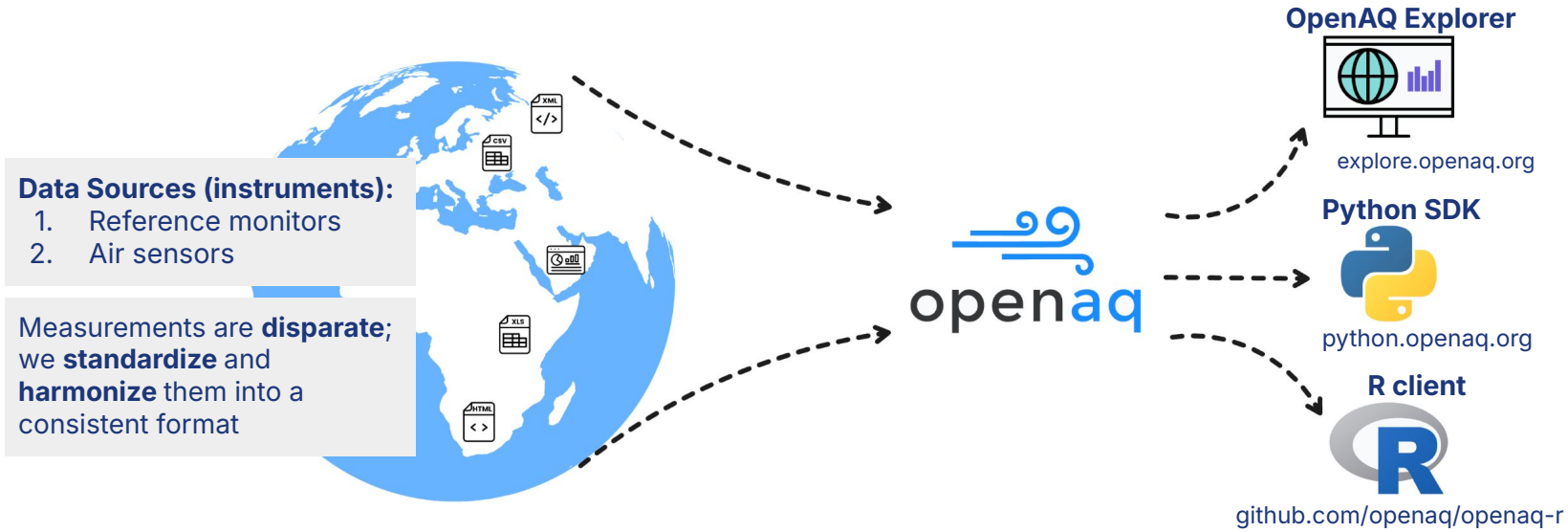
Bridging the Gap: Complementing Satellite-Derived Air Quality Data with Open-Source Ground Data

Minh Nghiem
Applied Data Fellow, OpenAQ
minh@openaq.org

About OpenAQ

We are the world's largest open-source air quality data aggregator and harmonizer

"By providing universal access to air quality data, OpenAQ empowers a global community of changemakers to solve air inequality."



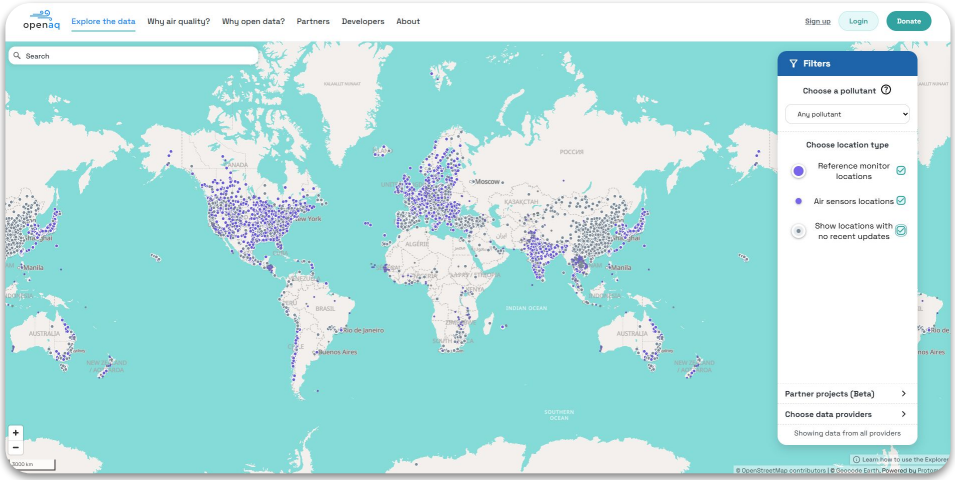
- Founded in 2015: see origin story at **link.openaq.org/origin** and **openaq.org/ten-years**
- U.S. 501(c)(3) non-profit organization as of 2018

