Monitoring ground-level particulate matter concentrations from the synergism of space-borne measurements & machine learning techniques

J. C. Péré[1*], J. Staufer[1], B. Gratadoux[1] , N. Amsellem[1], M. Lucas[1], J. Cuesta[2] , G. Dufour[2], S. Tanguy[1], L. Chaumat[1] and A. Zabukovec[3]

[1] Thales Services Numériques, Labège France
[2] Laboratoire Inter-Universitaires des Systèmes Atmosphériques, Paris, France
[3] Centre National d'Etudes Spatiales, Toulouse, France
* ***jean-christophe.pere@thalesgroup.com***

Motivation

Comprehensive monitoring and mapping of ground-level air pollutants is **essential** for evaluating the population's exposure and assessing the resulting health effects. Satellites provide a **cost-effective way** for area-wide pollution monitoring. However, air quality products from space-borne sensors provide **only a vertically integrated information** of air pollutants in terms of optical properties and thus reflect only partly their concentrations at ground-level. We present a **novel approach so-called FLORIA** *(FLexible algORIthm for the monitoring of Air pollution based on Artificial Intelligence and satellite observations)* to calculate the concentration of fine particles with diameters less than 10 µm (PM10) at ground-level **using machine learning** (ML) techniques. The ML algorithm has been **trained with several years of high-resolution simulation of the CHIMERE air quality model** (version 2020). FLORIA has been applied to Sentinel-3/SYN aerosol optical thickness (AOT) at 550 nm and ERA-5 meteorological re-analysis to **provide maps of PM10 at ground-level** and at the spatial resolution of the satellite.

Presentation of FLORIA

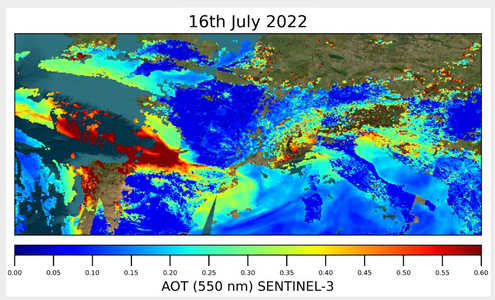
Data Hub for satellite measurements of Pollutants from TOA + Weather Data

https://github.com/thalesgroup/dhemeter

DHEMETER → Input → FLORIA (**Cat Boost ML algorithm**) → Output

Amount of pollution at ground level

Monthly PM10 maps are freely available at: https://floria-dev.pf-services-bu-aes.fr/



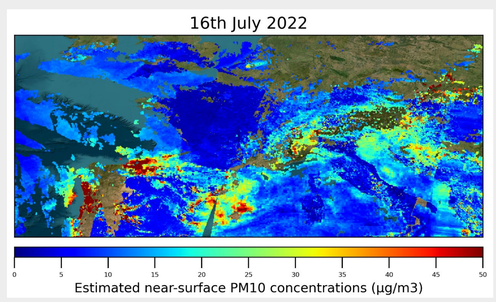


Figure 1 – Methodology to estimate near-surface PM10 concentration from SENTINEL-3/SYN AOT

Validation of PM10 product

FLORIA performance - estimated PM10 vs measured PM10 from 09/2023 – 08/2024

	Bordeaux	Lille	Paris	Lyon	Marseille	Nantes	Nice	Toulouse
Mean FLORIA (µg/m^3)	8.6	8.3	8.2	11.8	13.9	5.6	8.3	14.8
Mean Obs.(µg/m^3)	18.3	17.6	22.9	17.9	19.0	14.5	16.0	14.2
Bias (%)	-53	-53	-64	-34	-27	-61	-48	4
Error (%)	65	50	59	76	56	57	53	77
Temporal Correlation	0.43	0.44	0.19	0	0.11	0.23	0.36	0.28

Table 1 – Annual scores between FLORIA and in-situ PM10 for different background sites

- Near-surface **PM10 concentrations inferred by FLORIA are compared with *in situ* measurements** for 8 stations of the French AASQA/GEOD'AIR network for 09/23–08/24
- Scores per **season** (not shown) are **similar** to annual scores.
- The temporal correlation is **moderate** (*R* is up to 0.44).
- FLORIA **underestimates** the near-surface PM10 concentration (between 50 % and 77 %).

The CHIMERE model (v2020) used to generate the training database is known to have **a systematic negative bias on PM10** (6 µg/m^3 on average over Europe)

ON-GOING: The **ML training database** is being **consolidated with an optimized version** of the air quality model CHIMERE (**v2023**, Menut et al. 2024) where **PM10 biases have been reduced** and the correlation with in situ data is significantly improved **.**

Menut, L., Cholakian, A., Pennel, R., Siour, G., Mailler, S., Valari, M., Lugon, L., and Meurdesoif, Y.: The CHIMERE chemistry-transport model v2023r1, Geosci. Model Dev., 17, 5431–5457, https://doi.org/10.5194/gmd-17-5431-2024, 2024.

