EGU25-1317, updated on 21 May 2025
https://doi.org/10.5194/egusphere-egu25-1317
EGU General Assembly 2025

Empowering Data-Driven Air Quality Management through low-cost sensors: Insights from the Indo-Gangetic Plains

Sachchida Tripathi
Indian Institute of Technology Kanpur, Department of Civil Engineering and Department of Sustainable Energy Engineering, Kanpur, India (snt@iitk.ac.in)

In India, the sparse air quality monitoring infrastructure beyond urban centers limits our understanding of fine particulate matter ($PM_{2.5}$) spatiotemporal variability, exposure dynamics, and their health and ecological impacts. To address this challenge, the Ambient Air Quality Monitoring in Rural Areas using Indigenous Technology (AMRIT) project was launched, deploying 1,400 low-cost air quality sensors (LCS) across diverse demographics and land-use settings in the states of Uttar Pradesh and Bihar, within the Indo-Gangetic plains (IGP). Additionally, the Dynamic Hyper-local Source Apportionment (DHSA) project was initiated to explore the potential of LCS for real-time source apportionment, facilitating pollution source identification and providing policymakers with critical data for effective air pollution control. Together, these initiatives aim to capture regional pollution dynamics and impacts beyond urban centers for the first time in India.

Our investigation showed that the AMRIT-LCS network effectively captured fine-scale pollution patterns across urban and rural areas of Bihar, with a mean concentration of 135 $\mu g/m^3$. Rural areas exhibited 6-8% higher pollution levels than urban settings. Within urban areas, pollution levels were notably higher in suburban and small-town regions, highlighting the need for targeted pollution management in these emerging cities. Leveraging machine learning (ML) frameworks, the village-level exposure assessment revealed significant spatiotemporal heterogeneity in $PM_{2.5}$ exposure, with consistently high levels in the northwestern districts and a mortality rate (per lakh) of 50 (95% CI: 33–64) in the region. Data-driven techniques were applied to delineate air sheds for future air quality management. Five distinct air-sheds were identified, with three (in the north) and two (in the south) of the Ganges River, flowing west to east and dividing Bihar into two unequal parts. Consistent with the exposure model, the air-sheds over the northwestern and north-central regions were more polluted. Additionally, the LCS network was used to identify pollution hotspots by leveraging satellite-driven datasets within a deep learning framework, across various temporal scales.

These findings underscore the critical need for high-density monitoring networks to capture local variability and generate spatially resolved estimates. Under the DHSA project, ML techniques enabled real-time, accurate predictions of pollution source contributions using LCS datasets, with a mean absolute error of less than 5%. This approach strengthens the potential for deploying widespread LCS networks capable of high-resolution source apportionment with spatiotemporal

precision. The data-driven approach provides a solid foundation for designing decentralized mitigation strategies and comprehensive policy frameworks aimed at improving air quality and public health in the IGP. Thus, the portable LCS technology offers a promising alternative to conventional air quality monitoring methods, presenting a viable solution not only for India but also for data-sparse regions globally, contributing to the achievement of sustainable development goals.

Keywords: $PM_{2.5}$, Low-Cost Sensors, Indo-Gangetic Plains, Machine Learning, Exposure Assessment, Urban-Rural Disparity, Airsheds, Hotspot Analysis, Source Apportionment