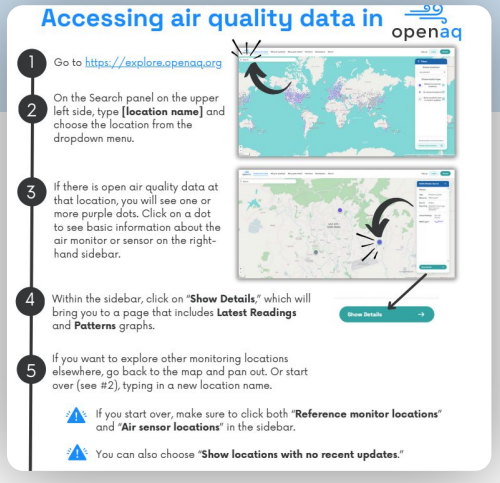
OpenAQ Explorer

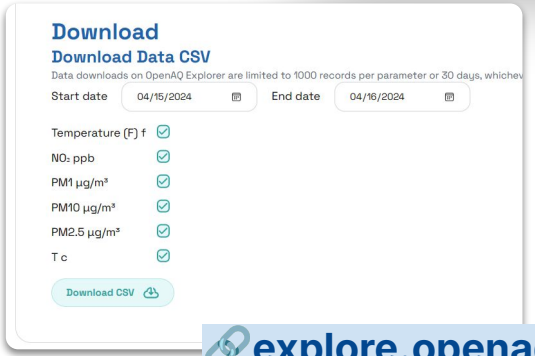
🌐 **explore.openaq.org**

🔗 **explore.openaq.org/getting-started**

- An interactive, open-source, user-friendly web application that allows users to visualize global air quality data, including real-time and historical measurements of pollutants
- Browse, filter by location or sensor type, and download data from government and low-cost sensors



