EGU26-2252, updated on 13 Jun 2026
https://doi.org/10.5194/egusphere-egu26-2252
EGU General Assembly 2026

Regression Based Apportionment with Low-Cost Air Quality Sensors Using Machine Learning and Emission Inventories

Rajat Sharma, Andry Razakamanantsoa, Erwann Rayssac, and Agnes Julian
Université Gustave-Eiffel, GERS-GIE, France (rajat.sharma@univ-eiffel.fr)

Effective air quality management depends on the availability of reliable, locally resolved pollution data. Low-cost sensors (LCS) offer high spatial and temporal resolution but are often constrained by data reliability challenges, primarily due to their reliance on field calibration. Conventional calibration methods typically require colocation with reference stations, a limitation in regions with sparse monitoring infrastructure. This study presents a model that integrates emission inventory (EI) data with machine learning (ML) to achieve source apportionment (SA) using LCS, demonstrated through a case study in Fianarantsoa, Madagascar. Unlike conventional ML calibration approaches benchmarked solely against collocated reference monitors, the proposed method exploits a distributed sensor network in which each device is cross-validated by at least two neighbouring sensors within 500 m. Conventional calibrations frequently suffer from sensor- and site-specific biases, provide limited source-specific information, and are often hindered by proprietary algorithms. To address these issues, a Data Reliability Indicator (DRI) is introduced to evaluate LCS performance across high-, middle-, and low-income country contexts. The findings demonstrate that LCS, when supported by emission inventories and network-based cross-validation, can deliver reliable source apportionment and high-resolution air quality insights, even in regions with minimal reference-grade monitoring.