Quality of input data: Sentinel-3/SYN versus AERONET AOT

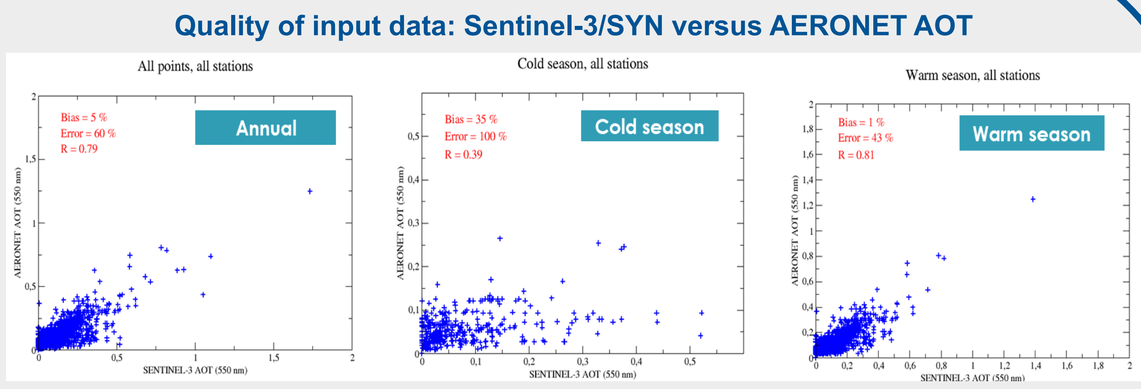


Figure 2 – Scores between Sentinel-3 and AERONET AOT (550 nm) averaged over 8 French sites

- Sentinel-3/SYN AOT (550 nm) are compared with photometric measurements from 8 **AERONET** sites close to PM10 stations to **evaluate the quality of input data.**
- The **annual average error** of Sentinel-3 AOT compared to photometric measurements is **60 %**.

ON-GOING: Data fusion of Sentinel-3, SENTINEL-5P and MODIS-MAIAC AOT to reduce uncertainties in input data and maximise the spatial coverage.

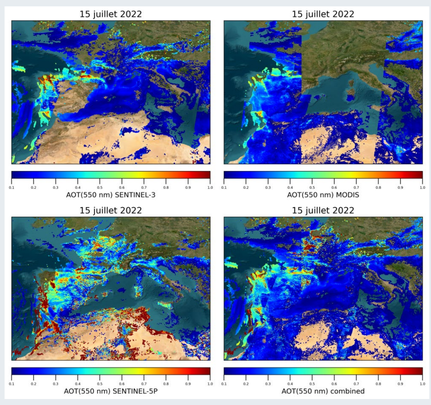


Figure 3 – Example of AOT (550 nm) for the 15th of July 2022 from the combination of SENTINEL-5P, SENTINEL-3 and MODIS satellite sensors.

The merging technique relies on a least squares fitting of the available τ_a spectra onto a 2nd order polynomial relation (Mélin et al., 2007):

$$\ln \tau_a(\lambda) = a_0 + a_1 \ln \lambda + a_2 \ln^2 \lambda$$

with $\tau_a(\lambda)$ the AOT at wavelength λ. Coefficients a_i are computed by least squares curve fitting.

Melin F, Zibordi G, Djavidnia S. Development and Validation of a Technique for Merging Satellite Derived Aerosol Optical Depth from SeaWIS and MODIS.. REMOTE SENSING OF ENVIRONMENT 108, p. 436-450, 2007.

Conclusion and perspectives

- Development of **4-yr training base** of atmospheric composition and meteorological variables over Europe/Northern Africa using the chemistry-transport model **WRF-CHIMERE**
- Implementation and testing of **Cat Boost ML algorithms to deduce ground-level PM10 concentrations from space-borne AOT**
- **PM10 was inferred from Sentinel-3 AOT data** at 5 km x 5 km resolution from 09/2023 – 08/20024. Comparison to in situ data in France showed **errors of 50 – 77 %.** Consequently, two major **areas for improvement** were identified and are being implemented:
  - **Update of ML training data base** with most recent CHIMERE version
  - **Data fusion** to reduce errors and improve spatial coverage (e.g. combining Sentinel-3, SENTINEL-5P and MODIS)

Présenté aux
RENCONTRES TECHNIQUES ET NUMERIQUES
2 décembre 2025
Toulouse, Centre de Congrès Pierre Baudis