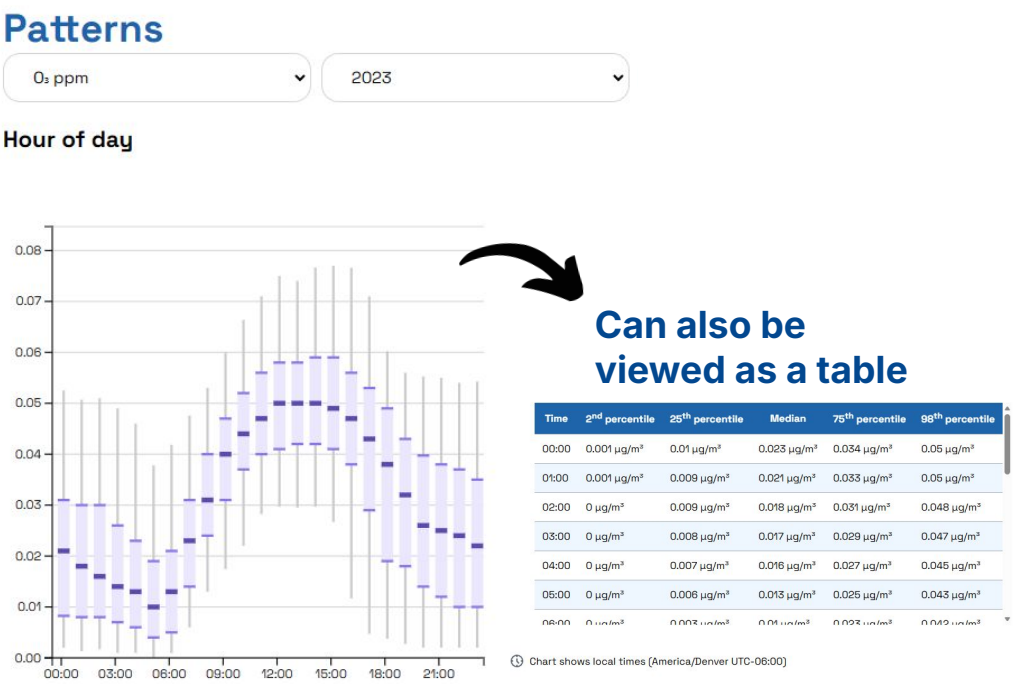




🔗 **explore.openaq.org/register**

Diurnal plots

Q: Which air pollutants have a dependence on time, day, and season, and why?

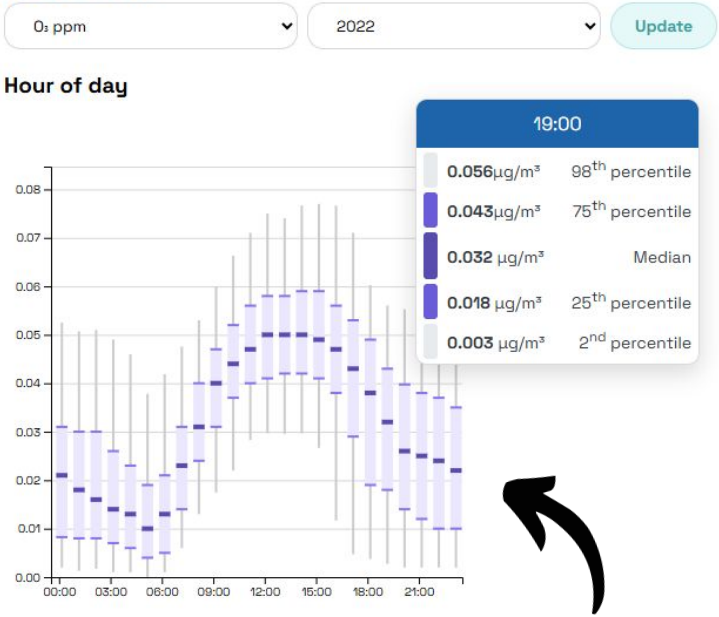


Can also be viewed as a table

Time	2nd percentile	25th percentile	Median	75th percentile	98th percentile
00:00	0.001 µg/m³	0.01 µg/m³	0.023 µg/m³	0.034 µg/m³	0.05 µg/m³
01:00	0.001 µg/m³	0.009 µg/m³	0.021 µg/m³	0.033 µg/m³	0.05 µg/m³
02:00	0 µg/m³	0.009 µg/m³	0.018 µg/m³	0.031 µg/m³	0.048 µg/m³
03:00	0 µg/m³	0.008 µg/m³	0.017 µg/m³	0.029 µg/m³	0.047 µg/m³
04:00	0 µg/m³	0.007 µg/m³	0.016 µg/m³	0.027 µg/m³	0.045 µg/m³
05:00	0 µg/m³	0.006 µg/m³	0.013 µg/m³	0.025 µg/m³	0.043 µg/m³

Chart shows local times (America/Denver UTC-06:00)

Five-number summary



Q: What is the median O_3 concentration during a certain day/time in my location?

