Final Project Report

Project bb1260

Project: 1260 (bb1260)
Project title: Megacity Aerosol Composition by Satellite: A tool to study anthropogenic Emissions, Climate change and human Health
Principal investigator: Adrien Deroubaix
Allocation period: 2025-01-01 to 2025-12-31

Project Summary

Anthropogenic aerosol emissions from growing urban agglomerations contribute to air pollution with impacts on climate and human health. This project analysed the composition of carbonaceous aerosols (black carbon, BC; organic aerosol, OA) in megacity plumes by combining airborne observations from the two EMeRGe campaigns (Europe 2017; East Asia 2018) with complementary modelling and statistical approaches. The scientific objective was to quantify how far co-measured, satellite-observable trace gases (CO, NO_2, HCHO, O_3, SO_2) can constrain BC and OA, and to identify the conditions under which non-linear regressions provide added value over linear models.

Main Findings

Using 1-minute collocated aircraft observations across both campaigns, we find that CO is the most robust single proxy for BC (linear correlation $R^2 \approx 0.6$). OA is less constrained by trace gases overall ($R^2 \approx 0.3$), with improved association in urban plumes where OA shows moderate correlation with O_3 ($R^2 \approx 0.5$). These results highlight a strong contrast between BC (primarily emitted) and OA (mixed primary/secondary with stronger regime dependence).

Multiple linear regression remains limited for separating pollution types and chemical regimes. In contrast, non-linear machine-learning models (Random Forest in particular) achieve substantially higher predictive skill (reported in the submitted manuscript as $R^2 > 0.9$ for BC and $R^2 > 0.8$ for OA on standard cross-validation), consistent with their ability to represent non-linear gas–aerosol relationships. Across both campaigns, CO remains the dominant predictor for BC, while OA requires multiple predictors (at least CO, O_3, and NO_2).

Work Performed

The allocation supported (i) curation of a harmonised, collocated dataset across aerosols and trace gases at 1-minute resolution; (ii) identification of urban plume segments along flight tracks

using passive tracer simulations; (iii) implementation and comparison of linear and non-linear regression models with consistent metrics (R^2, MAE); and (iv) preparation of figures, tables, and documentation needed for journal submission and reproducibility.

Manuscript Status and Next Steps

A manuscript titled "Tracing carbonaceous aerosols through trace gas–aerosol relationships" is currently under major revision at npj Climate and Atmospheric Science (Nature Portfolio). The major-revision decision was received on 23 January 2026.

Following internal discussions, we will implement a minimal but essential revision focused on (1) robust validation that accounts for spatio-temporal autocorrelation, and (2) interpretability of the Random Forest results using SHAP diagnostics to document non-linearities and predictor interactions.

Publications and Data

- **2026.1.23: Major-revision manuscript (npj Climate and Atmospheric Science):** Tracing carbonaceous aerosols through trace gas–aerosol relationships.
- **Previous ACP submissions / preprints underpinning this effort (model evaluation and linear relationships):** egusphere-2024-516 and egusphere-2024-521.
- **Observational data access:** The original EMeRGe aircraft observations are available via the HALO database (registration required): https://halo-db.pa.op.dlr.de/. The processed matrices used in this study (combined gas–aerosol arrays used for the statistical and machine-learning analyses) are available at: **10.5281/zenodo.18098564**

Data Archival

Core scripts, processed collocation matrices, and figure outputs are preserved within the project workspace to ensure the revision can be completed reproducibly. Larger model outputs have been curated to retain only the variables required to regenerate the published diagnostics.

Request for Limited Continued Access (6 months)

To complete the major revision, I request a 6-month extension of access limited to interactive computing resources (analysis and plotting only; no new intensive simulations) and continued access to the existing project data and directories.

Kind regards,

Adrien Deroubaix
Max Planck Institute for Meteorology
adrien.deroubaix@mpimet.mpg.de