OpenAQ R package

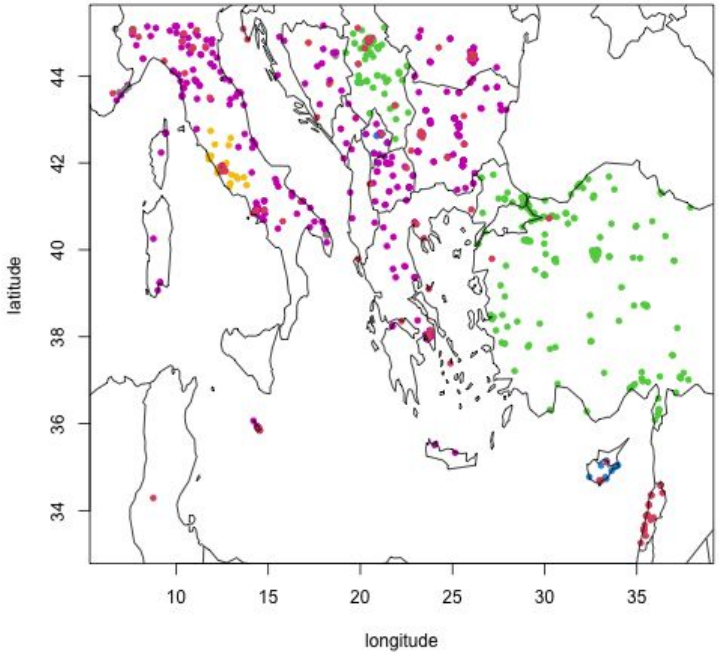
🔗openaq.github.io/openaq-r

```
library(openaq)

locations <- list_locations(
  bbox = c(xmin = 6.291990, ymin = 32.823129,
           xmax = 37.932615, ymax = 45.181129),
  parameters_id = 2,
  limit = 1000
)

plot(locations, col = locations$providers_id, pch = 20)
```

Code excerpt from OpenAQ R package documentation:
https://openaq.github.io/openaq-r/articles/plotting.html



OpenAQ Python SDK

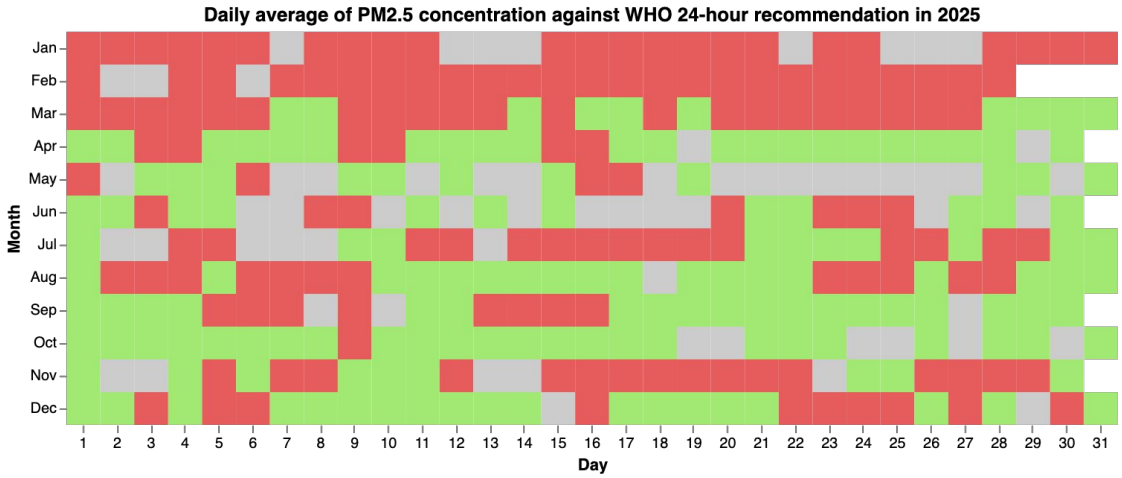
🔗python.openaq.org

Latest version: v1.0.0rc2

Features:

- **User-friendly helpers:** expressive error handling, preemptive error catching, automatic rate limiting
- **More intuitive data access:** Python objects-oriented instead of API json responses
- **Fully typed:** autocomplete, inline documentation

Python tutorials coming soon!



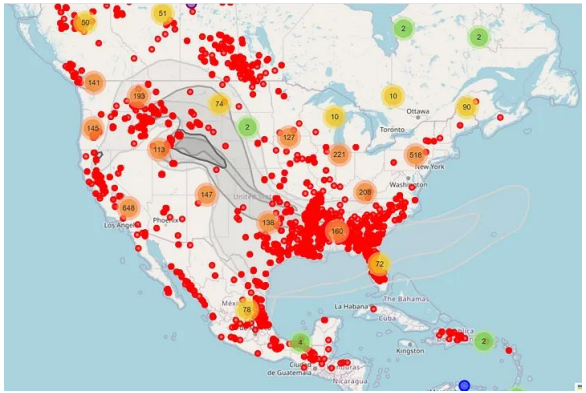
Visualization from OpenAQ Python SDK tutorials

Link to visualization

Use Cases: Ground Data from OpenAQ + Satellite Data

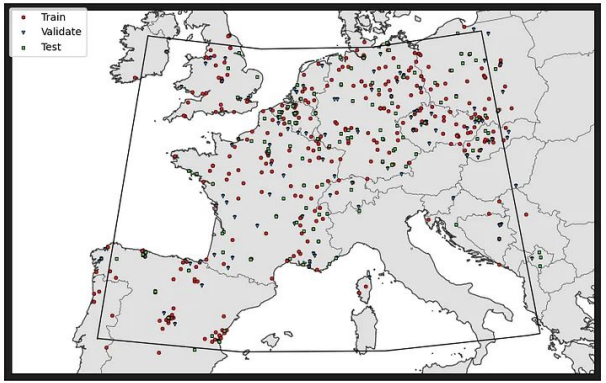
Feature Article: Using Satellites Plus Ground Sensing To Tackle Air Pollution

Read more: openaq.org/about/use-cases/validating-satellite-data



MELODIES MONET (NSF, NCAR and NOAA)

A toolkit that helps scientists more easily and efficiently conduct air quality research while also integrating remotely-sensed atmospheric plus ground data.



University of Eastern Finland

Researchers used OpenAQ to improve the conversion accuracy of satellite atmospheric aerosol optical depth (AOD) data

Use Cases: Ground Data from OpenAQ + Satellite Data

Global validation and hybrid calibration of CAMS and MERRA-2 $PM_{2.5}$ reanalysis products based on OpenAQ platform

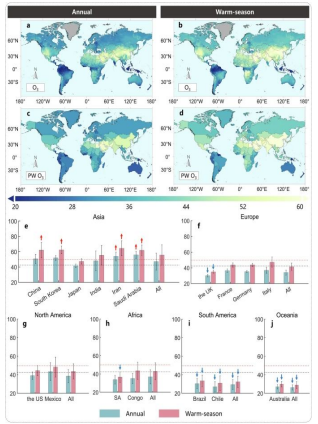
C Jin, Y Wang, T Li, Q Yuan. Atmos Environ, 2022. 118972. doi.org/10.1016/j.atmosenv.2022.118972

CAMS and MERRA-2 $PM_{2.5}$ reanalysis products were validated on both the regional and global scale using global in-situ $PM_{2.5}$ observations from 2017-2019 obtained from OpenAQ.

- Researchers applied machine learning models (specifically, Extremely Randomized Trees or ERT) to correct biases in CAMS and MERRA-2 using data in OpenAQ as the reference.
- Gap-free global daily $PM_{2.5}$ concentration maps were created for spatiotemporal analysis.

Substantially underestimated global health risks of current ozone pollution

Y Wang, et al. Nature Communications, 2025. 16, Article number: 102.
doi.org/10.1038/s41467-024-55450-0



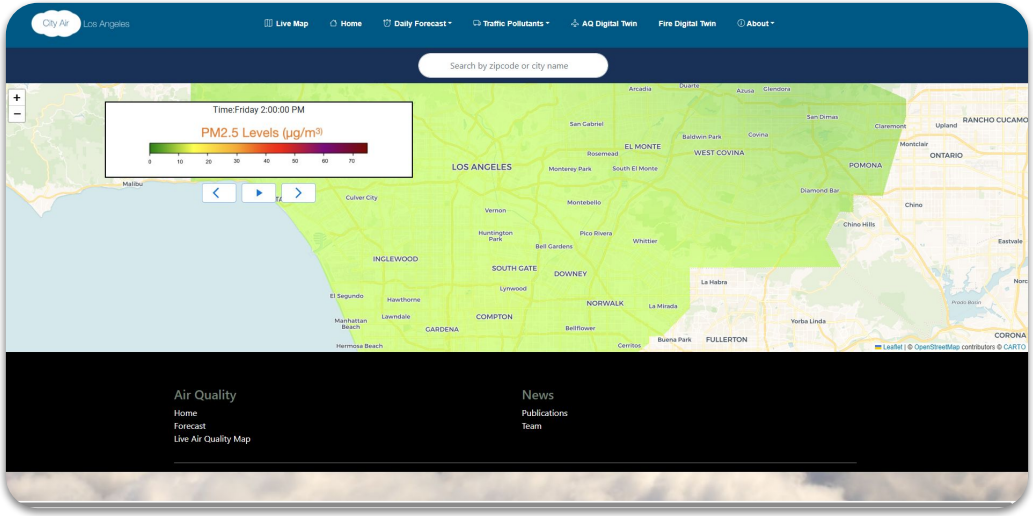
- OpenAQ as the primary source of in-situ MDA8 ozone measurements (2019-2021) to train and validate its global modelling framework.
- Estimated that 66.2% of the global population is exposed to excess ozone for short term (> 30 days per year), and 94.2% suffers from long-term exposure.

Use Cases: Ground Data from OpenAQ + Satellite Data

City of Los Angeles and Cal. State LA's
Predicting What We Breathe Forecasting System

airquality.lacity.gov
ai-aq.com/MapForecastCA

- NASA-funded
- Forecasting algorithm for air pollution events developed based on time-series measurements of satellite & ground data + machine learning (accuracy = 80-93%, compared to ground data alone)



OpenAQ as Project Partner:

Scaling Data Fusion Tools to Support Local Air Quality Managers in Latin America

NASA HAQAST Project PI: Carl Malings

Project description from the NASA HAQAST website: *haqast.org/staff/malings-dr-carl*

- Use of NASA satellite data and tools towards stakeholder needs such as optimizing air monitor placement, air quality forecasting, and developing datasets compatible with epidemiological studies.
- The team will leverage current and future NASA instruments, models and data products including Goddard Earth Observing System Composition Forecast (GEOS-CF), VIIRS, MAIA, AERONET, Pandora ground-based spectrometers, as well as non-NASA data from TROPOMI and local air quality monitoring systems.

Project goals / deliverables:

- Optimize air quality monitoring network expansion in Rio, Brazil
- Prepare for and supplement the NASA MAIA mission in Maule, Chile
- Enable a new air quality forecasting capability in Bolivia

openaq

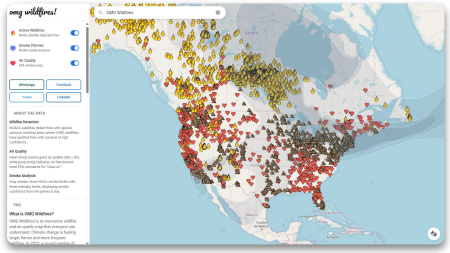
🌐 **openaq.org**

🔗 **links.openaq.org**

📧 **info@openaq.org**

Help support the OpenAQ platform through a sponsorship or a grant collaboration!

Read more use-case examples:
openaq.medium.com



OMG Wildfires
omgwildfires.com

Deciding on procuring air sensors or expanding your monitoring network?

Download:

Before You Buy:
A Guide to Choosing Air Sensors that Allow Fully Open Data Sharing

link.openaq.org/beforeyoubuy

aqihub.info

Documentation for air quality indices (AQI) in one convenient platform